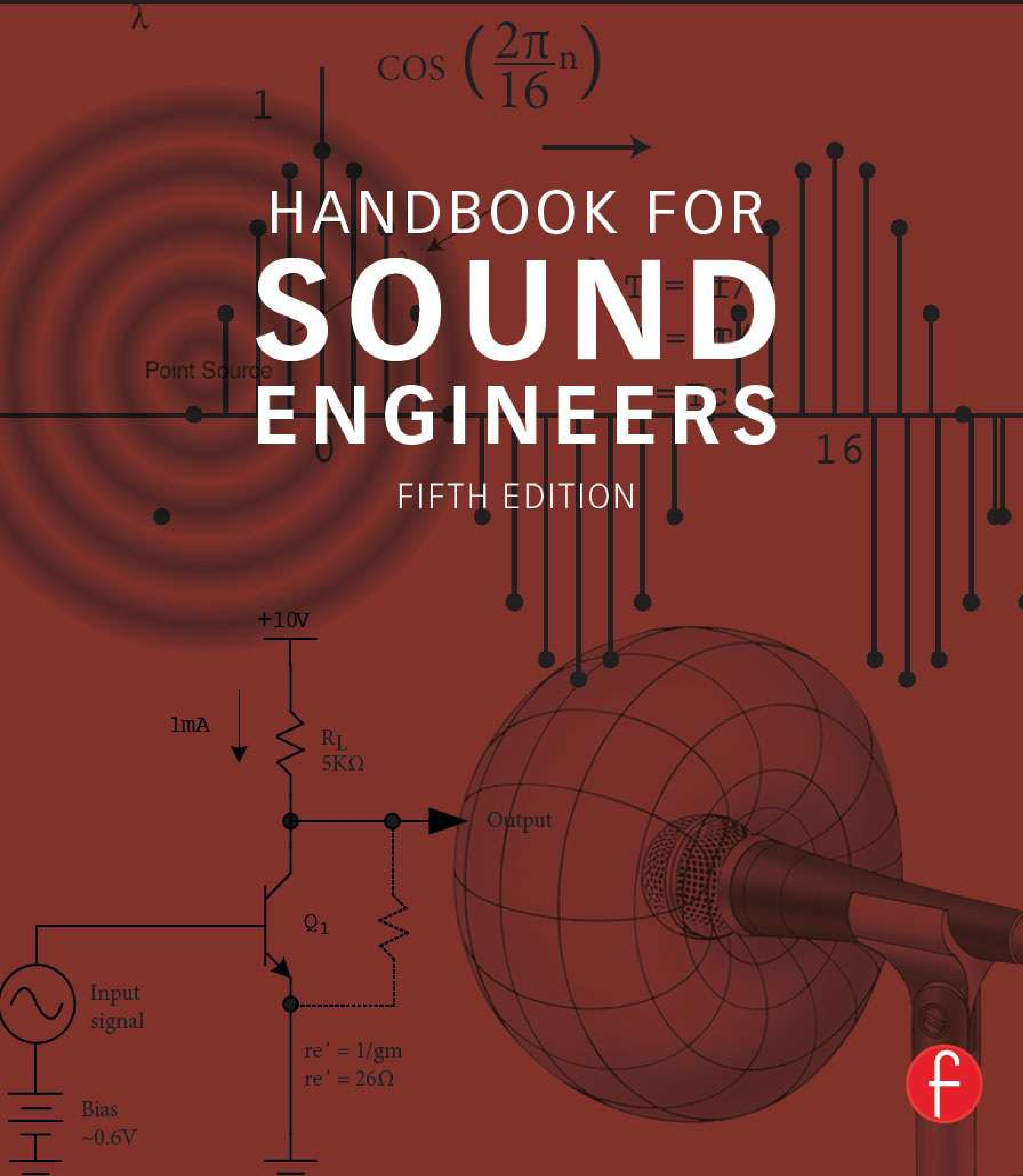
EDITED BY GLEN BALLOU

Audio Engineering Society Presents

HANDBOOK FOR SOUND ENGINEERS

FIFTH EDITION



Handbook for Sound Engineers

Fifth Edition

Handbook for Sound Engineers

Fifth Edition

Glen M. Ballou
Editor

Focal Press
Taylor & Francis Group
NEW YORK AND LONDON

First published 1987 by Howard W. Sams & Co., Inc.

This edition published 2015 by Focal Press
70 Blanchard Road, Suite 402, Burlington, MA 01803

and by Focal Press
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Focal Press is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data

Handbook for sound engineers / Glen M. Ballou, editor. -- Fifth edition.
pages cm
ISBN 978-0-415-84293-8 (hardback)
1. Acoustical engineering--Handbooks, manuals, etc. 2. Sound--Recording and reproducing--Digital techniques.

I. Ballou, Glen.
TK7881.4.H36 2015
621.389'3--dc23
2014036047

ISBN: 978-0-415-84293-8 (hbk)
ISBN: 978-0-203-75828-1 (ebk)

Typeset in Times New Roman and Optimum
By Glen Ballou

Contents

Chapter 50

Preface

When the first edition of the *Handbook for Sound Engineers* came out in 1987, it was subtitled *The New Audio Cyclopedia* so that people who were familiar with Howard Tremain's *Audio Cyclopedia* would understand that this is an updated version of it. When the third edition came out in 2002, the subtitle was dropped and the *Handbook for Sound Engineers* has stood on its own since.

We have seen a tremendous change in the field of sound and acoustics since 1987. Digital has certainly made its place in all forms of audio, but analog systems will still be around for a long time. After all, sound is analog and the transfer of sound waves to a microphone is analog, and from the loudspeaker to our ears is analog. However, as Steve Dove said, "digital won"; and with all of the new digital circuitry in microphones, electronic circuits, loudspeakers, and test equipment, digital has, without a doubt, revolutionized the way we produce, reproduce, and measure sound.

The *Handbook for Sound Engineers* discusses sound through eight sections: *Introduction to Audio and Acoustics, Acoustics, Electronic Components, Electro-Acoustic Devices, Audio Electronic Circuits and Equipment, Recording and Playback, Design Application*, and *Measurements*.

Much has changed in this, the fifth edition of the *Handbook for Sound Engineers*. Four new chapters have been added: *Subjective Methods for Evaluating Sound Quality, Hearing Physiology—Disorders—Conservation, Worship Styles in the Christian Church*, and *Surround Sound for Cinema*.

We have new authors for six original chapters or sections of chapters: *Stadiums and Outdoor Venues, Loudspeakers, Preamplifiers and Mixers, Amplifier Design, Sound System Design*, and *Message Repeaters, Museum and Tour Group Systems, Voice Evacuation/Mass Notification Systems*.

The book also covers message repeaters, interpretation systems, assistive listening systems, intercoms, modeling and auralization, surround sound, and personal monitoring, subjects that are not often covered in audio books.

This edition is not a replacement for the previous editions as the information in the past four editions is relevant today; however the fifth edition brings us into the digital age and numerous new ideas. A good audio library will contain all five editions of this book plus books written by many other authors including Don and Carolyn Davis, Pat Brown, Ken Pohlmann, Bob Cordell, Doug Jones, Wolfgang Ahnert, Stefan Feistel, David Miles Huber, Steve Lampen, Peter Mapp, and Craig Richardson, all authors who have contributed in this book.

Technical authors write to share their knowledge that they spent a lifetime gathering so we can learn from them. No one person can be knowledgeable in all the fields of sound and acoustics. I have been extremely fortunate to have exceptional authors, who are considered by many the most knowledgeable in their field, give of their time to contribute so much in this book.

Glen Ballou

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks are listed below. In addition, terms suspected of being trademarks or service marks have been appropriately capitalized. Focal Press cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

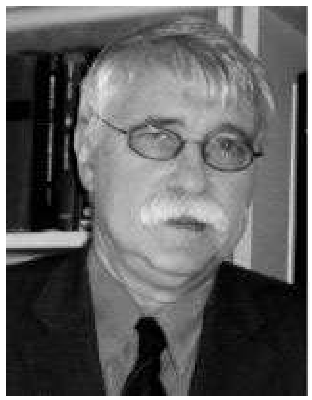
Abffusor is a trademark of RPG Diffuser Systems, Inc.
Audio Spotlight is a registered trademark of Holosonic Research Labs, Inc.
AutoCAD is a trademark of Autodesk, Inc.
Beldfoil is a registered trademark of Belden Electronics Division.
Bidiffusor is a trademark of RPG Diffuser Systems, Inc.
Biffusor is a trademark of RPG Diffuser Systems, Inc.
Boogie is a registered trademark of Mesa Engineering.
C-Quam is a registered trademark of Motorola, Inc.
Co-Netic® is a registered trademark of Magnetic Shield Corp.
CobraCAD is a registered trademark of Peak Audio.
CobraNet is a registered trademark of Peak Audio.
dbx is a trademark of DBX Professional Products, AKG, Inc.
Diffractal is a trademark of RPG Diffuser Systems, Inc.
Diffusor is a trademark of RPG Diffuser Systems, Inc.
Dolby is a trademark of Dolby Laboratories Licensing Corporation.
Dolby SR is a trademark of Dolby Laboratories Licensing Corporation.

Electro-Voice, a product brand of Telex Communications, Inc.
Enkasonic is a registered trademark Colbond.
Fiberglas is a registered trademark of Owens Corning.
Flutterfree is a trademark of RPG Diffuser Systems, Inc.
FMX is a registered trademark of Broadcast Technology Partners.
Freon-11 is a trademark of E.I. Dupont de Nemours & Co., Inc.
Holophone® is a registered trademark of Rising Sun Productions Limited.
Hot Spot is a trademark of Galaxy Audio.
HyMu® is a registered trademark of Carpenter Technology Corp.
InGenius® is a registered trademark of THAT Corporation.
ISO-MAX® is a registered trademark of Jensen Transformers, Inc.
Isoloop is a registered trademark of 3M Company.
ITE is a trademark of Synergetic Audio Concepts.
Jensen Twin-Servo is a registered trademark of Jensen Transformers, Inc.
Jensen-Hardy Twin-Servo® is a registered trademark of Jensen Transformers, Inc.
Korner-Killer is a trademark of RPG Diffuser Systems, Inc.
LEDE is a trademark of Synergetic Audio Concepts.
Magneplanar is a registered trademark of Magnepan, Inc.
Mumetal® is a registered trademark of Telcon Metals, Ltd.
Mylar is a trademark of E.I. Dupont de Nemours & Co., Inc.
Omniffusor is a trademark of RPG Diffuser Systems, Inc.
OpticalCon® is a registered trademark of Neutrik AG.
OptiFiber® is a registered trademark of Fluke Networks.
PAR is a trademark of Synergetic Audio Concepts.
PCC is a registered trademark of Crown International, Inc.
Performer is a trademark of Mark of the Unicorn, Inc.
Permalloy® is a registered trademark of B & D Industrial & Mining

Services, Inc.
PZM is a registered trademark of Crown International, Inc.
QRD is a trademark of RPG Diffuser Systems, Inc.
QRDX is a trademark of RPG Diffuser Systems, Inc.
Romex® is a registered trademark of General Cable Industries, Inc.
RPG is a trademark of RPG Diffuser Systems, Inc.
SASS is a trademark of Crown International, Inc.
Scotchguard is a trademark of 3M Co.
Series Manifold Technology is a trademark of Electro-Voice, Inc.
Source Independent Measurements is a trademark of Meyer Sound Laboratories, Inc.
Stereo Ambient Sampling System is a trademark of Crown International, Inc.
TEF is a trademark of Gold Line.
Teflon is a trademark of E.I. Dupont de Nemours & Co., Inc.
Tektronix is a registered trademark of Tektronix.
Time Align is a trademark of Ed Long Associates.
Triffusor is a trademark of RPG Diffuser Systems, Inc.
UniCam® is a registered trademark of Corning, Inc.
Variac is a registered trademark of Gen Rad.
Voice Development System is a trademark of Musicode.

Contributors

Professor Dr. Ing. Habil. Wolfgang Ahnert



Dr. Wolfgang Ahnert graduated in Dresden in 1975 and from 1975 to 1990 he worked in an engineering office in Berlin and in 1990 founded ADA Acoustic Design Ahnert. In 2000 he founded SDA Software Design Ahnert GmbH to increase the output in writing source codes for acoustic and electronic applications. In 2002 the ADA foundation was established to support universities and colleges with acoustic software including EASE and the measurement tools EASERA and EASERA SysTune. The Ahnert Feistel Media Group was established in 2006 to coordinate all the above activities.

In 1993 he became an honorary professor at the Hochschule fuer Film und Fernsehen Potsdam, and in 2001 he became an honorary professor at Lomonossov University in Moscow. Since 2005 he has been a visiting professor at the Rensselaer Institute for Architecture in Troy, New York.

Dr. Ahnert has been a member of the AES since 1988 and was made a fellow of the Audio Engineering Society in 1995. He has been a fellow of the Acoustical Society of America since 2005. He is a member of the German DEGA and the British Institute of Acoustics. More than sixty five scientific lectures by Dr. Ahnert have been published.

Dr. Ahnert is one of the authors of the book *Fundamentals of Sound Reinforcement* (1981, and 1984 in Russian) and the book *Sound Reinforcement—Basics and Practice* (1993, and fully updated in English in 2000, in Chinese in 2002, and in Russian in 2003). He wrote with coauthors chapters in the 3rd edition of *Handbook for Sound Engineer, [Handbuch der Audiotechnik]* (2008 Springer) and *Akustische Messtechnik* (2008 Springer), both in German.

Ronald G. Ajemian

Ronald G. Ajemian is an instructor at the Institute of Audio Research, New York, NY, where he has been teaching audio technology for over 30 years. Ron was employed with the Switched Services Department of Verizon for 32 years in New York City and recently took an early retirement.

Ron is a graduate of RCA Institutes and attended the school of electrical engineering at Pratt Institute of Brooklyn, NY. He is an author, and contributed many articles in the field of audio electronics, telephony, and fiber optics.

He is a member of many professional organizations, including the Audio Engineering Society (AES), Society of Motion Pictures and Television Engineers (SMPTE), the Telephone Pioneers of America, Communication Workers of America (CWA), and the Optical Society of America, to name a few. Ron is sometimes referred to as Dr. FO (fiber optics), for his expertise in the field of fiber optics. Mr. Ajemian is also a guest lecturer at NYU and for the Audio Engineering Society. Mr. Ajemian held the position of the Chair of the AES New York Section in 2000-2001, and is currently on the AES New York Section Committee. He is currently the Chair of the AES Standards Task Group SC-05-02-F on Fiber Optic Connections and Co-Chair of the AES Technical Committee on Fiber Optics for Audio.

Mr. Ajemian is owner and consultant of Owl Fiber Optics in New York, specializing in fiber optic technology and education for pro-audio/video and broadcast.

George Alexandrovich

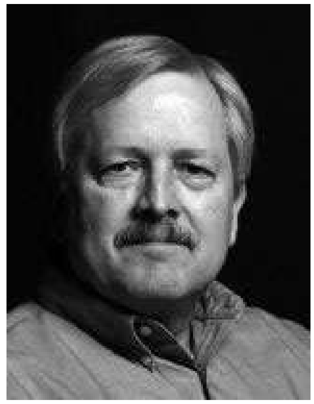
George Alexandrovich was born in Yugoslavia and attended schools in Yugoslavia, Hungary, and Germany. In the United States, he studied at the RCA Institute and at Brooklyn Polytech, earning a B.S.E.E. At Telectro Industries Corp., he was involved in the design and development of the first tape recorders and specialized military electronic test and communications equipment.

After service in the Korean war, he ran Sherman Fairchild's private research lab. At Fairchild Recording Equipment Corp., he designed and manufactured turntables, tonearms, pickups, mixing consoles, amplifiers, equalizers, reverberation chambers, the first light-activated compander, Autoten, Lumiten compressors, limiters, and a line of remote-controlled audio components. He also designed the first professional multichannel portable console, a disk-cutting lathe, and stereo cutters. As vice president and general manager his responsibilities included designing and manufacturing the Voice of America recording facilities, NBC-TV studio consoles for Johnny Carson, Huntley-Brinkley Newsroom, KNX, KCBS, and other radio stations.

At Stanton Magnetics, Inc., as vice president of field engineering and the professional products manager for phonograph cartridge research, he traveled extensively, holding seminars, giving lectures, and conducting conferences.

George is a fellow and a past governor of the AES. He holds eighteen patents in the audio field and is past chairman of the Electronics Industry Association (EIA) P8.2 Standards Committee.

Ron Baker



Ron Baker is a principal of Wrightson, Johnson, Haddon and Williams (WJHW), a multi-discipline consulting design firm based in Dallas, Texas, where he heads up the audio design department. Ron has served as primary audio designer for dozens of stadiums and arenas all over the world, as well as extensive work on large scale hospitality projects. In 1972, Ron began his work in the audio industry as a sound system installer and service technician for a division of Comcast. In 1981, he made the transition to consulting by working as an audio designer for Joiner, Pelton, Rose; later known as the Joiner Rose Group. It was there Ron began designing systems for profession sports stadiums and collaborating with manufacturers to develop product better suited to the stadium market. Ron joined WJHW in 1990 as senior audio designer, where he continues to provide design solutions for numerous professional and collegiate sporting facilities.

Glen Ballou

Glen Ballou graduated from General Motors Institute in 1958 with a bachelor's degree in Industrial Engineering and joined the Plant Engineering Department of the Pratt & Whitney Aircraft Division of United Technologies Corporation. There he designed special circuits for the newly developed tape control machine tools and was responsible for the design, installation, and operation of the 5,000,000 ft^2 plant public address and two-way communication system.

In 1970, Glen transferred to the Technical Presentations and Orientation section of United Technologies' corporate office, where he was responsible for the design and installation of electronics, audio-visual, sound, and acoustics for corporate and division conference rooms and auditoriums. He was also responsible for audiovisual and special effects required for the corporation's trade show program. Glen transferred to the Sikorsky Aircraft division of United Technologies as manager of Marketing Communications in 1980, where his responsibilities included the Sikorsky trade shows and special events program, plus the operation and design of all conference rooms.

After his retirement from Sikorsky, Glen and his wife, Debra, opened Innovative Communications, a company specializing in sound system design and installation, and technical writing.

Glen is the editor/author of the 1st, 2nd, 3rd, and 4th and 5th editions of the *Handbook for Sound Engineers*. He also was a contributing author for *The Electrical Engineering Handbook* (CRC Press). Glen has written many article for Sound and Communications, Sound and Video Contractor, and Church Production magazines.

He has been active in the Audio Engineering Society (AES) as Governor, three times Papers Chairman and four times Facilities Chairman, Vice Chairman, and Chairman of the 1989 AES Convention, for which he received the Governor's award. He was also a member of SMPTE and the IEA.

Steve Barbar

Steve Barbar has been working professionally in the audio industry for a little more than 25 years, and has been infatuated with sound and sound technology throughout his life. He received his first open reel tape recorder while in the third grade, his first electric guitar as a gift at age 10, and started playing in bands at 15. While studying at the University of Maryland, he joined the technical staff at the student union and assisted in maintaining the audio systems for film and public address, as well as the broadcast equipment in the campus radio station. He helped to organize rentals of large format public address systems for music concerts on campus, and operated these systems for many performers.

In 1981 he joined Systems Wireless Ltd. who specialized in providing large multi-channel wireless systems and frequency coordination for large events. These included the Emmy Awards, Grammy Awards, Country Music Awards, Tony Awards, Super Bowl, Pro Bowl, Miss America Pageant, Liberty Weekend, and many more. This also included work for facilities such as NBC, CBS, ABC, PTL, CBN, USA Today, CNN, NFL Films Walt, Disney, Universal, and others.

Steve joined Lexicon Inc. in 1984 as Advanced and Broadcast Products Manager to work with a then emerging technology—digital audio. He oversaw development of Lexicon's broadcast quality time compression systems which earned an Emmy award. In addition, he worked on development of the Lexicon 480L—the first commercial signal processor with digital interconnectivity, which became an industry standard for both digital reverberation as well as all digital mastering. In 1989, he and Dr. David Griesinger developed LARES—Lexicon Acoustic Reinforcement and Enhancement System that alters room acoustics electronically. This system was nominated for a Tech award by Mix magazine in 1993.

In 1995 Mr. Barbar formed LARES Associates, which designs, integrates, and develops comprehensive systems for electronic acoustic enhancement. In addition, LARES Associates performs ongoing research in the fields of acoustics, electro-acoustics and human neurology. Since that time, LARES Associates has installed hundreds of acoustic enhancement systems worldwide in concert halls, opera houses, performing arts centers, houses of worship, sports arenas, recording facilities, soundstages, rehearsal rooms, and outdoor venues. These include the Netherlands Opera, Berlin Staatsoper, Royal Danish Opera, the Bolshoi Theatre, Sunset Center (home to the Carmel Bach Festival), Milwaukee Symphony, Indianapolis Symphony, Millennium Park in Chicago, LDS Conference Center (home to the Mormon Tabernacle Choir), and many others. The LARES system has also been used in many prominent events and festivals, such as the first staging of Puccini's Turandot in the Forbidden City in1998, and Vienna Fest featuring the Vienna Philharmonic Orchestra.

He is a member of the Audio Engineering Society, the Acoustic Society of America, the Institute of Acoustics, and the Society of Motion Picture and Television Engineers. He has been an invited panelist for these organizations, in addition to presenting scientific papers and lectures. He has been an invited instructor for broadcast audio training seminars at NBC; has participated in the Stereo Audio Technical Training Workshops for NPR, and has also been invited to lecture at numerous regional AES meetings and Universities.

Alan C. Brawn CTS, ISF, ISF-C

Alan C. Brawn is a principal of Brawn Consulting an audio visual and digital signage consulting, educational development, and market intelligence firm with national exposure to major manufacturers, distributors, and integrators. Brawn was previously president of Telanetix and National Business Development and Product Marketing Director, Samsung Electronics.

Alan is an AV industry veteran with experience spanning three decades and he was one of the founding members of Hughes-JVC. He is a recognized author for leading AV industry magazines such as Systems Contractor News, Digital Signage Magazine, and AV Technology.

Mr. Brawn is a member of the Imaging Science Foundation and Managing Director of ISF Commercial. Alan is CTS certified and a senior faculty member of InfoComm and was moderator of the ANSI Projected Images Task Group creating a new contrast standard for the industry.

He is a member of the International Commission on Display Metrology for the Society of Information Display. Brawn is a Director of the Digital Signage Experts Group certifying professionals in the digital signage industry and the past chairman of the Digital Signage Federation. He was awarded the coveted InfoComm Volunteer of the Year Award for 2011 and Hall of Fame recognition from rAVe in 2004.

Pat Brown

Pat Brown is a 1978 graduate of the University of Louisville with a degree in electrical engineering technology. With a background as a musician, sound technician, retail music store owner, contractor, and consultant, he is familiar with the many aspects of audio. He served as an associate instructor to audio legend and SynAudCon founder, Don Davis. Pat along with his wife Brenda own and operate SynAudCon and have been providing training for the audio industry since 1996. Pat, the main instructor for SynAudCon, has the unique gift of being able to share his knowledge of complex technical topics in an easy-to-understand real world manner through multi-media demonstrations. SynAudCon offers both in-person and web-based audio training. SynAudCon is recognized as a industry leader in audio education.

Pat writes articles for several publications. He authored two chapters in the 3rd edition, three chapters in the 4th edition, and four chapters for the 5th edition for Glen Ballou's *Handbook for Sound Engineers* and is co-author of *Sound System Engineering, 4th Edition*. He considers SynAudCon web-based audio training courses as his most complete "book," being able to explain the concepts and principles with animations and demonstrations.

He was chosen as NSCA "Educator of the Year" in 2005, NSCA "Mover and Shaker" Award in 2011, and the "Peter Barnett" Award in 2012.

SynAudCon has given private seminars to Kennedy Space Center, Disney World, Sea World, IMAX, Purdue University, the US military, and to many audio integrators and manufacturers. He has conducted seminars in Europe, the Middle East, Africa, Asia, South America, Australia and Canada.

Pat and Brenda Brown founded a second business—Electro-acoustic Testing Company (ETC). ETC, Inc. measures loudspeakers to produce data files for use in room acoustics modeling programs. ETC, Inc. was the first company of its kind in the USA. This experience of developing a measurement facility and measuring hundreds of loudspeakers from dozens of manufacturers has allowed him to develop some methodologies and specifications that have influenced a number of industry standards.

Dominique J. Chéenne, Ph.D., INCE

Dominique J. Chéenne immigrated to the United States in 1979 and opened C & C Consultants, a consulting practice specializing in architectural acoustics and environmental noise control. He received his Masters and his Ph.D. in Electrical Engineering from the University of Nebraska (Lincoln), and he also holds an undergraduate degree from the University of Caen (France) as well as continuing education certificates in the area of industrial control systems. In 1995, Dr. Chéenne joined the faculty of Columbia College Chicago where he currently directs the acoustics program in the Audio Arts & Acoustics Department where in holds the rank of Full Professor.

Since its inception C & C Consultants has provided design services on hundreds of projects including individual residences, performing arts centers, schools, offices, factories, churches, as well as local, state and Federal government facilities. The firm's clients include more than a hundred architectural, engineering, development, and legal firms, and its portfolio spans over seventeen states and three continents. C & C Consultants operates from Lincoln (NE), and Chicago (IL).

A member of the Partage Group LLC, C & C Consultants has conducted business since 1979. The firm does not represent any manufacturer and/or products, does not derive any income from commissions or product sales, and solely represents the Owner's interests while working on a project. The firm's designs and solutions are based on solid scientific principles combined with decades of practical experience and an appreciation of the costs, regulatory, and aesthetics issues to be considered.

From initial site survey to the proof of performance, C & C Consultants uses state-of-the-art tools during all phases of the design process. The firm's portfolio is a testimony that it has the ability and experience to competitively meet the needs of a wide range of clients by providing excellence in design and measurable results.

Joe Ciaudelli

Joe Ciaudelli leads Spectrum Affairs, consulting for Sennheiser USA. Upon graduating from Columbia University with an electrical engineering degree, Joe was hired by Sennheiser in 1987. He provided frequency coordination for large multi-channel wireless microphone systems used by Broadway productions, major theme parks, and broadcast networks. He wrote the Turbo-RF software which became the industry standard planning tool of the time. He also wrote the white paper "Large Multi-Channel Wireless Mic Systems", published by industry trade journals, the NAB and InfoComm proceedings, and the *Handbook for Sound Engineers*. Joe has served in various capacities at Sennheiser USA including: Director of Marketing, Director of Market Development & Education, and Director of Advanced Projects & Engineering. Joe also holds patents for the narrow angle hologram technology.

Bob Cordell

Bob Cordell is an electrical engineer who has been deeply involved in audio for over four decades. He began his professional career at Bell Laboratories where he designed linear integrated circuits and fiber optic communications systems. Bob is a prolific designer of amplifiers, audio test equipment, loudspeakers and other audio gear. In 1983 he published a power amplifier design combining vertical power MOSFETs with error correction, achieving distortion levels of less than 0.001% at 20 kHz. He has published articles on power amplifier design in the popular press and in the Journal of the Audio Engineering Society. Bob holds 17 patents, including one for a loudspeaker “Equalized Quasi Sealed System” (EQSS). In 2010 he authored the book *Designing Audio Power Amplifiers*, published by McGraw-Hill. Bob is a member of the JAES Review Board and he maintains an audiophile website at www.cordellaudio.com.

Tom Danley

Tom Danley has been interested in radio, electronics, loudspeakers and “hifi” since childhood. In 1979 he joined Intersonics Inc, a NASA hardware contractor working with acoustic levitation. One of his first inventions was a levitation sound source, which was 100 times more efficient than the St. Clair source that had previously been used.

While at Intersonics, Tom was awarded 17 patents for acoustic and electromagnetic levitation devices, as well as the Servodrive subwoofer; an air cooling system that eliminated power compression, and the Rotary driver used in the Phoenix Cyclone. In addition to his inventions, Tom designed and built major portions of hardware flown on space shuttle flights STS-7 and STS-51a.

In 2005 Tom and Mike Hedden founded Danley Sound Labs. Since starting DSL he has developed:

- The Synergy Horn, an improvement over the Unity Horn that preserves waveshape and radiates as a single source.
- The Tapped Horn, a low frequency horn configuration using both sides of the radiator, allowing a smaller package than a conventional bass horn with similar cutoff.
- The Paraline, an acoustic lens that converts a point source (compression driver) into a “line source” in a minimum physical depth.

- The Shaded Amplitude Horn, a technique which allows the amplitude vs. angle to be adjusted allowing a point source to have equal or less fall-off of SPL vs. distance than a line source.
- The layered combiner which is used in DSL's J4 to combine the output of 64 compression drivers into one horn.

Tom has presented papers at and has been an invited speaker at the Audio Engineering Society, the Acoustical Society of America, and he has presented on both acoustic and electromagnetic levitation at the Jet Propulsion Labs, NASA HQ and was a presenter at the International Committee on Space Research in The Hague on stabilized electromagnetic levitation.

Tom was one of the early TEF-10 users and is thankful that his boss had the insight to send him to a few SynAudCon seminars and to join AES.

Don and Carolyn Davis

Don and Carolyn Davis form a unique husband and wife team working in audio and acoustics. Celebrating their 65th wedding anniversary in 2014; they started their first business in 1951 with the Golden Ear in Lafayette, Indiana selling eclectic high fidelity equipment. Selling their three stores in 1955, they took a three months trip to Europe in order to try out their Porsche on European racetracks. Returning home they served at the Christian Science Church headquarters in Boston until 1958. Returning again to Europe to acquire another Porsche, they observed an inadequate display of sound equipment at the world's fair in Brussels. They returned to the United States to convince the State Department to allow them to take a full representation of the best component high-fidelity equipment made in the United States at that time. Returning again to Europe with two friends, they demonstrated live recording versus playback in the Pavilion Theatre at the Fair. This early travel led to over 22 trips abroad during the ensuing years.

Upon returning from the second trip Don became Paul Klipsch's "President in Charge of Vice" in Hope, Arkansas, where he was able to accomplish the incorporation of Klipsch and Associates. In 1959 Don went to work for Altec Lansing, eventually becoming one of the Vice President's of the company. Don was co-inventor of the Altec one-third of an octave equalizer system and trained their contractors in advanced system design and utilization of equalizers.

The Davises founded Synergetic Audio Concepts in 1973 in response to the growing need in the audio industry for training in the fundamentals of sound reinforcement. Their work in equalization, speech intelligibility, and recording technologies provided the backbone for developing training seminars and workshops. When they turned Syn-Aud-Con over to Pat and Brenda Brown in 1995, Don and Carolyn had been responsible for the education (training yes, but more important, a new mindset which did bespeak education) of more than 10,000 sound contractors, designers, and consultants.

Don and Carolyn have authored four books, *Acoustical Tests and*

Measurements in 1965; *How to Build Speaker Enclosures* in 1968, co-authored with Alex Badmaieff, which sold over 200,000 copies, *Sound System Engineering*, available from 1973 to 1975 in a ring binder, and in 1975 as a bound edition, and in 2013 a fourth edition, and in 2004, *If Bad Sound were Fatal, Audio would be the Leading Cause of Death*, containing non-technical items from over 23 years of Newsletters along with current comments on the excerpts. (this book contains personal experiences with their sports cars such as how Don got an international competition license from the FIA and how Carolyn decided to attend classes at Gunsight which led to a personal friendship with Col. and Mrs. Jeff Cooper the founder of the school). In the process of communicating with the grads of their seminars on audio and acoustics, a quarterly Newsletter was established. Most of the Newsletter was technical in nature, but it also contained the evolving mindset that bonded teachers and grads into something not originally expected—a fraternity of people dedicated to changing an industry.

The Davises spent their professional careers writing and lecturing on sound system engineering. The audio industry has generously recognized their efforts as instrumental in the better sound quality we enjoy today. Don and Carolyn are both Fellows of the Audio Engineering Society and have received many awards in the audio industry including the distinguished award in Sound Design and Technology from USITT. Don has a Gold certificate from the Acoustical Society of America, and is a senior member of the IEEE. Both Don and Carolyn received the 1999 Heyser award, and most recently the Adele De Berri Pioneers of AV award from InfoComm International in 2010.

Don and Carolyn remain active shooters with Don having been a

Provost coach at the famous Gunsight Academy, Inc. in Arizona. This training has enabled Don to train members of his local sheriff's department at the shooting range on their farm. Don and Carolyn continue to work on his "Great American novel" which may never be published but they'll of had the pleasure of writing it. Fellow shooters include Dr. Petronas, Pat and Brenda Brown, and various young friends they have trained over the years. Don and Carolyn are both Life Members of the NRA and strong supporters of the Second Amendment to the Constitution of the United States.

Steve Dove

Native to Oxfordshire, England, much traveled and now screeched to a halt on a beach in Florida, Steve Dove has designed mixing consoles and signal processing in every genre from valves to DSP, and plug-ins since before they were even called that.

On the basis of an early foray with ICs, he was made a wireman (until fired for his atrocious soldering), a designer, and later a director of Alice, a then major manufacturer of broadcast, film, and recording consoles. Concurrently he provided engineering and acoustics expertise to major rock bands, theatres, and studios worldwide.

He is “Minister of Algorithms” for Wheatstone Corporation, a prominent broadcast and signal processing manufacturer. Former clients as a design consultant include Sony Broadcast, Shure Bros., Solid State Logic, Altec Lansing, Clair Bros., Harman/JBL, Crest, and Peavey.

A widely published author with a long list of innovative design techniques and award-winning products to his credit, he is one of a very few to have designed large-format consoles in both analog and digital.

Dipl.-Phys. Stefan Feistel

Stefan Feistel studied physics at the University of Rostock, Germany, and at the Humboldt University, Berlin, Germany, and received a Master’s degree in theoretical physics in 2004. He defended his PhD thesis on computational modeling of sound systems at the RWTH Aachen University in 2013.

In 2000, together with Wolfgang Ahnert, he founded SDA Software Design Ahnert GmbH, which is dedicated to the development of acoustic modeling and measuring software. In 2002 the nonprofit organization ADA Foundation was established to support universities and colleges with software products such as the simulation software EASE and the measurement tool EASERA. To coordinate these activities, the Ahnert Feistel Media Group (AFMG) was established in 2006. Stefan Feistel is a member of the AES, the ASA, and of the DEGA.

Mr. Feistel authored or coauthored more than 60 papers focusing on software projects and the related mathematical, numerical, and experimental background studies. The JAES article on *Methods and Limitations of Line Source Simulation* has been distinguished with the AES Publications Award 2010. Stefan Feistel is the author of the book *Modeling the Radiation of Modern Sound Systems in High Resolution*, and a coauthor of the books *Messtechnik der Akustik*, edited by M. Möser, and *Handbook for Sound Engineers*, edited by Glen Ballou.

Ralph Heinz

As the son of Renkus-Heinz founder and president, Harro Heinz, you could say that Ralph Heinz was born to design loudspeakers. With a background in mechanical engineering and manufacturing prior to joining his father's company, Ralph Heinz has been the principle loudspeaker designer for Renkus-Heinz since 1992. His accomplishments include the development of the company's proprietary and patented “Complex Conic” and “Co Entrant” waveguides and transducers. Most recently, Ralph has turned his attentions towards “steerable” line arrays, and the fundamental principles that enable this new technology in the Renkus-Heinz “Iconyx” series loudspeakers.

Ralph Heinz resides in Norco, CA with his wife and two sons. Along with designing loudspeakers, Ralph enjoys photography and music (preferably live).

David Miles Huber

David Miles Huber is a three times Grammy-nominated producer and musician in the electronic dance and surround-sound genres, whose music has sold over the million mark. His style is energized and balanced out by lush beats and live acoustic instruments that combine to create a “Zen-meets-Tech experience”. His latest music and collaborations can be heard on www.davidmileshuber.com.

DMH received his degree in music technology from Indiana University and the University of Surrey in Guildford, England. His most prominent book *Modern Recording Techniques* is the standard recording industry text worldwide.

Doug Jones

Doug Jones is Professor of Acoustics Emeritus and the founder of the Audio Arts and Acoustics department at Columbia College Chicago. He has worked in acoustics on six continents on projects ranging from radio stations in developing countries to internationally recognized recording studios. He is still waiting for that hot gig in Antarctica!

Doug is the author of *Sound of Worship* (Focal Press 2010) a book which examines the relationship between theology and the acoustics of worship spaces.

He is also a studio musician, engineer and producer with numerous album credits. He performs regularly with a number of bands in the Chicago area, as well as weekly gigs playing Irish folk music.

Doug is responsible for Danley University, the education arm of Danley Sound Labs, as well as product development and technical support. Doug is a member of the AES and the ASA.

S. Benjamin Kanters

Benj has been on the faculty of Columbia College Chicago since 1993 as an Associate Professor and Associate Chair, Department of Audio Arts & Acoustics, where he directs the Audio Design & Production Program. His teaching responsibilities include recording technology, audio theory and hearing physiology.

Prior to Columbia, he spent 20 years in the audio and music industries, including fourteen years as an adjunct professor of audio in the Schools of Music and Communications of Northwestern University. Through the 1970s, he was partner and sound engineer with the Chicago area concert club Amazingrace. During the 1980s, he was partner and chief managing engineer of Studiomedia Recording Company in Evanston.

Since studying hearing physiology during his graduate studies in 2000, he continues to research developments in the field, including hearing disorders and conservation. In 2007, he founded The Hearing Conservation Workshop, visiting colleges and universities to teach hearing awareness to future audio and music industry professionals. More recently, he has been invited to speak to students and professionals in audiology, offering new perspectives in hearing health care.

He has a BS Speech, Northwestern University and a MM Music Technology, Northwestern University. Benj is a member of the

Audio Engineering Society, National Hearing Conservation Association, and the Performing Arts Medicine Association. He is an advisory board member of the Hugh Knowles Hearing Center of Northwestern University and The Foundation for Hearing & Speech Rehabilitation.

Wayne Kirkwood

Wayne Kirkwood began his career in audio at age eight when he was given a tape recorder. The following Christmas he received a Radio Shack “50 in 1” Science Fair™ kit which ended his recording career and sparked a life-long interest in audio circuit design. Mr. Kirkwood is self-taught and has worked in the sound recording and broadcast industries since his youth. He moderates the popular Pro Audio Design Forum (www.proaudiodesignforum.com) which showcases his and others’ circuit designs for use in recording and mastering. He has co-authored several application notes for THAT Corporation and the AES paper “The 48V Phantom Menace Returns” with Rosalfonso Bortoni.

Mr. Kirkwood quips: “I particularly enjoyed working on the Phantom Menace because it gave me the opportunity to identify the various failure modes from stored charges on the input capacitors. I had a lot of fun with that project." Partly as a result of that research, Mr. Kirkwood has since designed a phantom-tolerant active microphone preamp which has no input capacitors.

Steve Lampen

Steve Lampen has worked for Belden for twenty-three years and is currently Multimedia Technology Manager and also Product Line Manager for Entertainment Products. Prior to Belden, Steve had an extensive career in radio broadcast engineering and installation, film production, and electronic distribution. Steve holds an FCC Lifetime General License (formerly a First Class FCC License) and is an SBE Certified Broadcast Radio Engineer. On the data side he is a BICSI Registered Communication Distribution Designer. In 2010, he was named "Educator of the Year" by the National Systems Contractors Association (NSCA), and in 2011 was named "Educator of the Year" by the Society of Broadcast Engineers. His book, *The Audio-Video Cable Installer's Pocket Guide* is published by McGraw-Hill. His blog appears weekly at www.belden.com/blog. He can be reached at steve.lampen@belden.com

Peter Mapp, BSc, MSc, FIOA, FASA, FAES, CPhys, CEng,

MinstP, FinstSCE, MIEE.

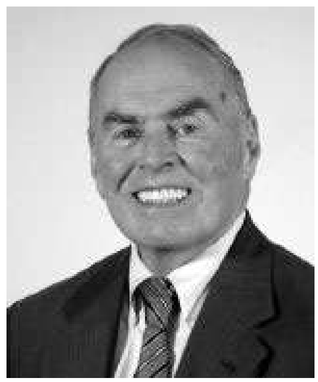
Peter Mapp is principal of Peter Mapp Associates, an acoustic consultancy, based in Colchester, England, that specializes in the fields of room acoustics, electro-acoustics, and sound system design. He holds an honors degree in applied physics and a master's degree in acoustics. He is a Fellow of the Institute of Acoustics, the Acoustical Society of America, and the Audio Engineering Society.

Peter has a special interest in speech intelligibility prediction and measurement and has authored and presented numerous papers and articles on this subject both in Europe and the United States. He currently serves on both British and International Standards committees concerning the sound system design and speech intelligibility.

Peter has been responsible for the design and commissioning of over 500 sound systems, varying from concert halls and theatres, to churches, cathedrals, and other religious buildings, to arenas, stadiums, power stations, and transportation terminals. He is also known for his extensive research work into Distributed Mode Loudspeakers and their application to sound reinforcement.

Peter is a regular contributor to the audio technical press, having written over 100 articles and papers. He is also a contributing author to a number of international reference books including the *Loudspeaker Handbook*. He is currently the chairman of the AES international working group on speech intelligibility and convenor of IEC 60268-16, the standard relating to the measurement of STI and Speech Intelligibility.

Hardy Martin



Musical talent came naturally for Hardy Martin and guitar was his instrument of choice playing with many bands from country to rock and roll. He is a lifetime member of the American Federation of Musicians.

With a world of knowledge learned from experience with audio products, Hardy formed a company and began to design and manufacture mixing consoles for recording studios, radio and TV production, and showrooms from 1965–1978.

Innovative Electronic Designs, LLC (IED) followed in 1978, and Hardy being one of the founders, was named president of the company and remained until 2008.

Through the following years, IED has been credited for many firsts in the audio industry: 1981–Digital Record and Playback System; 1982–Computer Controlled Paging System; 1984–

Computer Controlled Stadium/Arena and Convention Center System; 1985–Computer Controlled Automatic Mixer; 1986–Modular Class D Power Amplifier; 1987–Multi-Channel Software Controlled Ambient Analysis System; 1990–Automated Flight Announcement System (FAS); 1990-DSP Sound System; 1995-Integrated Transit Station Audio and Visual Controller; 1996–Ethernet Network Audio/Visual Paging System; 1997–Courtesy Announcement System; 2003–Integrated Airport Paging and MUFIDS/GIDS System; 2004–Network (CobraNet) based Microphone Station with Dynamic Routing; 2006–Networked DSP/Modular Power Amplifier System; 2007–Paging System that was 100% digital and networked from Paging Station to Power Amplifier; 2010–Virtual Announcement Control System.

IED's Announcement Control System can be found in 200+ airports around the world (75% of the top 100 U.S. airports). Mass Transit Systems are installed in the New York City Transit, Chicago Transit Authority, Los Angeles Metro Red Line, and many more. Government installations can be found at NASA and AFB and many similar venues. Industrial systems include automotive, pharmaceutical and healthcare facilities.

Hardy is the co-author of many patents in audio/visual related topics. He is currently Chief Technology Officer (CTO) at IED. He is a past member of the AES, and current member of the National Fire Protection Association, NFPA.

Hardy enjoys life with many hobbies from sports cars, airplanes to senior sports events. He holds a commercial pilot's certificate with both instrument and multi-engine ratings. He is currently a player on the Louisville Thunder Senior Softball Team and plays in tournaments throughout the United States. He is a proud Eagle

Scout and supports the Boy Scout Organization of America.

Steven McManus

Steven McManus graduated from the University of Edinburgh in 1990 with a B. E. in Electrical and Electronic Engineering. During this time he also worked on a computer-aided navigation and charting project through the University of Dundee.

An early interest in sound recording expanded into a production and installation company that he ran for 8 years. Steve moved to the United States in 1999 after returning to Herriot-Watt University for an update in programming skills.

He was with the TEF division of Gold Line for 6 years working on both software and hardware. He is currently working for Teledyne Benthos in Massachusetts with underwater acoustic communication and navigation systems.

Ken C. Pohlmann

Ken C. Pohlmann has worked extensively in the research, development, and testing of new audio technology. He serves as a consultant in the development and sound quality evaluation of digital audio products and sound systems for audio and automobile manufacturers. He also serves as a consultant and expert witness in patent infringement litigation. Some of his consulting clients have included: Alpine Electronics, Analog Devices, Apple Computer, Bertlesmann Music Group, Blockbuster Entertainment, BMW, Canadian Broadcasting Corporation, Cirrus Logic, Daimler-Chrysler, Eclipse, Ford, Fujitsu Ten, Harman International, Hewlett-Packard, Hughes Electronics, Hyundai, IBM, Kia Motors, Lexus, Lucent Technologies, Microsoft, Mitsubishi Electronics, Motorola, Nippon Columbia, Onkyo America, Philips, Real-Networks, Recording Industry Association of America, Samsung, Sensormatic, Sonopress, Sony, SRS Labs, TDK, Time Warner, Toyota, and United Technologies.

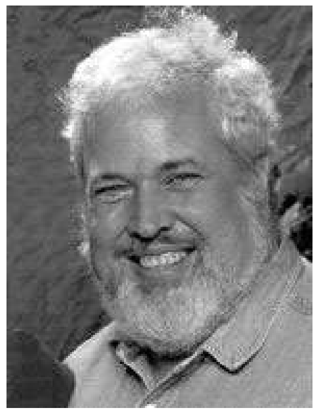
Ken is a professor emeritus at the University of Miami in Coral Gables, Florida where he served as a tenured full professor and director of the Music Engineering Technology program in the Frost School of Music. He initiated new undergraduate and graduate courses in digital audio, advanced digital audio, internet audio, acoustics and psychoacoustics, and studio production. In 1986, he founded the first master's degree program in Music Engineering Technology in the United States. Mr. Pohlmann holds Bachelor of Science and Master of Science degrees in Electrical Engineering from the University of Illinois in Urbana–Champaign.

He is the author of *Principles of Digital Audio* (McGraw-Hill); this book has appeared in six editions and has been translated into Dutch, Spanish, and Chinese. He is also author of *The Compact Disc Handbook* (A-R Editions); this book has appeared in two editions and has been translated into German. He is co-author of the sixth edition of *The Master Handbook of Acoustics* (McGraw-Hill) and the second edition of *Sound Studio Construction* (McGraw-Hill). He is also co-author of *Writing for New Media* (John Wiley & Sons), and editor and co-author of *Advanced Digital Audio* (Howard W. Sams). Since 1982, he has written over 3000 articles for audio publications and is columnist and blogger for Sound & Vision magazine.

Mr. Pohlmann chaired the Audio Engineering Society's International Conference on Digital Audio in Toronto in 1989 and co-chaired the Society's International Conference on Internet Audio in Seattle in 1997. He was presented AES Board of Governor's Awards in 1989 and 1998 and was named an AES Fellow in 1990 for his work as an educator and author in the field of audio engineering. He was elected to the AES Board of Governors in 1991. He served as AES convention papers chairman in 1984 and papers co-chairman in 1993. He was elected as the AES Vice President of the Eastern U.S. and Canada Region in 1993. He was presented the University of Miami Philip Frost Award for Excellence in Teaching and Scholarship in 1992. He served as a Non-Board Member of the National Public Radio Distribution/Interconnection Committee

(2000–2003). He served on the Board of Directors of the New World Symphony (2000–2005). He served on the Advisory Board of SRS Labs as the charter member (2009–2012).

Ray A. Rayburn



Ray A. Rayburn is Principal Consultant with Sound First, LLC. He is a Life Fellow of the Audio Engineering Society, and an emeritus member the Acoustical Society of America. He is Chairman of the AES Standards Subcommittee on Interconnections, and the working group on Audio Connectors. He is a member and former Chairman of the AES Technical Committee on Signal Processing and is a member of the AES Standards working group on Microphone Measurement and Characteristics. He was one of the developers of the ANSI/InfoComm Standard on Audio Coverage Uniformity, and currently leads the InfoComm Standards working group on Spectral Balance.

Ray has an ASET from New York Institute of Technology and has worked in the fields of audio, electroacoustics and telecommunications for over 43 years. He has taught audio at Eastman School of Music, Institute of Audio Research, and Info-Comm. He designed recording facilities for Broadway Video, Eurosound, Photo-Magnetic Sound, db Studios, and *Saturday Night Live*. His recording credits range from the Philadelphia Orchestra to Frank Zappa. Equipment he has designed includes film dubbers, tape duplicators, specialty telecommunication equipment for the stock brokerage industry, and a polymer analyzer. He has been a consultant on sound systems for the U.S. Senate, U.S. House, Wyoming Senate and House, Georgia Dome, Texas House of Representatives, and University of Hawaii Special Events Arena, among many others.

Dr. Craig Richardson

Dr. Craig Richardson is a vice president and general manager of Polycom's Installed Voice Business. In this role, Richardson leads the development of installed voice products and recently introduced the Sound Structure audio conferencing products—the first installed conferencing products that provide both mono and stereo echo cancellation capabilities for an immersive conferencing experience.

Prior to joining Polycom, Craig was president and CEO of ASPI Digital and focused the company on creating the EchoFree™ teleconferencing products for the audio/video integrator marketplace. ASPI's products allowed users to experience full-duplex communication in situations never before thought possible, such as distance learning, telemedicine and courtroom applications. In 2001, ASPI Digital was acquired by Polycom Inc., the leader in unified collaborative communication.

Dr. Richardson led algorithm development at ASPI Digital as the director of Algorithm Development working on an advanced low-bit-rate video coder, the MELP military standard voice coder, digital filter design products, computer telephony products, multimedia algorithms for speech, audio, and image processing applications for TMS320 family of digital signal processors. He has written numerous papers and book chapters, has been granted four patents, and is the co-author (with Thomas P. Barnwell, III, and Kambiz Nayebi) of *Speech Coding A Computer Laboratory Textbook*, a title in the Georgia Tech Digital Signal Processing Laboratory Series published by John Wiley & Sons.

He is a member of Tau Beta Pi, Sigma Xi, and a senior member of the IEEE. He was graduated from Brown University with a bachelor's degree in electrical engineering and received both his master's and doctoral degrees in electrical engineering from the Georgia Institute of Technology.

Gino Sigismondi

Gino Sigismondi, a Chicago native and Shure Associate since 1997, has been active in the music and audio industry for nearly 15 years. Currently managing the Technical Training division, Gino brings his years of practical experience in professional audio to the product training seminars he conducts for Shure customers, dealers, distribution centers, and internal staff. Gino spent over 9 years as a member of Applications Engineering, assisting Shure customers with choosing and using the company's vast array of products, and is the author of the Shure educational publications *Selection and Operation of Personal Monitors, Audio Systems Guide for Music Educators*, and *Selection and Operation of Audio Signal Processors*. He was recently awarded status as an Adjunct Instructor by the InfoComm Academy.

Gino earned his B.S. degree in Music Business from Elmhurst College, where he was a member of the jazz band as both guitar player and sound technician. After college, he spent several years working as a live sound engineer for Chicago-area sound companies, night clubs, and several local acts. Gino continues to remain active as a musician and sound engineer, consulting musicians on transitioning to in-ear monitors and expanding his horizons beyond live music to include sound design for modern dance and church sound.

Jeff D. Szymanski

Jeff D. Szymanski is an Acoustical Engineer with Black & Veatch Corporation, a leading global engineering, consulting, and construction company specializing in infrastructure development in energy, water, information, and government markets. Jeff's experience covers many areas of acoustics, including architectural acoustics design, industrial noise and vibration control, environmental noise control, and A/V systems design. He has over 19 years of experience in the acoustical manufacturing and consulting industries. Jeff has written and presented extensively on the subjects of architectural acoustics and noise control. He is a full member of the Acoustical Society of America and a Board Certified Member of the Institute of Noise Control Engineering. Jeff holds two US patents for acoustical treatment products, is a licensed professional engineer, and spends his free time jamming in the fantastic cover band "The Heather Stones."

Hans-Peter Tennhardt

Hans-Peter Tennhardt was born in Annaberg/Erzgebirge, Germany in 1942. From 1962–1968, Mr. Tennhardt studied at the Technical University Dresden, Department of Electrotechnics—Low-current Engineering. His field of study was Electroacoustics with Professor Reichardt.

Mr. Tennhardt graduated in 1968 as a Diploma'd Engineer for Electrotechnics—Low-current Engineering at the TU Dresden, with an extension of the basic study at the Academy of Music Dresden. His Diploma thesis was on the subject "Model Investigations in Town Planning." From 1968–1991, Mr. Tennhardt was a Scientific Collaborator in the Department Building and Room Acoustics at the Building Academy in Berlin, with Professor Fasold. He became Deputy Head of the Department Building and Room Acoustics of the Institute for Heating, Ventilation, and Fundamentals of Structural Engineering at the Building Academy in 1991. In 1992 Mr. Tennhardt became Group Leader for room acoustics at the Institute of Building Physics (IBP) of the Frauenhofer Institute Stuttgart, Berlin Branch.

Since then he has been Head of the Department Building Physics and of the Special Section Building and Room Acoustics at the Institute for Maintenance and Modernization of Buildings (IEMB) of the Technical University Berlin.

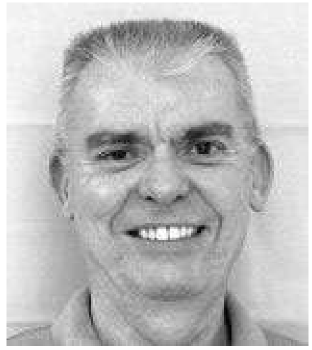
Leslie B. Tyler

Leslie “Les” Tyler is president of THAT Corporation, which provides high-performance audio technology in the form of integrated circuits to the professional audio industry, and licensed intellectual property to the consumer audio industry. Mr. Tyler co-founded the company in April of 1989 along with other senior managers and engineers from dbx, Inc.

Immediately prior to founding THAT, Mr. Tyler spent 10 years at various engineering and engineering management positions at dbx, including vice president for engineering and vice president for technology. Prior to dbx, he spent two years as the chief recording engineer at Pyramid Sound in Ithaca NY.

Mr. Tyler holds three U.S. patents. He attended Cornell University where he obtained a B.S. in electrical engineering. He plays several instruments, especially acoustic jazz piano.

Bill Whitlock



Bill Whitlock was born in 1944 and was building vacuum-tube electronics at the age of 8 and operating his own radio repair business at the age of 10. He grew up in Florida, attended St. Petersburg Junior College, and graduated with honors from Pinellas County Technical Institute in 1965. He held various engineering positions with EMR/Schlumberger Telemetry, General Electric Neutron Devices, and RCA Missile Test Project (on the USNS Vandenberg, now an artificial reef off Key West) before moving to California in 1971.

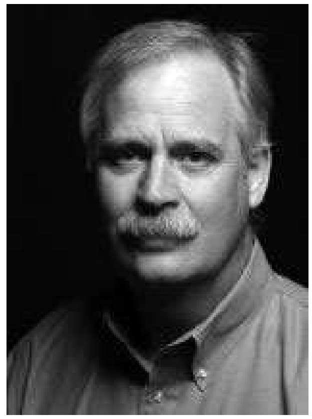
His professional audio career began in 1972 when he was interviewed by Deane Jensen and hired as chief engineer by custom console maker Quad-Eight. There he developed Compumix®, a pioneering console automation system, and other innovations. From 1974 to 1981, he designed automated image synthesis and control systems, several theater sound systems, and patented a multichannel PCM audio recording system for producers of the Laserium® laser light show. In 1981, Bill became manager of Electronic Development Engineering for Capitol Records/EMI where he designed high-speed cassette duplicator electronics and other specialized audio equipment. He left Capitol in 1988 to team with friend and colleague Deane Jensen, developing hardware for Spatializer Audio Labs and other clients. After Deane's tragic death in 1989, Bill became President and Chief Engineer of Jensen

Transformers, a position he held until his semi-retirement in 2014.

His landmark paper on balanced interfaces was published in the June 1995 AES Journal. He is an active member and former chairman of the AES Standards Committee Working Group that produced AES48-2005. Over the years, Bill has presented many tutorial seminars and master classes for the AES as well as presentations to local AES chapters around the world. He suggested major changes to CMRR measurement methods for IEC standards, which the IEC adopted in 2000. He has written numerous magazine articles and columns for Mix, EDN, S&VC, System Contractor News, Live Sound, Multi-Media Manufacturer, and others. Since 1994, he has taught myth-busting seminars on grounding and interfacing to tens of thousands at industry trade shows and invited lectures at SynAudCon workshops, private companies, and universities, including MIT.

Bill is a Life Fellow of the Audio Engineering Society and a Life Senior Member of the Institute of Electrical and Electronic Engineers. His patents include a bootstrapped balanced input stage, available as the InGenius® IC from THAT Corporation, and a mechanically and electrically compatible balanced version of the ubiquitous RCA connector. Bill currently does system troubleshooting and analog circuit design work as Whitlock Consulting. He is married and lives in Oxnard, CA. In his leisure time, Bill enjoys travel, hiking, music, and restoring vintage radio and TV receivers.

Jack Wrightson



Jack Wrightson is one of the founding partners of Wrightson, Johnson, Haddon and Williams (WJHW), a multi-discipline consulting design firm based in Dallas, Texas, where he oversees all aspects of the company's operations. Following graduate school in 1982, Jack began his career with Joiner, Pelton Rose; later known as the Joiner Rose Group where he was involved in a variety of acoustical and electroacoustical sports projects. In 1990, Jack along with three colleagues founded WJHW to provide A/V and acoustical consulting services to both sports and non-sports clients. Today, Jack continues to advise sports clients worldwide on audio and technology enhancements for stadiums and arenas.

Jack earned a BA degree in Biopsychology from Rutgers University, a Masters in Psychoacoustics from the University of Wisconsin—Milwaukee and a Masters of Business Administration from Southern Methodist University.

Dr. Peter Xinya Zhang

Dr. Peter Xinya Zhang is a tenured associate professor at Columbia College Chicago. His research interests include psychoacoustics (especially binaural hearing, spatial hearing, 3D sound simulation, and virtual reality), auditory physiology, singing voice, acoustics and its applications in media arts. He received his degree of Bachelor of Science in physics from Peking University, P. R. China, and received his Ph.D. in physics from Michigan State University, U.S.A. In his doctoral thesis, he investigated human binaural pitch effects to test various binaural models, and developed a new technique to simulate 3D sound field for virtual reality with loudspeakers.

Dr. Zhang served as president of the Chicago Chapter, Acoustical Society of America from 2010 to 2013. He is a member of the Acoustical Society of America, and of the Audio Engineering Society. Dr. Zhang has published papers at journals on acoustics and hearing, such as the Journal of Acoustical Society of America and Hearing Research. He has been the author of the chapter on psychoacoustics in this handbook since the 4th edition. Dr. Zhang has presented at various conferences, and served as co-chair in the session of psychological and physiological acoustics at the 4th joint conference of Acoustical Society of America and Acoustical Society of Japan. He has lectured on psychoacoustics and hearing at various institutions including Columbia College Chicago, Loyola University, Peking University, and Institute of Acoustics at Chinese Academy of Sciences.

Part 1

Introduction to Sound and Acoustics

Chapter 1

Audio and Acoustic DNA—Past and Present

by Don and Carolyn Davis

1.1 Introduction

This chapter is the DNA of my ancestors, the giants who inspired and influenced our lives. If you or a hundred other people wrote this chapter, your ancestors would be different. We hope you find reading the DNA of our ancestors worthwhile and that it will provoke you into learning more about them.

Interes t in my audio and acoustic ancestors came about by Carolyn and I starting the first independent Hi-Fi shop, The Golden Ear, in Lafayette, Indiana in early 1952. The audio enthusiasts from Purdue came to our shop to meet the great men of hi-fi; Paul Klipsch, Frank McIntosh, Gordon Gow, H.H. Scott, Saul Marantz, Rudy Bozak, Avery Fisher—manufacturers who exhibited in the Hi-Fi shows at the Hollywood Roosevelt and the Hilton in New York City. We sold our shops in Indianapolis and Lafayette in 1955, and took an extended trip to Europe. In 1958 I went to work for Paul Klipsch as his “President in Charge of Vice.” Mr. Klipsch introduced me to the writings of Lord Kelvin, the Bell Labs West Street personnel, as well as to his untrammeled genius.

Don and Carolyn Davis and "The Golden Ear"

Altec was the next stop, with my immediate manager being "the man who made the motion picture talk." At Altec I rubbed against and was rubbed against by the greats and those who knew the Greats and those who knew the Greats from the inception of the Art. This resulted in our awareness of the rich sense of history we have been a part of and we hope that sharing our remembrance will help you become alert to the richness of your own present era.

In 1972 Carolyn and I were privileged to work with the leaders in our industry who came forward to support the first independent attempt at audio education, Synergetic Audio Concepts (SynAudCon). These manufacturers represented the best of their era and they shared freely with us and our students without ever trying to "put strings on us."

Thoughts for Today. I want to discuss many of my contemporaries in this new section because in conversations with newcomers to the field I have shared several of those stories with them and they have found them inspiring. Each of these individuals led my thinking in a new direction and often, as indicated, were a direct assistance to the task at hand. They all possessed traits of character, that made them valuable members of society. Carolyn and I have always sought alertness to talented individuals in any field. They are fun to know, a challenge to keep up with, and a guide to a fuller understanding of viewpoints not formerly considered.

Few of us ever chose audio and acoustics as a career but rather those subjects chose us as we were led by our innate experiences (love of music, ham radio, etc.) into appreciation of the how, what, and why of the technology behind the toys that gave us pleasure in our early lives.

1.2 Gestalt Theory

While men and women attracted to acoustics are often listeners to classical music and appreciators of visual art, their counterparts in audio are often more fastened on the scientific aspects of the material components used in their work.

Ernst Mach wrote, *We do not perceive the world in itself, if we did we would perceive chaos. Thus we have evolved senses that perceive contrasts of perception, relations of perception. Sensations by themselves can have no organic meaning; only the relations of sensations to one another can have meaning*.

Perception, Mach believed, is never perception of direct stimuli. Sensations are not simply raw experiences, but the interaction of experience with a preformed cognitive structure. For instance, when we hear a known melody, we recognize it no matter what key it is played in. It can be hummed, buzzed, or strummed on a guitar. Furthermore, even if one or more notes are incorrect, we still recognize it.

Ernst Mack

Mach asks, *What constitutes a melody?* It seems incorrect to say that the actual sound vibrations constitute the melody as we have just seen that numerous different sounds can make the same melody. But on the other hand, it seems empirically odd to say that a melody is not constituted out of its sounds. The actual melody, then, exists in our ability to recognize it. It is formed by experience of one or more examples of the melody, but it is an idealization of that experience. Significantly, the idealization captures not the actual sounds, but the relationships of the sounds to one another. For Mach, this process is at the basis of all perception.

Experience requires an “a priori,” but that “a priori” is itself formed by experience. Men and women who have, through experience, learned the deceptive nature of the five physical senses, learn more readily the nuances of acoustic measurements as well as the optimum number and placement of microphones in recording situations. It is the inept in recording work that resorts to close miking and asks “where shall I place the microphone” for a measurement.

They have ears that receive signals but the “listening capability” between them is relatively unprogrammed. There is a great deal of evidence that the current generation in audio have been exposed to highly undesirable “a priori” both in program material and cultural background.

In all professional work one should seek out the real standards that apply which are not necessarily those that are currently in use by the majority. To measure anything meaningfully requires:

1. Experience with similar devices.
2. Mathematical analysis of the device and its most likely performance.
3. Cut and try experimentation.

Obviously the mathematical approach is often the quickest. Experience with similar devices occurs when someone knowledgeable guides you through a process they are already very familiar with. Cut-and-try can overtime, hopefully with a minimum amount of destruction, lead to experience. In the long run, all measurement runs up against a trained listener's perception. Yes! I know any untrained listener can be satisfied with trash, but any professional's goal should be acceptance by the trained listener. When that occurs you have gained membership in a privileged peer group.

Ernst Mach 1838–1916, a contemporary of Gustav Fechner the psycho-acoustician, was acknowledged by Einstein as being the philosophical forerunner of relativity. Early in his life Mach defended Christian Doppler against two prominent physicists, Petzval and Angstrom who had challenged the "Doppler effect" by building an apparatus that consisted of a six foot tube with a whistle at one end that rotated in a vertical plane. When the listener stood in the plane of the axis of rotation no changes in pitch could be heard. But if the observer stood in the plane of rotation, fluctuations in pitch that corresponded to the speed of rotation could be heard.

Mach discovered that the eye has a mind of its own; we perceive not direct stimuli but relations of stimuli. The visual system operates through a process of continual adaptation of the present sensation to previous ones. We do not experience reality but rather experience the after effects of our nervous systems adaptations to new stimuli. Our cognitive structure is itself formed through previous experience, and our current experience is structured by it in turn. Mach claimed "We do not perceive the world in itself, if we did we would perceive chaos." From Gustav Fechner's work.

1.3 Music in Acoustics

The following is from Frederick V. Hunt, *The Origin of Acoustics*.

> *The broad implications of what Sarton called "The cause and cure of Scholasticism" may seem to be strange grist for the acoustic mill. There were at least two reasons, however, why the branch of acoustics dealing with music was able to make a unique contribution toward the ultimate conquest of scholasticism. The first was that music was insured a firm continuing hold on its place in the scholastic sun by virtue of its role as a part of the classical quadrivium (Astronomy, Geometry, Mathematics, Music.) Educators, philosophers, encyclopedist, and commentators alike had pre-force to deal with music and with the evolution of musical science. The second basis for the close relation between music and scholasticism stemmed from the fact that music is, sui generis, an epitome of experimental science. Objective in execution and humanistic in appreciation, its three aspects of composition, performance, and appreciation exemplified—and held up continuously for conscious or unconscious regard—the*

scientific credo of hypothesis, experiment, and conclusion.

> *Almost every medieval writer who considered the theory of music felt obliged to devote at least one section of his treatise to the production of sound and to the factors that influence pitch. Since the Greeks had already achieved an understanding of the physical nature of sound that represented almost as much sophistication as could be sustained without explicit formulation of the laws of dynamics and of the physical nature of the sound medium, it was left for the Moslem scientists of the time to sort out many of the false ideas expressed along with the true by the Greeks; that clarified if not augmented the Greek acoustical tradition and was thus conserved in Islam for retransmission—"on the wings of a song," as it were—to the West.*

I was studying the above when I was asked by a scholar of medieval history, "What was there on acoustics in the 1100s?" Just such exchanges were what jump started the Renaissance. The synergy generated by having literature arrive on the scene for which the readers had no "a priori" conditioning prods the human consciousness to a higher level of brain activity than that normal to familiar environments.

It was the Christian church in the period surrounding the Crusades that began the idea of Universities for the education of the chosen by means of the quadrivium (defined above) and trivium (grammar, logic, rhetoric) and the catalyst provided by the Hindu – Moslem number system, (where the missing number in the Roman numerals, Zero) led thinkers into more rational systems of thinking. The study of Western thought from the Crusades to the early 1900s is replete with acquired "a priori" from unexpected sources. I would suggest that a fruitful area of investigation for a doctorate candidate would be why this path was so one-way to the benefit of the West in so far as the physical sciences were concerned. The secular West when compared to the more religiously based East has advanced materially in science at the expense of spiritual values. The East has demonstrably fallen behind the material needs of their people. Perhaps both sides would benefit from a deep reevaluation of what the real goals are for a truly educated individual.

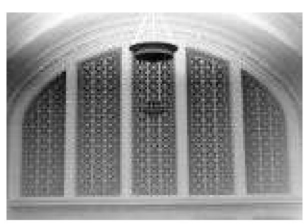
I have, since high school, preferred apprenticing myself to talented, thinking individuals willing to point out my shortcomings and provide me with tasks that advanced me. This allowed me to follow my enthusiasms which in turn provided the motivation to master subjects I otherwise would have avoided in the normal course of academic work.

I was very definitely a student that benefited from waiting until I was ready before I approached difficult subjects. The world as a whole seemingly can't allow masses that luxury. The best students I observed when I attended Purdue were returning veterans from World War II, empowered by the G. I. Bill to attend the university of their choice. The students were truly motivated, street-smart in an international sense, and knew how to work hard. Subsequently it is has been my opinion that attending and finishing a high-quality high school should be followed by multiple years of either military service or its equivalent involving manual labor in civilian life before being allowed to take university level classes. Manual labor provides clarity of thought regarding a career.

1.4 Sales or Communication?

Many times excellent technical solutions fail to achieve execution due to the lack of ability on the part of the engineer to communicate with the client. One such case was a well-designed church, from the standpoint of acoustics, that gave remarkable support to an extremely talented soloist and the pipe organ but needed assistance in terms of speech intelligibility for anyone speaking from the platform.

A prominent architectural feature was the ornate grill in front of the organ pipes. When the engineer proposed installing a large tencell multicellular horn in the center of the organ grill, a gasp of horror escaped the lips of the building committee. Being a resourceful communicator, the engineer suggested making a cardboard model of the front of the horn, painting it the same color as the grill, and installing it for one Sunday service on the front of the grill to see how the congregation would react to its presence. Interestingly, no one in the audience indicated that they had noticed the horn and when questioned, went back to look at it. The building committee immediately approved the installation of the horn.



The pictures show a close view of the grill with the real horn installed and a second view from the balcony area. The result was unsullied architecture, superb musical qualities, and acceptable speech intelligibility. The installed system containing theater quality electronics as well as theater quality loudspeakers had a life expectancy of at least 25 years.

1.5 Motional Impedance

From Frederick V. Hunt, *Electroacoustics*, pages 96-102.

> *The term Motional Impedance was first introduced by A. E. Kennelly and G. W. Pierce in 1912 when they were studying the variation of impedance with frequency for a telephone receiver. They discovered that the electrical impedance could be influenced by the motion of the coupled mechanical system. The circumstances of this discovery are not without interest. In making their measurements, one of the experimenters would balance the impedance bridge while the other tended the signal source, a not always reliable Vreeland oscillator that was located in a nearby room. Quite by chance, one of them would habitually lay the receiver on its side on the laboratory bench while adjusting the bridge. The other always turned it facedown, thereby altering the acoustic loading on the diaphragm, its motion, and the electrical impedance! It is easy to understand how alarmed they were when their measurements showed a complete lack of agreement at all frequencies in the neighborhood of resonance. In the course of pursuing the source of this discrepancy, they finally decided to*

abandon the careful nursing of the oscillator and watch each other balance the bridge, whereupon the difference in their procedures became apparent at once. Kennelly and Pierce both appreciated the physical significance of the effect immediately, and each succeeded in working out a substantially complete theoretical analysis of the phenomenon within a few hours.

The motional modifications of the mechanical and the electric impedances are seen to be proportional to the negative product of the two transduction coefficients. It follows that the magnitude and nature of the motional impedances will depend on the size of these coefficients and on whether they are real or complex.

Interestingly Dr. Hunt employed Nyquist-type vector plots to show all of the components including the admittance which is probably why Richard Heyser suggested using the admittance values in his last papers on the measurement of impedance. Hunt's book contains all the equations relative to impedance including the fleas on the back of fleas.

1.6 Genesis

The true history of audio consists of ideas, men who envisioned the ideas, and those rare products that represented the highest embodiment of those ideas. The men and women who first articulated new ideas are regarded as discoverers. Buckminster Fuller felt that the terms *realization* and *realizer* were more accurate.

Isaac Newton is credited with “We stand on the shoulders of giants” regarding the advancement of human thought. The word science was first coined in 1836 by Reverend William Hewell, the Master of Trinity College, Cambridge. He felt the term, natural philosopher, was too broad, and that physical science deserved a separate term. The interesting meaning of this word along with entrepreneur-tinkerer allows one a meaningful way to divide the pioneers whose work, stone by stone, built the edifice we call audio and acoustics.

Mathematics, once understood, is the simplest way to fully explore complex ideas but the tinkerer often was the one who found the “idea” first. In my youth I was aware of events such as Edwin Armstrong’s construction of the entire FM transmitting and reception system on breadboard circuits. A successful demonstration then occurred followed by detailed mathematical analysis by the same men who earlier had used mathematics to prove its impossibility. In fact, one of the mathematician’s papers on the impossibility of FM was directly followed at the same meeting by a working demonstration of an FM broadcast by Armstrong.

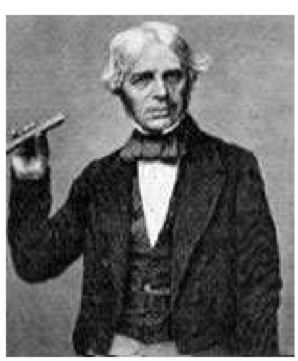
The other side of the coin is best illustrated by James Clerk Maxwell (1831–1879), working from the non-mathematical seminal work of Michael Faraday.

Michael Faraday had a brilliant mind that worked without the encumbrance of a formal education. His experiments were with an early Volta cell, given him by Volta when he traveled to Italy with Sir Humphry Davy as Davy’s assistant. This led to his experiments with the electric field and compasses. Faraday envisioned fields of force around wires where others saw some kind of electric fluid flowing through wires. Faraday was the first to use the terms electrolyte, anode, cathode, and ion. His examination of inductance led to the electric motor. His observations steered his good friend, James Clerk Maxwell, to his remarkable equations that defined electromagnetism for all time.

A conversation with William Thomson (later Lord Kelvin) when Thomson was 21 led Faraday to a series of experiments that showed that Thomson's question as to whether light was affected by passing through an electrolyte…it wasn't, led to Faraday's trying to pass polarized light past a powerful magnet to discover the magneto-optical effect (the Faraday effect). Diamagnetism demonstrated that magnetism was a property of all matter.

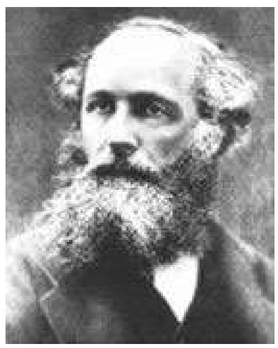
Faraday was the perfect example of not knowing mathematics freed him from the prejudices of the day.

James Clerk Maxwell was a youthful friend of Faraday and a mathematical genius on a level with Newton. Maxwell took Faraday's theories of electricity and magnetic lines of force into a mathematical formulation. He showed that an oscillating electric charge produces an electromagnetic field. The four partial differential equations were first published in 1873 and have since been thought of as the greatest achievement of the 19th century of physics.



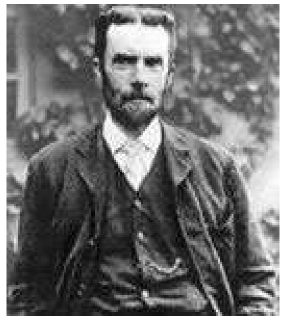
Michael Faraday

At the time of Maxwell's death in 1879, his theory of electromagnetism was only one among several and by no means the clear leader. By 1890 it had swept its rivals from the field and was taking its place as one of the most successful and fundamentals theories in all of physics. G. F. Fitzgerald, (1851–1901), Oliver Lodge (1851–1940), and Oliver Heaviside (1850–1925) recast the theory, confirmed it experimentally, and extended it in directions that Maxwell himself would have scarcely anticipated.



James Clerk Maxwell

Oliver Heaviside hit on the energy flux theorem independently a few months after J.H. Poynting and saw it as central to Maxwell's theory. Heaviside was inspired to recast the long list of equations Maxwell had given in his Treatise into the compact set of four vector equations now universally known as "Maxwell's equations." These equations led to the energy flux formula ($S = E \times H$) in a simple and direct way that clarified many other aspects of Maxwell's theory.



Oliver Heavyside

Maxwell's equations are the perfect example of mathematics predicting a phenomenon that was unknown at that time. That two such differing mind-sets as Faraday and Maxwell were close friends bespeaks the largeness of both men.

These equations brought the realization that, because charges can oscillate with any frequency, visible light itself would form only a small part of the entire spectrum of possible electromagnetic radiation. Maxwell's equations predicted transmittable radiation which led Hertz to build apparatus to demonstrate electromagnetic transmission.

J. Willard Gibbs, America's greatest contributor to electromagnetic theory, so impressed Maxwell with his papers on thermodynamics that Maxwell constructed a three-dimensional model of Gibbs's thermodynamic surface and, shortly before his death, sent the model to Gibbs.

G.S. Ohm, Alessandro Volta, Michael Faraday, Joseph Henry, Andre Marie Ampere, and G. R. Kirchhoff grace every circuit analysis done today as resistance in ohms, potential difference in volts, current in amperes, inductance in henrys, and capacity in farads and viewed as a Kirchhoff diagram. Their predecessors and contemporaries such as Joule (work, energy, heat), Charles A. Coulomb (electric charge), Isaac Newton (force), Hertz (frequency), Watt (power), Weber (magnetic flux), Tesla (magnetic flux density), and Siemens (conductance) are immortalized as international S.I. derived units. Lord Kelvin and Ampere have their names inscribed as an S.I. base unit.

As all of this worked its way into the organized thinking of humankind, the most important innovations were the technical societies formed around the time of Newton where ideas could be heard by a large receptive audience. Some of the world's best mathematicians struggled to quantify sound in air, in enclosures, and in all manner of confining pathways. Since the time of Euler (1707–1783), LaGrange (1736–1813), and d'Alembert (1717–1783), mathematical tools existed to analyze wave motion and develop field theory.

By the birth of the 20th century, workers in the telephone industry comprised the most talented mathematicians and experimenters. Oliver Heaviside's operational calculus had been superseded by LaPlace transforms at MIT (giving them an enviable technical lead in education).

Oliver Heaviside wrote in his book, *Electromagnetic Theory*, volume III page 519:

> *As the universe is boundless one-way, towards the great, so it is equally boundless the other way, towards the small; and important events may arise from what is going on in the inside of atoms, and again, in the inside of electrons. There is no energetic difficulty. Large amounts of energy can be very condensed by reason of great forces at small distances. How electrons are made has not yet been discovered. From the atom to the electron is a great step, but it is not finality.*

> *Living matter is sometimes, perhaps generally, left out of consideration when asserting the well known proposition that the course of events in the physical world is determined by its present state, and by the laws followed. But I do not see how living matter can be fairly left out. For we do not know where life begins, if it has a beginning. There may be and probably is no ultimate distinction between the living and the dead.*

Heaviside in these few words revealed his understanding of Markov processes, the potential energy that can be confined in extremely short distances, and that what we call life, which could be the defining principle of the universe, needs our consideration. Heaviside would've been quite comfortable with John Wheeler and the universe as "*Its from Bits*."

1.7 1893—The Magic Year

At the April 18, 1893 meeting of the American Institute of Electrical Engineers (AIEE) in New York City, Arthur Edwin Kennelly (1861–1939) gave a paper entitled "Impedance."

The truly extraordinary Arthur Edwin Kennelly left school at the age of thirteen and taught himself physics while working as a telegrapher. He is said to "have planned and used his time with great efficiency," which is evidenced by his becoming a member of the faculty at Harvard in 1902 while also holding a joint appointment at MIT from 1913–1924. He was the author of ten books and the co-author of eighteen more, as well as writing more than 350 technical papers.

Arthur Edwin Kennelly

Edison employed A.E. Kennelly to provide physics and mathematics to Edison's intuition and cut-and-try experimentation. His classic AIEE paper on impedance in 1893 is without parallel. The reflecting ionosphere theory is jointly credited to Kennelly and Heaviside and known as the Kennelly-Heaviside layer. One of Kennelly's Ph.D. students was Vannevar Bush, who ran American's WWII scientific endeavors.

In 1893 Kennelly proposed impedance for what had been called apparent resistance, and Steinmetz suggested reactance to replace inductance speed and wattless resistance. In the 1890 paper, Kennelly proposed the name Henry for the unit of inductance. A paper in 1892 that provided solutions for RLC circuits brought out the need for agreement on the names of circuit elements. Steinmetz, in a paper on hysteresis, proposed the term reluctance to replace magnetic resistance. Thus, by the turn of the 20th century the elements were in place for scientific circuit analysis and practical realization in communication systems.

That same year General Electric, at the insistence of Edwin W. Rice, bought Rudolph Eickemeyer's company for his transformer patents. The genius Charles Proteus Steinmetz (1865–1923) worked for Eickemeyer. In the saga of great ideas, I have always been as intrigued by the managers of great men as much as the great men themselves. E.W. Rice of General Electric personified true leadership when he looked past the misshapen dwarf that was Steinmetz to the mind present in the man. General Electric's engineering preeminence proceeded directly from Rice's extraordinary hiring of Steinmetz.

Charles Proteus Steinmetz

Arthur E. Kennelly's writings on impedance were meaningfully embellished by Charles Proteus Steinmetz's use of complex numbers. Michael Pupin, George A. Campbell, and their fellow engineers developed filter theory so thoroughly as to be worthwhile reading today.

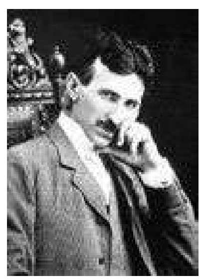
Steinmetz was not at the April 18, 1893 meeting, but sent in a letter of comment which included,

> *It is, however, the first instance here, so far as I know, that the attention is drawn by Mr. Kennelly to the correspondence between the electric term "impedance" and the complex numbers.*
>
> *The importance hereof lies in the following: The analysis of the complex plane is very well worked out, hence by reducing the technical problems to the analysis of complex quantities they are brought within the scope of a known and well understood science.*

The fallout from this seminal paper, its instantaneous acceptance by the other authorities of the day, its coalescing of the earlier work of others, and its utilization by the communication industry within a decade, makes it easily one of the greatest papers on audio ever published, even though Kennelly's purpose was to aid the electric power industry in its transmission of energy.

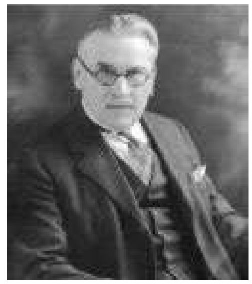
The generation, transmission, and distribution of electromagnetic energy today has no meaning in itself, but only gains meaning if information is conveyed, thus the tragedy of the use of mankind's precious resources to convey trash.

Nikola Tesla (1856–1943) working with Westinghouse designed the ac generator that was chosen in 1893 to power the Chicago World's Fair



Nikola Tesla

Dr. Michael I. Pupin of Columbia University was present at the Kennelly paper. Pupin mentioned Oliver Heaviside's use of the word impedance in 1887. This meeting established the correct definition of the word and established its use within the electric industry. Kennelly's paper, along with the ground-work laid by Oliver Heaviside in 1887, was instrumental in introducing the terms being established in the minds of Kennelly's peers.



Dr. Michael I. Pupin

1.8 Bell Laboratories and Western Electric

The University of Chicago, at the end of the 19th century and into the 20th century, had Robert Millikan, America's foremost physicist. Frank Jewett, who had a doctorate in physics from MIT, and now worked for Western Electric, was able to recruit Millikan's top students.

George A. Campbell (1870–1954) of the Bell Telephone Laboratories, had by 1899 developed successful "loading coils" capable of extending the range and quality of the, at that time, unamplified telephone circuits. Unfortunately, Professor Michael Pupin had also conceived the idea and beat him to the patent office. Bell Telephone paid Pupin $435,000 for the patent and by 1925 the Campbell-designed loading coils had saved Bell Telephone Co. $100,000,000 in the cost of copper wire alone.

Frank Jewett

To sense the ability of loading coils to extend the range of unamplified telephone circuits, Bell had reached New York to Denver by their means alone. Until Thomas B. Doolittle evolved a method in 1877 for the manufacture of hard drawn copper, the metal had been unusable for telephony due to its inability to support its own weight over usable distances. Copper wire went from a tensile strength of 28,000 lbs/in^2 with an elongation of 37% to a tensile strength of 65,000 lbs/in^2, an elongation of 1%.

George A. Campbell

Campbell's paper in 1922, "Physical Theory of the Electric Wave Filter" is still worthwhile reading today. I remember asking Dr. Thomas Stockham, "Do digital filters ring under transient conditions?" Dr. Stockham replied "Yes" and pointed out that it's the math and not the hardware that determines what filters do. Papers like Campbell's are pertinent to Quantum filters, when they arrive, for the same reasons Dr. Stockham's answer to my question about digital filters was valid.

Harold D. Arnold

Bell Telephone Laboratories made an immense step when H.D. Arnold designed the first successful electronic repeater amplifier in 1913.

H.D. Arnold at Bell Laboratories had taken DeForest's vacuum tube, discarded DeForest's totally false understanding of it, and, by establishing a true vacuum, improved materials and a correct electrical analysis of its properties enabled the electronic amplification of voice signals. DeForest is credited with putting a "grid" into a Fleming valve.

Sir John Ambrose Fleming (1848–1945) is the English engineer who invented the two-electrode rectifier which he called the thermionic valve. It later achieved fame as the Fleming valve and was patented in 1904. DeForest used the Fleming valve to place a grid element in between the filament and the plate. DeForest didn't understand how a triode operated, but fortunately Armstrong, Arnold, and Fleming did.

Sir Ambrose J. Fleming

Another Fleming—Sir Arthur (1881–1960)—invented the demountable high power thermionic valves that helped make possible the installation of the first radar stations in Great Britain just before the outbreak of WWII.

The facts are that DeForest never understood what he had done, and this remained true till his death. DeForest was never able, in court or out, to correctly describe how a triode operated. He did however; provide a way for large corporations to challenge in court the patents of men who did know.

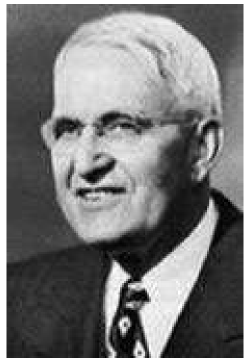
With the advent of copper wire, loading coils, and Harold D. Arnold's vacuum tube amplifier, transcontinental telephony was established in 1915 using 130,000 telephone poles, 2500 tons of copper wire, and three vacuum tube devices to strengthen the signal.

The Panama Pacific Exposition in San Francisco had originally been planned for 1914 to celebrate the completion of the Panama Canal but the canal was not completed until 1915. Bell provided not only the first transcontinental telephony, but also a public address system at those ceremonies.

The advances in telephony led into recording technologies and by 1926–1928 talking motion pictures. Almost in parallel was the development of radio. J.P.

Maxfield, H.C. Harrrison, A.C. Keller, D.G. Blattner were the Western Electric electrical recording pioneers. Edward Wente's 640A condenser microphone made that component as uniform as the amplifiers, thus insuring speech intelligibility and musical integrity.

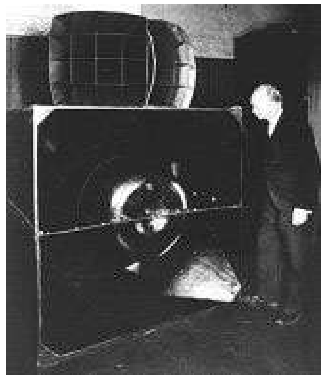
In 1933, Harvey Fletcher (1884–1981), Steinberg and Snow, Wente and Thuras and a host of other Bell Lab engineers gave birth to "Audio Perspective" demonstrations of three-channel stereophonic sound capable of exceeding the dynamic range of the live orchestra. In the late 60s, William Snow was working with John Hilliard at Ling Research, just down the street from Altec Lansing. It was a thrill to talk with him. He told me that hearing the orchestra level raised several dB was more astounding to him than the stereophonic part of the demonstration.



Harvey Fletcher

Edward C. Wente and Albert L. Thuras were responsible for full range, low distortion, high-powered sound reproduction using condenser microphones, compression drivers, multicellular exponential horns, heavy duty loaded low-frequency enclosures, the bass reflex enclosures, and both amplifiers and transmission lines, built to standards still challenging today. The Fletcher loudspeaker was a three-way unit consisting of an 18 inch low-frequency driver, horn loaded woofer, the incomparable W.E. 555 as a mid-range, and the W.E. 597A high-frequency unit.

Edward Wente



Albert L. Thuras

In 1959, I went with Paul W. Klipsch to Bell Labs where we jointly presented our redo of their 1933 Audio Perspective geometry tests. The demo was held in the Arnold Auditorium and afterward we were shown one of the original Fletcher loudspeakers. Western Electric components like the 555 and 597 are to be found today in Japan where originals sell for up to five figures. It is estimated that 99% of the existing units are in Japan. (As a side note, I genuinely earned a “Distinguished Fear of Flying Cross” with Paul Klipsch in his Cessna 180, the results of which entertained many SynAudCon classes.)

Paul Klipsch in the cockpit of his Cessna 180

The Western Electric 640A was superseded by the 640AA condenser microphone in 1942, still used today as a measurement standard by those fortunate enough to own one. The 640A was a key component in the reproduction of the full orchestra in 1933. When redesigned in 1942 as the 640AA, Bell Labs turned over the manufacturing of the capsule to Brüel and Kjaer as the B&K 4160.

Rice and Kellogg's seminal 1925 paper and Edward Wente's 1925 patent #1,333,744 (done without knowledge of Rice and Kellogg's work) established the basic principle of the direct-radiator loudspeaker with a small coil-driven mass controlled diaphragm in a baffle possessing a broad mid-frequency range of uniform response.

Paul Kilpsch and his assitant in his lab in Hope, Arkansas

Rice and Kellogg also contributed a more powerful amplifier design and the comment that for reproduced music the level should be that of the original intensity.

1.8.1 Negative Feedback—1927

In 1927 Harold S. Black, while watching a Hudson River ferry use reverse propellers to dock, conceived negative feedback for power amplifiers. With associates of the caliber of Harry Nyquist and Hendrik Bode, amplifier gain, phase, and stability, became a mathematical theory of immense use in remarkably diverse technical fields. Black's patent took nine years to issue because the U.S. Navy felt it revealed too much about how they adjusted their big guns and asked that its publication be delayed.

Harold S. Black

The output signal of an amplifier is fed back and compared with the input signal, developing a “difference signal” if the two signals are not alike. This signal, a measure of the error in amplification, is applied as additional input to correct the functioning of the amplifier, so as to reduce the error signal to zero. When the error signal is reduced to zero, the output corresponds to the input and no distortion has been introduced. Nyquist wrote the mathematics for allowable limits of gain and internal phase shift in negative feedback amplifiers, insuring their stability.

Hendrik Bode

Harry Nyquist (1889–1976) worked at AT&T’s Department of Development and Research from 1917 to 1934 and continued when it became Bell Telephone Laboratories in that year, until his retirement in 1954.

Harry Nyquist

The word “inspired” means “to have been touched by the hand of God.” Harry Nyquist’s 37 years and 138 U.S. patents while at Bell Telephone Laboratories personifies “inspired.” In acoustics the Nyquist plot is by far my favorite for first look at an environment driven by a known source. The men privileged to work with Harry Nyquist in thermal noise, data transmission, and negative feedback all became giants in their own right through that association.

Nyquist worked out the mathematics that allowed amplifier stability to be calculated leaving us the Nyquist plot as one of the most useful audio and acoustic analysis tools ever developed. His cohort, Hendrik Bode, gave us the frequency and phase plots as separate measurements.

Karl Kupfmuller (1897–1977) was a German engineer who paralleled Nyquist’s work independently, deriving fundamental results in information transmission and closed-loop modeling, including a stability criterion. Kupfmuller as early as 1928 used block diagrams to represent closed-loop linear circuits. He is believed to be the first to do so. As early as 1924 he had published papers on the dynamic response of linear filters. For those wishing to share the depth of understanding these men achieved, Ernst Guillemin’s book, *Introductory Circuit Theory*, contains clear steps to that goal.

Today's computers as well as digital audio devices were first envisioned in the mid-1800s by Charles Babbage and the mathematics discussed by Lady Lovelace, the only legitimate daughter of Lord Byron. Lady Lovelace even predicted the use of a computer to generate musical tones. Harry Nyquist later defined the necessity for the sampling rate for a digital system to be at least twice that of the highest frequency desired to be reproduced.

Nyquist and Shannon went from Nyquist's paper on the subject to develop "Information Theory." Today's audio still uses and requires Nyquist plotting, Nyquist frequency, the Nyquist-Shannon sampling theorem, the Nyquist stability criterion, and attention to the Johnson-Nyquist noise.

1.8.2 The dB, dBm and the VI

The development of the dB from the mile of standard cable by Bell Labs, their development and sharing of the decibel, dB, the dBm, and the VU via the design of VI devices changed system design into engineering design.

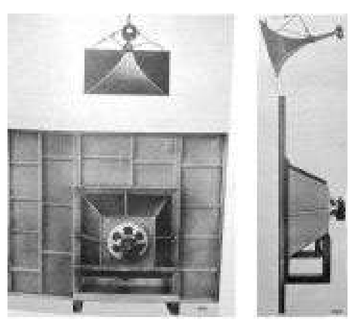
Of note here to this generation, the label VU is just that, VU, and has no other name, just as the instrument is called a volume indicator, or VI. In today's world, a majority of technicians do not understand the dBm and its remarkable usefulness in system design. An engineer must know this parameter to be taken seriously.

1.8.3 Bell Labs and Talking Motion Pictures

Bell Telephone Laboratories by the mid to late 1930s had from the inception of talking motion pictures in 1927–1928 brought forth the condenser microphone, exponential high frequency horns, exponential low frequency loudspeakers, compression drivers, the concepts of gain and loss, the dBm, the VU, in cooperation with the broadcasting industry, and installed sound in 80% of the existing theater market.

Yes, there were earlier dabblers thinking of such ideas but their ideas remained unfulfilled. What generated the explosive growth of motion picture sound—even through the deepest depression—was that only (1) entertainment, (2) tobacco, and (3) alcohol were affordable to the many and solaced their mental depression.

For physicists, motion picture sound was that age's "space race" and little boys followed the sound engineers down the street saying, "He made the movie talk." Dr. Eugene Patronis sent me a picture of the W.E. loudspeaker system installed in the late 1930s in which the engineer had actually aligned the H.F. and L.F. drivers. Dr. Patronis had worked in the projector booth as a teenager. He later designed an outstanding loudspeaker system for the AMC theater chain that was aligned and installed above rather than behind the screen, thereby allowing much brighter images. The system maintained complete spatial location screen-center for the audio.



W.E. loudspeaker system installed in the late 1930s

1.8.4 Motion Pictures—Visual versus Auditory

The first motion pictures were silent. Fortunes were made by actors who could convey visual emotion. When motion pictures acquired sound in 1928, a large number of these well-known personalities failed to make the transition from silent to sound. The faces and figures failed to match the voices the minds of the silent movie viewers had assigned them. Later, when radio became television, almost all the radio talent was able to make a transition because the familiar voices predominated over any mental visual image the radio listener had assigned to that performer.

Often, at the opera, the great voices will not look the part but, just a few notes nullify any negative visual impression for the true lover of opera, whereas appearance will not compensate for a really bad voice.

1.8.5 The Transition from Western Electric to Private Companies

A remarkable number of the giants in the explosion in precision audio products after WWII were alumni of Western Electric-Bell Labs, MIT, and General Radio, and in some cases, all three.

In 1928, a group of Western Electric engineers became the Electrical Research Products, Inc. (ERPI), to service the theaters. Finally a consent decree came down, as a result of litigation with RCA, for W.E. to divest itself of ERPI. At this point the engineers formed All Technical Services or Altec. That is why it is pronounced all-tech, not al-tech. They lived like kings in a depressed economy. As one of these pioneer engineers told me, “Those days were the equivalent of one ohm across Fort Knox.” They bought the W.E. Theater inventory for pennies on the dollar.

The motion picture company MGM had assembled, via Douglas

Shearer, head of the sound department, John Hilliard, Dr. John Blackburn, along with Jim Lansing, a machinist, and Robert Stephens, a draftsman, produced a proprietary theater loudspeaker which was named the Shearer horn. Dr. Blackburn and Jim Lansing did the high frequency units with Stephens, adapting the W.E. multicell to their use. It was this system that led to John Hilliard's correction of the blurred tapping of Eleanor Powell's very rapid tap dancing by signal aligning the high and low frequency horns. They found that a 3 inch misalignment was small enough to not smear the tapping. (Late in the 1980s, I demonstrated that from 0 to 3 inch misalignment resulted in a shift in the polar response.) Hilliard had previously found that there was on the order of 1500° in phase shift in the early studio amplification systems. He corrected the problem and published his results in the 1930s.

John Hilliard

After WWII, Hilliard and Blackburn, who both were at MIT doing radar work during the war, went their separate ways, with Hilliard joining Altec Lansing. Hilliard received an honorary Ph.D. from the Hollywood University run by Howard Tremaine, the author of the original *Audio Cyclopedia*, the forerunner to this present volume, *Handbook for Sound Engineers*.

Robert Lee Stephens left MGM in 1938 to found his own company. In the early 1950s I witnessed demonstrations of the Altec 604, the Stephens TruSonic co-axial and the Jensen Triaxial, side by side in my hi-fi shop, The Golden Ear. The Tru-Sonics was exceptionally clean and efficient. Stephens also made special 15 inch low-frequency drivers for the early Klipschorns. Hilliard, Stephens, Lansing and Shearer defined the theater loudspeaker for their era with much of the design of the Shearer multicells manufactured by Stephens.

In 1941, Altec bought Lansing Manufacturing Company and changed the Altec name to Altec Lansing Corp. James Lansing was enjoined by Altec to the use of JBL rather than Lansing for product names. He committed suicide in 1949, and JBL would have vanished except Edmond May, considered the most valuable engineer ever at JBL, stepped into the design breach with a complete series of high quality products.

In 1947, Altec purchased Peerless Electrical Products Co. This brought in not only the first designers of 20–20,000Hz output transformer, Ercel Harrison and his talented right-hand man, Bob Wolpert, but also the ability to manufacture what they designed. Ercel Harrison’s Peerless transformers are still without peer even today.

In 1949, Altec acquired the Western Electric Sound Products Division and began producing the W.E. product lines of microphones and loudspeakers. It was said that all the mechanical product tooling, such as turntables and camera items were dumped in the channel between Los Angeles and Catalina.

Jim Noble, H. S. Morris, Ercel Harrison, John Hilliard, Jim

Lansing, Bob Stevens and Alex Badmieff, my co-author for *How to Build Speaker Enclosures*, were among the giants who populated Altec and provided a glimpse into the late 1920s, the fabulous 1930s, and the final integration of W.E. Broadcasting and Recording technologies into Altec in the 1950s.

Jim Noble

Paul Klipsch in 1959 introduced me to Art Crawford, the owner of a Hollywood FM station, who developed the original duplex speaker. The Hollywood scene has always had many clever original designers whose ideas were for “one only” after which their ideas migrated to manufacturers on the West coast.

Alex Badmaieff

Running parallel through the 20s and 30s with the dramatic developments by Western Electric, Bell Labs, and RCA were the entrepreneurial start-ups by men like Sidney N. Shure of Shure Brothers, Lou Burroughs and Al Kahn of what became Electro-Voice, and E. Norman Rauland who from his early Chicago radio station WENR went on to become an innovator in cathode ray tubes for radar and early television.

When I first encountered these in men in the 50s, they sold their products largely through parts distributors. Starting the 1960s they sold to sound contractors. Stromberg-Carlson, DuKane, RCA, and Altec were all active in the rapidly expanding professional sound contractor market.

A nearly totally overlooked engineer in Altec Lansing history is Paul Veneklasen, famous in his own right for the Western Electro Acoustic Laboratory, WEAL. During WWII, Paul Veneklasen researched and designed, through extensive outdoor tests with elaborate towers, what became the Altec Voice of the Theater in postwar America. Veneklasen left Altec when this and other important work (the famed “wand” condenser microphone) were presented as Hilliard’s work in Hilliard’s role as a figurehead. Similar tactics were used at RCA with Harry Olson as the presenter of new technology. Peter Goldmark of the CBS Laboratories was given credit for the 33⅓ long playing record. Al Grundy was the engineer in charge of developing it, but was swept aside inasmuch as CBS used Goldmark as an icon for their introductions. Such practices were not uncommon when large companies attempted to put an “aura” around personnel who introduced their new products, to the chagrin and disgust of the actual engineers who had done the work.

From *Who Shot Liberty Valance*:

This is the west, sir, and when a legend and the facts conflict, go print the legend.

1.9 LEDE Studio

As Carolyn and I have looked back through our 66 years of travel we remember Starmusik Studios in Hamburg, Germany in 1985 as the scene of an international collection of "critical mass" of diverse audio and acoustic talent imaginable. Chips Davis had provided Ralf and Jenny Arnie, the owners, with an up-to-date LEDE control room and we were invited to conduct a seminar.

The staff was extraordinary and included Dr. Eugene Patronis, Richard Heyser, Ron McKay, along with Mr. and Mrs. Arnie. Starmusik Studios had produced the hit record in the studio "You're My Heart, You're My Soul" that year.

Dr. Eng Wolfgang Ahnert attended as our guest and brought along with him a "friend" from the East German broadcasting system who later turned out to be a STASI agent writing negative reports on Dr. Ahnert. When the Berlin wall came down the STASI files were opened to the victims.

1.10 Audio Publications

Prior to WWII, the IRE, Institute of Radio Engineers, and the AIEE, American Institute of Electrical Engineers, were the premier sources of technology applicable to audio. The Acoustical Society of America filled the role in matters of acoustics. I am one year older than the *JASA*, which was first published in 1929. In 1963, the IRE and AIEE merged to become the IEEE, the Institute of Electrical and Electronic Engineers.

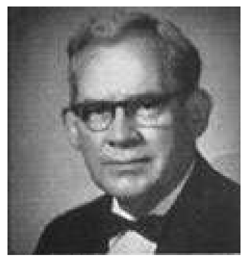
In 1947, C. G. McProud published *Audio Engineering* that featured construction articles relevant to Audio. Charles Fowler and Milton Sleeper started *High Fidelity* in 1954. Sleeper later published *Hi Fi Music at Home*. These magazines were important harbingers of the explosive growth of component sound equipment in the 1950s.

The Audio Engineering Society, AES, began publication of their journal in January 1953. The first issue contained an article written by Arthur C. Davis entitled, "Grounding, Shielding and Isolation."

Readers need to make a clear distinction in their minds between magazines designed as advertising media for "fashion design" sound products and magazines that have the necessary market information requiring the least reader screening of foolish claims. The right journals are splendid values and careful perusal of them can bring the disciplined student to the front of the envelope rapidly.

1.11 The "High" Fidelity Equipment Designers

By the beginning of WWII, Lincoln Walsh had designed what is still today considered the lowest distortion power amplifier using all triode 2A3s. Solid state devices, even today, have yet to match the perfection of amplifiers such as Lincoln Walsh's Brook with its all triode 2A3s or Marantz's EL34 all triode amplifier. The Walsh amplifiers, with the linearity and harmonic structure achieved by these seminal tube amplifiers, are still being constructed by devotees of fidelity who also know how to design reasonable efficiency loudspeakers. One engineer that I have a high regard for tells the story, *It wasn't that long ago I was sitting with the editor of a national audio magazine as his $15,000 transistor amplifier expired in a puff of smoke and took his $22,000 speakers along for the ride. I actually saw the tiny flash of light as the woofer voice coil vaporized from 30A of dc offset—true story folks.*



Lincoln Walsh

In the 1950s, a group of Purdue University engineers and I compared the Brook 10W amplifier to the then very exciting and unconventional 50W McIntosh. The majority preferred the 10W unit. Ralph Townsley, chief engineer at WBAA, loaned us his peak reading meter. This was an electronic marvel that weighed about 30lbs but could read the true full peak side-by-side with the VU reading on two beautiful VI instruments. We found that the ticks on a vinyl record caused clipping on both amplifiers but the Brook handled these transients with far more grace than the McIntosh.

We later acquired a 200W tube-type McIntosh and found that it had sufficient headroom to avoid clipping over the Klipschorns, Altec 820s, etc.

When Dr. R.A. Greiner of the University of Wisconsin published his measurements of just such effects, our little group were appreciative admirers of his extremely detailed measurements. Dr. Greiner could always be counted on for accurate, timely, and when necessary, myth-busting corrections. He was an impeccable source of truth. The home entertainment section of audio blithely ignored his devastating examination of their magical cables and went on to fortunes made on fables.

Music reproduction went through a phase of, to this writer, backward development, with the advent of extremely low efficiency book shelf loudspeaker packages with efficiencies of 20–40dB below the figures which were common for the horn loudspeakers that dominated the home market after WWII. Interestingly, power amplifiers today are only 10–20dB more powerful than a typical 1930s triode amplifier.

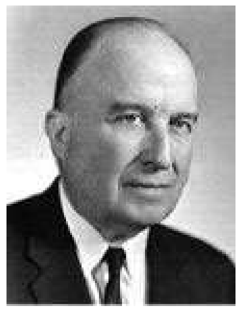
I had the good fortune to join Altec just as the fidelity home market did its best to self-destruct via totally unreliable transistor amplifiers trying to drive "sinkholes" for power loudspeakers in a marketing environment of spiffs, department store products, and the introduction of source material not attractive to trained music listeners.

I say "good fortune" as the professional sound was, in the years of the consumer hiatus, to expand and develop in remarkable ways. Here high efficiency was coupled to high power, dynamic growth in directional control of loudspeaker signals, and the growing awareness of the acoustic environment interface.

1.12 Sound System Equalization

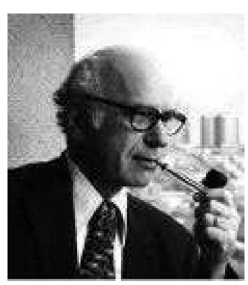
Harry Olson and John Volkmann at RCA made many advances with dynamical analogies, equalized loudspeakers, and an array of microphone designs.

Dr. Wayne Rudmose was the earliest researcher to perform meaningful sound system equalization. Dr. Rudmose published a truly remarkable paper in Noise Control (a supplementary journal of the Acoustical Society of America) in July 1958. At the AES session in the fall of 1967, I gave the first paper on the 1/3-octave contiguous equalizer. Wayne Rudmose was the chairman of the session.



Harry Olson

In 1969, a thorough discussion of acoustic feedback that possessed absolute relevance to real-life equalization appeared in the Australian Proceedings of the IREE. “A Feedback-Mode Analyzer/Suppressor Unit for Auditorium Sound System Stabilization” by J.E. Benson and D.F. Craig, illustrating the step-function behavior of the onset and decay of regeneration in sound systems.



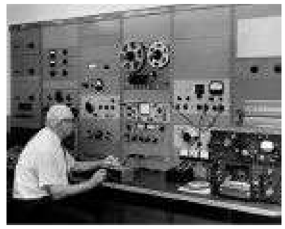
Dr. Wayne Rudmose

These four sources constitute the genesis of modern system equalization. Fixed equalization was employed by many early experimenters including Kellogg and Rice in the early 1920s, Volkmann of RCA in the 1930s, and Dr. Charles Boner in the 1960s.

Dr. Boner is shown here in the midst of installing filters hardwired one at a time "until the customer ran out of money"—was a famous quote. His demonstrations of major improvements in sound systems installed in difficult environments encouraged many to further investigate sound system design and installation practices, followed by custom ⅓ octave equalization. His view of himself was "that the sound system was the heart patient and he was the Dr. DeBakey of sound."

Dr. Charles Boner

The equalization system developed at Altec in 1967 by Art Davis (of Langevin fame), Jim Noble, chief electronics engineer, and myself was named Acousta-Voicing. This program, coupled precision measurement equipment and specially trained sound contractors, resulted in larger more powerful sound systems once acoustic feedback was tamed via band rejection filters spaced at 1/3-octave centers.



Art Davis in his lab at Altec

Equalization dramatically affected quality in recording studios and motion picture studios. I introduced variable system equalization in special sessions at the screening facilities in August 1969 to the sound heads of MGM—Fred Wilson, Disney—Herb Taylor, and Al Green—Warner Bros/7 Arts.

Sound system equalization, room treatment such as Manfred Schroeder's Residue Diffusers designed and manufactured by Peter D'Antonio, and the signal alignment of massive arrays led to previously unheard of live sound levels in large venues.

In the early days of equalization we never bothered to know the frequency, the wavelength, or the amplitude of the signals we were dealing with. We would bring the sound system into acoustic feedback, tune an uncalibrated oscillator to zero beat by ear or with Lissajous figures on an oscilloscope, followed by connecting the output of the oscillator to the filter to be tuned using a decade capacitor box to null the oscillator signal.

The filters were tapped inductors placed in the high side of an unbalanced line. By picking the correct taps we adjusted the depth of the filter. This tedious "one filter at a time" technique was replaced at the advent of one-tenth decade summing filter sets and the first real-time spectrum analyzers that were also set at one-tenth decade intervals. (The frequency centers were labeled in one-third octave intervals.)

One of the virtues of the old way was that you learned how to listen to the room/filter interaction. With the advent of real-time analysis, the focus of the user shifted to the visual display to the sometimes detriment of not listening to the step-by-step application of the filters. The orthogonal axes of the Lissajous display at a single frequency was a good preparation for understanding the Hilbert transform of all frequencies when Dick Heyser led us into the TEF world.

Measurement of any kind depends upon an observer. The most meaningful measurements are those made by an observer who has mentally conceived of what he wants to measure and consequently his training, sensitivity, and prejudices enter into the equation. The most interesting measurements one can encounter are those that occur unexpectedly in the midst of a carefully constructed "scientific" experiment. To quote Schrödinger,

> *So we come back to this strange state of affairs. While the direct sensual perception of the phenomenon tells us nothing as to its objective physical nature (or what we usually call so) and has to be discarded from the outset*
>
> *as a source of information, yet the theoretical picture we obtain eventually rests entirely on a complicated array of various informations, all obtained by direct sensual perception. It resides upon them, is pieced together from them,*

yet it cannot really be said to contain them. And using the picture we usually forget about them, except in the quite general way that we know our idea of a sound wave is not a haphazard invention of a crank but is based on experiment.

Democritus in the 5th century introduces the intellect having an argument with the senses about what is real. The former says: ostensibly there is color, ostensibly sweetness, obstinately bitterness, actually only Atoms and the void, to which the senses retort: poor intellect, do you hope to defeat us while from us you borrow your evidence? Your victory is your defeat."

Those of you with a love of the art and science of audio and acoustics and who can afford the luxury of the time it takes to stop and think as well as listen, the rewards are rich. Getting the job successfully completed using the tools you are already sure of is a question of craftsmanship, another highly regarded virtue by me and it's what pays the bills. I have full respect for those in the quest for either.

1.13 Real-time analysis RTA

Back in 1967 while attending a technical conference where Hewlett-Packard (H.P.) was demonstrating their 8051A loudness analyzer dedicated to the single function of measuring the loudness according to Zwicker, I realized that they had the capacity to build the one-third octave analyzer of my dream. I told them who I was and stated "I would purchase such an analyzer if they produced it." Roughly a year later they came to my office at Altec Lansing with their new HP model 8054A with a price tag of $9225.00. Altec was not willing to pay such a price, at that time, so I went down to the bank and I borrowed the money personally. I went out and remeasured each system I had equalized earlier using seconds instead of days to make the measurements along with the Altec one-third octave filter set I had developed with Art Davis (not related), thereby achieving greater acoustic gain and smoother amplitude response in every case—in some cases dramatically due to seeing for the first time in “real-time” what was actually going on in the environments we were working in. Altec soon reimbursed me and offered in a descending series of prices these real-time analyzer models to their sound contractors:

HP 8054A 1/3-octave analyer

- HP 8054A at $9225.00.
- HP H23 – 8054A act $5775.00.
- HP 8056A at $3500.00 which was designed to use any oscilloscope as the display.
- Altec/HP 8050A at $2660.00 which was low enough in price to sell over 500 units.

All of this within a two-year time period which left the Altec sound contractors involved with the AcoustaVoicing program in temporary charge of “audio planet Earth” as expressed by a contemporary student of the era. Today all forms of equalization are available, some to the improvement of quality, while others contribute to the confusion of newcomers. Proper use of equalization initially requires training and a full understanding of it requires education.

Altec/HP 8050A analyer

1.14 Altec Classes

The photograph of the first Acousta-Voicing contractor class shows them utilizing GR1650A impedance bridges, GR 1564 one-third of an octave analyzers, and the GR1382 random noise generators. These basic instruments allowed acoustic frequency response curves of loudspeakers in situ, and matching of the filter sets to the electronics.

A year later, Altec was able to provide the contractor’s with the special Altec-HP 8050A real-time analyzer plus an Altec-HP 8058A sound level meter. Shown in the photo is the original test set of filters that included a full set of one-third octave BRF filters from 50Hz to 12,500Hz plus a set of octave band filters from 62Hz to 8000Hz. Also in included were a pair of filters for high-frequency and low-frequency cutoff points. Provision was made for switching between having the filter set in and out of the circuit. A precision attenuator was provided to measure the increased acoustic gain before feedback with and without the filters.

Using this arrangement at a Los Angeles AES convention in the late 1960s, I was able, in a 20 minute session, to predict the probable acoustic gain for the environment we were in. We were using an HP9100A calculator, equalized the Altec A7 loudspeaker, and measured that we had slightly exceeded the expected gain.

One of the unintended consequences of this title wave of development was the almost complete overshadowing of Richard C. Heyser's TDS (time delay spectrometry) accomplishments where only one licensee was acquired in the first 10 years after Caltech first offered them. At that point my SynAudCon classes had begun to introduce Heyser's technology with the result that Caltech approached us to conduct the licensing in their name. We agreed to do so and in 6 months, renaming it the TEF (Time Envelope Frequency) process, we acquired 120 licensees. At that point our classes had the generous support of both Caltech and Dick Heyser and we requested that Caltech free us from the paperwork which they graciously did. Our remuneration for this work was the close association it gave us with Dick Heyser and his cornucopia of new ideas for presentation in our classes. Many of Heyser's techniques have been "borrowed" by today's FFT based programs such as spread spectrum, etc. while missing the deep significance inherent in greater signal-to-noise immunity.

Nothing in my experience better demonstrated the difference between training and education than these two measuring system. Few academicians have risen to the mental challenges Heyser proposed with the notable exception of Eugene Patronis at Georgia Tech. Now, 47 years later, the Heyser transform has yet to be exploited fully. In terms of real-time analysis two of the best publications by qualified people were Brüel and Kjaer's Technical Reviews No. 1, 1968 and No. 2, 1968 describing their efforts at utilizing some of Heyser's basic concepts. In studying the basics underlying TEF, Dick read Dennis Gabor's papers on "Theory of Communication" (1944), and especially Gabor's later paper "Communication Theory and Physics" (1953) wherein he stated,

> *Nor must one forget that vision, one of the most important paths of communication, is based essentially on quantum effects, and further I have been rightly criticized for having left out noise, which is an essential feature of all communication. In technical theories of thermal noise it is usually forgotten that it is not the noise power but it's fluctuations which caused a disturbance. This will be remedied here and at the same time will be brought in line with modern physics." Gabor remarked in his 1953 paper "We will limit the discussion to electric communications, though most of results will apply equally well to sound, or in short, to communication by any quantity which is considered as*

continuous in classical physics.

Papers like these have led to research by the military and intelligence communities into work that we won't hear about for years to come, witness the nearly 30+ years between Dr. Sidney Bertram's "Hells Bells" spread spectrum detection of mine fields by audio for the U.S. Navy in World War II and its publication in the IEEE literature 30+ years later. Perhaps it's not beyond reasonable hope that some really bright, well-educated, young man or woman, not willing to just accept the accepted, will resurrect Heyser's energy theorem and solve the "*its from bits*" potential therein.

For those in audio in need of a hint of where to start Dragan Samardzija (Bell Labs Holmdel, N.J.) paper entitled "Some Analogies Between Thermodynamics and Shannon Theory" wherein he comes to the conclusion that the average energy needed for the adiabatic compression of the ideal gas to 1/N of its initial volume is the same as the average energy needed to achieve the Capacity $C = \log_2 N$ of the equivalent *AWGN* (additive white Gaussian noise) channel by utilizing a piston in a tube, provides a useful link between acoustics and information theory. Ah to be young again.

1.15 JBL

MGM Studios used Jim Lansing's small company for the machining of the high-frequency drivers being designed by Hilliard and Blackburn. Among the key employees, that would later be taken over by the ERPI personnel when they became Altec, were Jim Lansing, Bill Martin (Jim Lansing's brother—they both changed their last names from Martinelli after observing that in Hollywood most of the people they met had changed their names as well), and

Ercel Harrison. Harrison was Peerless Transformers and had designed the first 20–20,000Hz audio output transformer for the electronics that John Hilliard had designed for MGM. Bill Martin became the head of the machine shop at Altec. Later during World War II the very knowledgeable Leo Beranek came to these man for advice in how to measure sound and ended up employing the early Western Electric sound level meter. The epitome of practical electro-acoustic achievements at that time lay on the West Coast. Altec, during World War II produced among other things, magnetic anomaly detectors for the detection of mines. One of the "other things" was the "Voice of the Theater" loudspeaker system that dominated the sound in postwar motion picture theaters.

1.16 Acoustics

As Kelvin was to electrical theory so was John William Strutt, Third Baron Rayleigh, to acoustics. He was known to later generations as Lord Rayleigh (1842–1919). I was employed by Paul W. Klipsch, a designer and manufacturer of high quality loudspeaker systems in the late 1950s. He told me to obtain and read Lord Rayleigh's *The Theory of Sound*. I did so to my immense long term benefit. This remarkable three-volume tome remains the ultimate example of what a gentleman researcher can achieve in a home laboratory. Lord Rayleigh wrote, *The knowledge of external things which we derive from the indications of our senses is for the most part the result of inference.*

The illusionary nature of reproduced sound, the paper cone moving back and forth being inferred to be a musical instrument, a voice, or other auditory stimuli, was vividly reinforced by the famous Chapter 1 from *The Theory of Sound*.

In terms of room acoustics, Wallace Clement Sabine was the founder of the science of architectural acoustics. He was the acoustician for Boston Symphony Hall, which is considered to be one of the three finest concert halls in the world. He was the mountain surrounded by men like Hermann, L.F. von Helmholtz, Lord Rayleigh, and others—early insights into how we listen and perceive. As one both researches and recalls from experience the movers and shakers of the audio-acoustic industry, the necessity to publish ideas is paramount.

Wallace Clement Sabine

Modern communication theory has revealed to us a little of the complexity of the human listener. The human brain has from 10^{15} to 10^{17} bits of storage and we are told an operating rate of 100,000 Teraflops per second. No wonder some “sensitives” found difficulties in early digital recordings and even today attendance at a live unamplified concert quickly dispels the notion that reproduced sound has successfully modeled live sound.

Boston Symphony Hall

We have arrived in the 21st century with not only fraudulent claims for products (an ancient art) but deliberately fraudulent technical society papers hoping to deceive the reader. I once witnessed a faulty technical article in a popular audio magazine that caused Mel Sprinkle (authority on the gain and loss of audio circuits) to write a Letter to the Editor. The Editor wrote saying Mel must be the one in error as a majority of the Letters to the Editor sided with the original author—a case of engineering democracy. We pray that no river bridges will be designed by this democratic method.

Frederick Vinton Hunt of Harvard was one of the intellectual offspring of men like Wallace Clement Sabine. As Leo Beranek wrote,

> *At Harvard, Hunt worked amid a spectacular array of physicists and engineers. There was George Washington Pierce, inventor of the crystal oscillator and of magnetostriction transducers for underwater sound; Edwin H. Hall of the Hall effect; Percy Bridgeman, Nobel Laureate, whose wife had been secretary to Wallace Sabine; A.E. Kennelly of the Kennelly-Heaviside layer; W.F. Osgood, the*

mathematician; O.D. Kellog of potential theory; and F.A. Saunders, who was the technical heir at Harvard to Sabine.

Hunt's success in 1938 of producing a wide range 5gram phonograph pickup that replaced the 5 oz units then in use led to Hunt and Beranek building large exponentially folded horns, a very high power amplifier and the introduction of much higher fidelity than had previously been available.

Dr. Hunt attended the technical session at the Los Angeles AES meeting in 1970 when I demonstrated the computation of acoustic gain for the sound system at hand, followed by Acousta-Voicing equalization in real time on the first H.P. Real Time Analyzer, all in 20 minutes. Dr. Hunt's remark to the audience following the demonstration insured the immediate acceptance of what we had achieved without any questions from the previous doubters. Dr. Hunt took genuine interest in the technology and was generous in his praise of our application of it. He said, "I don't fully understand how you have done it, but it certainly works."

The primary use of audio equipment is to take an acoustic signal (speech, music, noises, etc.), transform it into an electric signal to be amplified, recorded, measured, etc., and in most cases transformed back into an acoustic signal.

In this process the input and the output of the system goes from acoustic to acoustic and real-world environments such as acoustic noise, reverberation, echoes, etc. become an unwanted part of either the input or the output or both. While this is true for spaces as small as a phone booth all away up to miles of speech warning systems outdoors and mammoth arenas this is the one factor most beginners fail to recognize. Experience, especially failures, teaches that the acoustic source and the acoustic output are step number one in any sound system design.

Don and Carolyn Davis and Ron McKay

I have been privileged in my career to have worked with many talented acoustic consultants such as Ron McKay, Rolly Brook, Dave Klepper, Victor Peutz, and others that truly understood the complexities of acoustic environments as well as the complexities of electro-acoustic apparatus. These men many times helped us teach our classes in many parts of the world. At the same time they were designing world class acoustic spaces such as the Walt Disney Music Hall in Los

Angeles, and IRCAM in Paris. These men and others made it possible for us to make TEF measurements in Sabine's Boston Symphony Hall, the Grosser Musikvereinsaal in Vienna, the Concertgebouw in Amsterdam, and the Troy Savings Bank Music Hall in Troy, New York. It was the late George Szell, who said that if the Troy Savings Bank Music Hall was ever in danger of being torn down, he would rush to Troy and stand outside the entrance, arms outstretched. There may be halls that come close to equaling these four but none that will surpass them. The experience of hearing live performances in each of these halls followed by detailed measurements provided the baseline for what we call Fidelity.

Dave Klepper

A highly respected music critic for the New York Times, Harold Schonberg, wrote after hearing a concert in the Troy Hall,

> *Everyone In the audience was saturated with rich glowing sound—with strong tone of velvet sheen of incredible presence and sweetness, with a spectacular bass response that never sounded booming or artificial, with solo instruments coming through bigger than life and the various orchestral choirs perfectly balanced.*

Dorian Records, true pioneers in solving quality problems in early digital recordings, uses Troy for many of their recordings of exceptional young artists.

1.17 Professional-Level Audio Equipment Scaled to Home Use

World War II had two major consequences in my life (I just missed it by one year). The first was going to college with the returning G.I.s and discovering the difference in maturity between a gung-ho kid and a real veteran only one or two years older. The chasm was unbridgeable and left a lifelong respect for anyone who has served their country in the armed services.

As a young ham operator, I had obtained a very small oscilloscope, McMillan, for use as a modulation monitor. I had seen the General Radio type 525A at Purdue University, without realizing until many years later, the genius it embodied by Professor Bedell of Cornell, inventor of the linear sweep circuit, and H.H. Scott while working on it as a student at MIT with a job at General Radio as well.

The second was the pent-up explosion of talent in the audio industry especially that part misnamed *hi-fidelity*. Precision high quality it was, fidelity we have yet to achieve.

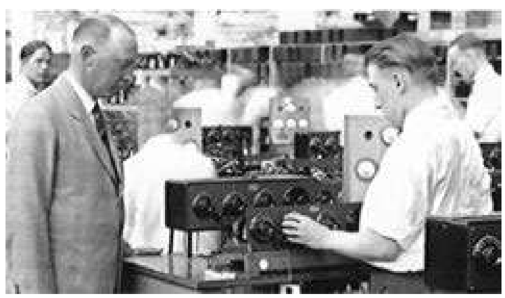
Directly after WWII a demand arose for professional level sound equipment scaled to "in the home use." Innovators such as Paul Klipsch, Lincoln Walsh, Frank McIntosh, Herman Hosmer Scott, Rudy Bozak, Avery Fisher, Saul Marantz, Alex Badmieff, Bob Stevens, and James B. Lansing met the needs of those desiring quality sound capable of reproducing the FM broadcasts and the fuller range that the advent of $33^1/_3$ vinyl records brought about.

Great audio over AM radio in the late 1920s and early 1930s ran from the really well-engineered Atwater Kent tuned radio frequency receiver (still the best way to receive AM signals via such classics as the Sargent Rayment TRF tuner) to the absolutely remarkable, for its time, E.H. Scott's Quaranta (not to be confused with the equally famous H.H. Scott of postwar years).

E.H Scott Quaranta reciver

Five special loudspeakers, automatic record changer, recording equipment, and ribbon microphone

This was a 48 tube superheterodyne receiver built on six chrome chassis weighing 620lbs with five loudspeakers (two woofers, midrange, and high frequency units) biamped with 50W for the low frequencies and 40W for the high frequencies. My first view of one of these in the late 1930s revealed that wealth could provide a cultural life.



Atwalter Kent in his factory

1.17.1 The Invention of Radio and Fidelity

The technical history of radio is best realized by the inventor/engineer Edwin Howard Armstrong (1890–1954). Other prominent figures were political and other engineers were dwarfed by comparison to Armstrong.

Edwin Armstrong

In the summer of 1912, Armstrong, using the new triode vacuum tube, devised a new regenerative circuit in which part of the signal at the plate was fed back to the grid to strengthen incoming signals. In spite of his youth, Armstrong had his own pass to the famous West Street Bell Labs because of his regenerative circuit work. The regenerative circuit allowed great amplification of the received signal and also was an oscillator, if desired, making continuous wave transmission possible. This single circuit became not only the first radio amplifier, but also the first continuous wave transmitter that is still the heart of all radio operations.

In 1912–1913 Armstrong received his engineering degree from Columbia University, filed for a patent, and then returned to the university as assistant to professor and inventor Michael Pupin. Dr. Pupin was a mentor to Armstrong and a great teacher to generations at Columbia University.

World War I intervened and Armstrong was commissioned as an officer in the U.S. Army Signal Corps and sent to Paris. While there and in the pursuit of weak enemy wireless signals, he designed a complex eight tube receiver called the *superheterodyne circuit*, the circuit still used in 98% of all radio and television receivers.

In 1933 Armstrong invented and demonstrated wide-band frequency modulation that in field tests gave clear reception through the most violent storms and the greatest fidelity yet witnessed. The carrier was constant power while the frequency was modulated over the bandpass chosen.

He had built the entire FM transmitter and receiver on breadboard circuits of Columbia University. After the fact of physical construction, he did the mathematics.

Armstrong, in developing FM, got beyond the equations of the period which in turn laid the foundations for information theory, which quantifies how bandwidth can be exchanged for noise immunity.

In 1922, John R. Carson of AT&T had written an IRE paper that discussed modulation mathematically. He showed that FM could not reduce the station bandwidth to less than twice the frequency range of the audio signal, "Since FM could not be used to narrow the transmitted band, it was not useful."

Edwin Armstrong ignored narrowband FM and moved his experiments to 41MHz and used a 200kHz channel for wideband, noiseless reproduction. FM broadcasting allowed the transmitter to operate at full power all the time and used a limiter to strip off all amplitude noise in the receiver. A detector was designed to convert frequency variations into amplitude variations.

Paul Klipsch was a personal friend of Edwin Armstrong: Mr. Klipsch had supplied Klipschorns for the early FM demonstration just after WWII. This was when Armstrong had been forced to move FM from 44–50MHz to 88–108MHz, requiring a complete redesign of all equipment. It was a stark lesson on how the courts, the media, and really big money can destroy genuine genius. Armstrong had literally created radio: the transmitters, the receivers for AM-FM-microwave in their most efficient forms. David Sarnoff made billions out of Armstrong's inventions, as well as an economic-political empire via the AM radio networks. No court or any politician should ever be allowed to make a technical judgment. Those judgments should be left to the technical societies as the "least worst" choice.

Edward Armstrong's breadboard system

The history of audio is not the forum for discussing the violent political consequences—Sarnoff of RCA totally controlled the powerful AM networks of the time. In 1954 attorneys for RCA and AT&T led to Armstrong's death by suicide. The current AM programming quality put on FM leaves quality FM radio a rare luxury in some limited areas.

1905s music system

The few, myself included, who heard the live broadcasts of the Boston Symphony Orchestra over the FM transmitter given them by Armstrong and received on the unparalleled, even today, Precedent

FM receivers know what remarkable transparency can be achieved between art and technology.

1.18 Acoustic Measurements

Plato said, “God ever geometrizes.” Richard C. Heyser (1931–1987), the geometer, should feel at ease with God. To those whose minds respond to the visual, Heyser’s measurements shed a bright light on difficult mathematical concepts. The Heyser Spiral displays the concepts of the complex plane in a single visual flash. Heyser was a scientist in the purest sense of the word, employed by NASA, and audio was his hobby. I am quite sure that the great scientists of the past were waiting at the door for him when he passed through. His transform has yet to be fully understood. As with Maxwell, we may have to wait a hundred years.

Richard Heyser

When I first met Richard C. Heyser in the mid-1960s, Richard worked for Jet Propulsion Labs as a senior scientist. He invited me to go to his basement at his home to see his personal laboratory. The first thing he showed me on his Time Delay Spectrometry equipment was the Nyquist plot of a crossover network he was examining. I gave the display a quick look and said,

“That looks like a Nyquist plot!”

He replied, “It is.”

“But,” I said, “No one makes a Nyquist analyzer.”

“That’s right,” he replied.

At this point I entered the modern age of audio analysis. Watching Dick tune in the signal delay between his microphone and the loudspeaker he was testing until the correct bandpass filter Nyquist display appeared on the screen was a revelation. Seeing the epicycles caused by resonances in the loudspeaker and the passage of non-minimum phase responses back through all quadrants opened a million questions.

Dick then showed me the Bode plots of both frequency and phase for the same loudspeaker but I was to remain a fan of seeing everything at once via the Nyquist plot.

To put all this in perspective (I worked at Altec at the time) I knew of no manufacturer in audio capable of making any of these measurements. We all had Brüel and Kjaer or General Radio frequency analyzers and good Tektronics oscilloscopes, but zero true acoustic phase measurement capabilities. I do not mean to imply that the technology didn’t exist because Wente calculated the phase response of 555A in the 1920s, but rather that commercial instruments available in audio did not exist until Richard Heyser demonstrated the usefulness of the measurements and Gerald Stanley of Crown International actually built a commercially available device. Heyser’s remarkable work became the Time, Envelope, Frequency (TEF) system, first in the hands of Crown International, and later as a Gold Line instrument.

The early giants of audio computed theoretical phase responses for minimum phase devices. A few pure scientists actually measured phase—Weiner, Ewask, Marivardi and Stroh, but their results had failed to go beyond their laboratories.

From 1966 until today, 48 years later, such analysis can now be embodied in software in fast, large memory computers. Dennis Gabor's (1900–1979) analytic signal theory appeared in Heyser's work as amplitude response, phase response, and Envelope Time Curves (ETC). One glance at the Heyser Spiral for impedance reveals Gabor's analytic signal and the complex numbers as real, imaginary, and Nyquist plot. The correlation of what seems first to be separate components into one component is a revelation to the first time viewer of this display. The unwinding of the Nyquist plot along the frequency axis provides a defining perspective.

Dennis Gabor

Heyser's work led to loudspeakers with vastly improved spatial response, something totally unrecognized in the amplitude-only days. Arrays became predictable and coherent. Signal alignment entered the thought of designers. The ETC technology resulted in the chance to meaningfully study loudspeaker–room interactions.

Because the most widely taught mathematical tools proceed from impulse responses, Heyser's transform is perceived "through a glass darkly." It is left in the hands of practitioners to further the research into the transient behavior of loudspeakers. The decades-long lag of academia will eventually apply the lessons of the Heyser transform to transducer signal delay and signal delay interaction.

I have always held Harry Olson of RCA in high regard because, as editor of the Audio Engineering Society Journal in 1969, he found Richard C. Heyser's original paper in the waste basket—it had been rejected by means of the idiot system of non-peer review used by the AES Journal.

1.19 Calculators and Computers

In the late 1960s, I was invited to Hewlett Packard to view a new calculator they were planning to market. I was working at this time with Arthur C. Davis (not a relative) at Altec, and Art was a friend of William Hewlett. Art had purchased one of the very first RC oscillators made in the fabled HP garage. He had used them for the audio gear that he had designed for the movie—Fantasia.

Don Davis and Tom Osborne

The 9100 calculator–computer was the first brainchild that Tom Osborne took to HP, after having been turned down by SCM, IBM, Friden and Monroe. (I purchased one; it cost me $5100. I used it to program the first acoustic design programs.) In 1966, a friend introduced Osborne to Barney Oliver, who after receiving Osborne's work arranged for him to meet Dave and Bill, to which Osborne said, "Who?" After one meeting with "Dave & Bill," Osborne knew he had found a home for his 9100. Soon Bill Hewlett turned to Tom Osborne, Dave Cochran, and Tom Whitney, who worked under the direction of Barney Oliver, and said, "I want one in a tenth the volume (the 9100 was IBM typewriter size), ten times as fast, and at a tenth of the price." Later he added that he "wanted it to be a shirt pocket machine."

HP 9100 calculator

The first HP 35 cost $395, was 3.2 × 5.8 × 1.3 in and weighed 9oz with batteries. It also fit into Bill Hewlett's shirt pocket. (Bill Hewlett named the calculator the HP 35 because it had 35 keys.) Art Davis took me to lunch one day with Mr. Hewlett. Because I had been an ardent user of the HP 9100 units, I was selected to preview the HP 35 during its initial tests in Palo Alto.

HP 35 calculator

In my mind, these calculators revolutionized audio education, especially for those without advanced university educations. The ability to quickly and accurately work with logarithm, trigonometric functions, complex numbers, etc., freed us from the tyranny of books of tables, slide rules, and carefully hoarded volumes such as Massa's acoustic design charts and Vegas's ten place log tables.

For the multitude of us who had experienced difficulty in engineering courses with misplaced decimal points and slide rule manipulation and extrapolation, the HP 35 released inherent talents we didn't realize we possessed. The x^y key allowed instant K numbers. The ten-place log tables became historical artifacts.

When I suggested to the then president of Altec that we should negotiate being the one to sell the HP 35s to the electronics industry (Altec then owned Allied Radio,) his reply stunned me, "We are not in the calculator business." I thought as he said it, "Neither is Hewlett Packard." His decision made it easy for me to consider leaving Altec.

I soon left Altec and started Synergetic Audio Concepts, teaching seminars in audio education. I gave each person attending a seminar an HP 35 to use during the 3-day seminar. I know that many of those attending immediately purchased an HP calculator, which changed their whole approach to audio system design. As Tom Osborne wrote, "The HP 35 and HP 65 changed the world we live in."

Since the political demise of the Soviet Union, "Mozarts-without-a-piano" have been freed to express their brilliance. Dr. Wolfgang Ahnert, from former East Germany, was enabled to use his mathematical skills with matching computer tools to dominate the audio-acoustic design market place.

Observing the current use of iPhones containing calculator, spectrum analyzer, impedance bridges, and other measurement devices all in the package you can hold in your hand led me to realize what the young engineers of today are missing from the past.

Many of the scientist/engineers who trained me often had built their own instrument in order to achieve the accuracy they desired in their measurements. Atwater Kent's Unisparker ignition system made him a multimillionaire (they were the automobile industry standard for almost 50 years) and he then proceeded to build the most prestigious broadcast receiver factory in the 1920s and early 1930s. His TRF receivers were the ideal design for receiving AM broadcasts. The factory he built led the way to modern well lit industrial work places. His exploitation of Bakelite introduced the idea of plastics to an entire industry.

1.20 Other Giants in Our Lives

Having spent the past 64 years actively engaged in audio and acoustic experiences, and at the very real risk of leaving worthy participants out of the list, we'd like to remember some very human personalities that made contributions to our understanding of the diversity of backgrounds that brought about our current industries.

The list includes inventors, engineers, artists, physicists, psychologists and psychiatrists, ministers, and military personnel all of whom we came to regard as good friends.

Dr. Eugene Patronis Jr. Dr. Patronis is a gifted teacher with a Ph.D. in Physics (started teaching at Georgia Tech when he was 19 years old). His Ph.D. in Physics is worn lightly by this extremely capable engineer, scientist, and extraordinary communicator. Those who interacted with him during the development years of TEF technology gained a deeper appreciation both of the man and the technology. After the passing of Dick Heyser, Gerald Stanley of Crown International and Gene Patronis were the only two fully equipped to understand both the electronic technology and acoustic applications it offered. I'll give but one example of Gene's effect on one of our young friends. This young man was made aware of Gene at a SynAudCon class and later when the Navy sent him to Georgia Tech he made sure to attend Gene's course there. It was a rigorous course that he wasn't required to take, but he felt the desire to have studied under such a teacher. His association with Gene at Georgia Tech led him in 2012 becoming the captain of the USS New York active in the Persian Gulf and the Mediterranean Sea.

Gene Patronis, Jr.

Gene is one of those gifted men who can envision at the highest mathematical levels, build it with his own two hands, and install it professionally while at the same time inspiring young men to strive to utilize their gifts to the fullest. A treasured mentor to me and a close personal friend in the highest meaning of the word.

Dr. Eugene Patronis, Jr. with his Pataxial loudspeaker

Gene's accomplishments span from working for the Manhattan project to designing a unique loudspeaker system for use in the White House. This unit was designed to be an integral part of the architectural room feature it was mounted in, and provided full range to the users. It has been reliably reported that distinguished users thought the microphone was the entire system and that the rest was accomplished by magic.

Ken Wahrenbrock. Ken Wahrenbrock, an ordained minister of a mainstream denomination, as well as a highly skilled computer draftsman-engineer had been instrumental in introducing computer drafting to one of the West Coast airplane manufacturers. Having served as a crypto analysts during World War II he personified the scholar and gentlemen the world is so in need of.

Ken Wahrenbrock with a PZM microphone

Witnessing Ed Long's demonstration of a condenser microphone capsule placed facedown a few millimeters from a hard floor, Ken searched out a suitable electret capsules to mount on the head of a precision screw that could be raised and lowered above the plate with the front of the capsule facing down towards the surface of the plate. He quickly found the optimum spacing for obtaining a pressure zone that continued well above the range of human hearing, which turned out to be the thickness of his business card.

The first practical Pressure Zone Microphone, PZM, quickly spread worldwide and Ken was soon recognized as the indefatigable demonstrator of their uses to enthusiastic listeners.

Dr. John Diamond. Dr. John Diamond, psychologist, psychiatrist, and medical Dr., whose unforgettable demonstrations at an AES convention in Los Angeles that provided a real non-Markovan moment for engineers in the audience. He later presented an unforgettable demonstrations of the use of the deltoid muscle in diagnosing ailments at a SynAudCon meeting in New York at the Waldorf-Astoria.

Chips Davis and Russ Berger. Two men who revolutionized recording studio control rooms all over the world. This picture was taken early one morning at our ranch seminar center in the Santa Ana Mountains after they had walked over from the bunkhouse in the pouring down rain to attend the class. This was taken before their fame but we've always felt represented the goodwill, humor, and the ability to overcome adverse situations that characterized both of them. These two men working with the concepts of live end – dead end, LEDE, reflection free zone RFZ, and the use of diffusers of the quadratic residue type change control room design forever.

Chips Davis and Russ Berger

Deward Timothy. Deward Timothy, a devout Mormon, and an intense audio and acoustic scholar. He has repeatedly demonstrated the usefulness of PET analysis in obtaining optimum coverage of the direct sound and minimizing the excitation of reflected sound in difficult spaces. It is quiet men, like Deward, consistently designing and building superior sound systems as witness to what can be accomplished.

Fred Fredericks. Fred Fredericks, a genuine military hero, who served all over the globe in highly diverse, dangerous surveillance work after having gained recognition for a distinguished fighting career in Korea. Fred, a skilled pilot, flew into and successfully out of many denied areas during his military career. He went on to a series of successful careers in audio. The photograph shows him presenting me with an original Western Electric high-frequency driver from his personal collection. Those fortunate to know him remember what a "command presence" voice sounds like.

Don Davis and Fred Fredericks

Dr. Thomas Stockham. Dr. Stockham, a man whose wife worshiped the air he walked on, and who did the first successful truly high-quality digital recordings and laid pioneer groundwork for the century to follow.

Dr. Thomas Stockham

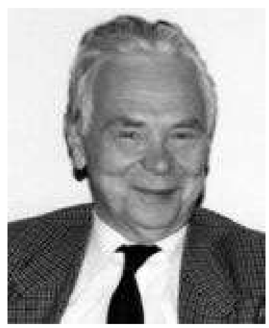
Gerald Stanley. A giant capable of straddling both the analog and digital world when he designed the world's first dedicated TEF analyzer utilizing the best of both worlds, Gerald Stanley personifies the epitome of late 20th century engineering that provided the tools for driving progress in the first decade of the 21st century. Known worldwide as the most capable power amplifier designer for his work at Crown International, Gerald had a part in many projects that the founders of Crown chose to give the world.

Gerald Stanley

Dr. Sidney Bertram. Dr. Sidney Bertram, who in World War II, designed the mine detection system called “Hells Bells” that enabled seven U.S. submarines to enter the Sea of Japan through an extremely dense minefield by using a swept spectrum sonar that provided the operator with a bearing and range to the mines and consequently played a key part in denying food supplies from mainland China getting to Japan. When Dick Heyser passed away, we gathered a panel to attempt the analysis of Dick’s energy theorem. I’ll never forget being present when Dr. Patronis, Gerald Stanley, and Sidney Bertram demonstrated what truly unique minds can do with analysis.

Dr. Sidney Bertram

Victor Peutz. Victor Peutz of Nijmegen Holland and Harvey Fletcher of Bell telephone laboratories were the two-men that played a major part in the 20th century to our understanding of speech intelligibility. Peutz's major contribution was the programming of the major components, which turned out to be ambient noise, and the percent loss of consonants, for the new TEF analyzers thus allowing direct measurements of these parameters thereby providing a speech intelligibility estimate that correlated closely with large groups of live listeners. While today some popular systems based primarily on ambient noise and reverberation are widely used due to the speed of measurement only the Peutz system allowed the user to not only obtain a score but also to see what caused the problem if one was present.



Victor Peutz

Mr. Peutz participated in the very large workshop, of over 100 participants, in live measurements of intelligibility versus the TEF estimates and tirelessly led the workshop to a fuller understanding of both the classic measurements and the modern measurement estimates.

Mr. Peutz later led an extensive workshop in Europe where we measured multiple concert halls, attended performances in them, and later interviewed the orchestra conductors as to their judgment of the halls. This gentle, erudite, thoroughly competent scientist shared freely his storehouse of knowledge with those ready to receive advanced understanding of acoustic spaces. Mr. Peutz was one of a small handful capable of discussing Dick Heyser's mathematical energy theories and as a result became a close friend of Dick. Victor personified the classic European gentlemen of the old-school.

Bill Hazlitt. We need to be reminded occasionally what kind of man served in World War II that never wore a uniform or could've passed a physical. I had attempted for months to obtain an interview with the head of an enormous distribution company in order to present Altec products to him. Finally I took Bill Hazlitt,

Altec Eastern Regional Manager, with me to the firm. As we were sitting in the waiting room an executive of the firm passed us, stopped in his tracks, and exclaimed "Bill Hazlitt" and immediately conducted Bill into the president's office. When the executive came out of the office I asked him how did you come to know Bill Hazlitt. He replied, "I was part of the crew of a B-17 bomber during the war over Europe and Bill was the Fairchild camera representative to the squadron. Because he drank with us he decided he should fly with us and did so for many hazardous missions over Germany. I could never forget him, he didn't have to go."

Bill Hazlitt

Bill, in his younger years, had been a steamboat captain on the Mississippi and a "Wobbly" in Montana at a most dangerous time to be involved in such a labor movement. When I had the privilege of working with him he was a distinguished older gentleman.

Manny Mohageri. Emilar loudspeakers were the result of a very synergetic partnership between Harold Lindsay, of Ampex fame, Jonas Renkus, a skilled designer of high-frequency compression drivers, and Manny Mohageri, a very talented engineer from Iran.

Manny Mohageri

Their first product, the EA–175 compression driver proved superior to its competitors with its smoother frequency response and lower distortion.

Manny was an ardent aviator and flew his Mooney single-engine aircraft solo around the world without ever mentioning it to the media, but only to his close friends. His home in California was a recreation of 19th century Persian culture. He had been trained as a petroleum engineer in the United States. His brother in Iran was a Bahai and when he and his family were threatened with death if they stayed in Iran they were smuggled across the border to Afghanistan by wearing sheepskins and crawling across the border with a herd of sheep. When their guide told them, “You can stand up now!” they were startled to find that they were not the only ones in the herd. Having difficulty getting them visas into the United States, Manny bought an island off the coast of Panama that included its own airstrip, and situated them there until they could apply for citizenship in the U.S.A.

Ralph Townsley. In my youth Ralph Townsley was the chief engineer for radio station WBAA at Purdue University in *West* Lafayette, Indiana. *WBAA* had the distinction of being Indiana's first broadcasting station and as a result had a clear channel dedicated to it. One famous blooper by an Ralph Townsley announcer, that delighted the student body, was "Indiana's first broadcasing station."

Ralph Townsley

What Ralph didn't know about audio wasn't worth learning and his basic test given to graduate electrical engineers wishing to work at the station was rigorous, but very basic, and often failed.

Ralph designed and built a peak reading VU side-by-side with a standard VU. It was capable of reading the true peak in 1Hz out of 20,000Hz. He loaned it to us in the 1950s to study the true peaks produced by ticks in vinyl records at the output of power amplifiers, Ralph's wide dynamic range level meter and his improved General Radio decade box (Ralph had primary standards available to him that allowed such improvements in what were already considered the best instruments available) were typical of the man's quest for accuracy.

WBAA when listened to on Sargent Rayment TRF tuners had AM signals undistinguishable from FM, except during electrical storms. The control room and studio was, for that early time, a model of automation via pushbutton relay controlled signal paths. Whenever faced with a "high fidelity myth" a visit to Ralph in his workshop at the station always led to a gentle explanation of what was actually true. When in the 1980s he visited one of our classes I told the attendees that a man was present who knew a great deal more than I did and that he was free at any time to correct or amplify any of my statements and I meant every word of it.

In 1965 Ralph used his access to an IBM 360 to expand the "K" number charts I was doing for Art Davis at Altec to 20 places as a gift, along with considerable insight into their use in the design of both simple and complex networks. It was later published as a book. While use of tables was replaced by computers available to all, the insights that Ralph had regarding the networks is still a valuable purchase.

James Moir. James Moir, an English patriot, had in the 1930s developed a pulse technique for measuring the influence of reflections on the direct sound level in theaters. Because of his combined electronic—acoustic experience, he was a natural choice for the team that came over to Western Electric to guide them in the production of the radar equipment that won the "Battle of Britain" in the skies in 1940. All of this was highly secret work in as much as it was in violation of our "neutrality." It was the British who brought the magnetron to us as a necessary component to do the job that couldn't be completed, due to bombing, industrially in Britain at that time. James Moir made many trips to the USA for AES conventions. We attended an AES panel with Mr. Moir. The panel consisted of American mixers and European tone meisters. The tone meisters related how they had to study the history of the music as well as the music itself plus a great deal of technical training in both electronics and acoustics, whereas the American mixer replied he couldn't see any relevance in all that training just to sit at the mixing console. Mr. Moir leaned over and whispered in my ear *"I found knowing how to hoe potatoes useful in my audio career*."

James Moir

Mead Killian. Mead Killian of Etymotic Research is the man that provided us with unique probe microphones that allows measurement directly on the surface of the eardrum. The accessibility of Mead Killian such instruments allowed our exploration of the in-the-ear (ITE) recordings, Pinnae measurements, and on to in-the-ear monitors of the present era.

Mead Killian

Nat Norman. I had the privilege of meeting Nat Norman at his apartment in Greenwich Village in 1958. He worked at the famous West Street Laboratories of Bell Telephone. Nat Norman held, at that time, seven U.S. patents, including one on the program amplifiers—the Western Electric 1126—used in the majority of radio and television stations of that day.

Nat Norman

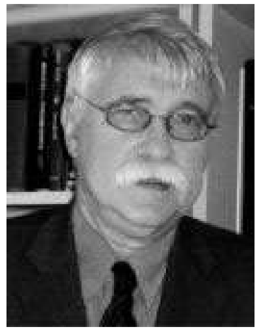
Mr. Norman had hand built Nat Norman his own Klipschorns and had used the W.E. 555 midrange driver along with a Bostwick high-frequency unit on top of the standard Klipsch woofer horn.

Mr. Klipsch and I lent appreciative ears to the concert he provided us, while out on the street our parked car was being burglarized of my German 35mm Canon cameras and Paul's priceless notebooks that were in his briefcase. When the police were called their only comment was "you *should've known better than to park on the street in Greenwich Village with an out-of-state license plate"*. It was apparent that they intended to do nothing about it and felt that they had fulfilled their duty when they gave us a copy of their report for our Insurance purposes.

Nat had worked with such sound pioneers as Dr. Harvey Fletcher, Will Munson, E.C. Wente, A.L. Thuras, J. P. Maxfield, H.C. Harrison, H.G. Bostwick, and Major Armstrong, the inventor of FM. Nat and Paul Klipsch had supplied the audio equipment for Armstrong's early demonstrations of superiority of FM for the broadcasting of classical music. It was interesting to me that Paul was always accepted as a Peer among these pioneers, not as a merchant promoting his loudspeaker.

Nat's living quarters in Greenwich Village included his photography studio where, using his Linhof Technika camera he photographed nudes and other interesting inhabitants of the area. His bohemian lifestyle, unlike the early hippies in the area, was well-financed, artistic, and creative.

Dr. Wolfgang Ahnert. Dr. Wolfgang Ahnert came into our lives at an AES conference in the Netherlands when he came up and introduced himself to us with the request "You will invite me to America". After a short discussion I realized that I was talking to someone from the other side of the Berlin Wall. Naturally my curiosity was aroused and I soon found myself gathering the necessary information to try and enable such a visit.



Dr. Wolfgang Ahnert

Because Carolyn and I were engaged with consulting work for the Navy at that time we felt it wise to obtain proper clearance before proceeding to invite him over to the United States. Steve Simpson, a Syn-Aud-Con graduate, and a former OSS hero of mine was contacted and asked if he still had any contacts in Washington that could check out Dr. Ahnert for us. It turned out that his contact in Washington was his former partner in London during World War II, William Casey, the present head of the CIA at that point in time. We rapidly heard from CIA officials that we should go ahead and invite him over. At that point we gave Dr. Ahnert the date of the class we would hold in Florida in the fall of the year. That class had three employees of the CIA openly register for the class as audiovisual personnel. Ahnert simply never showed up. We later found out that the work he was doing for an Austrian company that brought dollars to the East hadn't been completed in time for him to get out for the trip. We did however enjoy the CIA employees who it turned out had some very practical shooting experiences.

A year later Dr. Ahnert called us from Chicago saying that he was in the country and would like for us to meet him at the Louisville airport so he could travel with us to our classes on the East Coast. We picked him up at the airport and brought him back to the farm in Indiana. He had already had one experience on the flight to

Louisville that contradicted the propaganda he had been taught in East Berlin. The stewardess was a young black girl. Because the cattle were bawling he wouldn't eat supper until they had been fed, revealing an unexpected side of his nature. The next day we went up to meet Ralph Townsley and he discovered that a private citizen could own a laboratory more elaborate than the best available to him in Germany at the time. Traveling east, with our truck and trailer, we became involved with the edge of a tornado in western Pennsylvania which caused us to stop short of our planned destination and put the good Dr. in the motel for the night. In the morning we found he had discovered the motel's Gideon Bible and was reading it. New York City allowed us to introduce the good Dr. to the "Beard", David Andrews, who became his guide to the city. He couldn't believe that David Andrews apparently had a key to the back doors of every major building in the city and that he knew everything about sound in the city.

Our next stop was at Glen and Debbie Ballou's home in Connecticut where we discovered that sitting with his back to an open window made him very uncomfortable and the drapes had to be closed before he could eat. The following day was a Sunday and Glen invited him to go to church with them and he accepted. The sermon that day was by a female minister and was entitled, I'm not joking, "So you want to nuke those commie bastards". The ministers message, of course, was on brotherly love and learning to live with the diversity that was present in our world.

We had a class scheduled in Framingham, Massachusetts and took the opportunity to introduce Dr. Ahnert to the Bose operation. Later we found out that the class had the disguised CIA agents in attendance.

The next stop was at Phil Clark's where the good Dr. experienced his first ride on a deluxe riding lawnmower. We all drove over the next day to the Troy auditorium to do acoustical measurements of that extraordinary hall, then on into Canada for a Toronto class and a visit with Neil Muncy. At this point we discovered an interesting thing about people from the other side of the curtain. They lack most of the amenities of life and were unabashed about asking us to provide some of them. To make a long story short we bought him the largest chainsaw we could find which he packed in his luggage and flew home with it.

In thinking back on it he met a lot of key people in the United States that reflected the true power we have at hand. I'll never forget his first sight of an automobile junkyard with thousands of wrecks spread across acres of ground and the emotional effect it had on him. He couldn't imagine that we discarded valuable articles so casually and left them in the open where anyone could steal from them.

On subsequent visits we found he had absorbed the lessons America had provided and it was our turn to discover how really clever he was, truly a Mozart without a piano, because of the restricted access to computers east of the Wall. Eventually after the Wall came down and he acquired proper computers the EASE programs were the result.

Before the Wall came down we made a visit to Berlin and Dr. Ahnert entertained us at his home in East Berlin. We took Helmuth Kolbe with us because of his knowledge of German and his familiarity with the East from his recording work on historic organs behind the wall for Columbia records. At this time Dr. Ahnert's life was in greater danger than he knew because a pretended friend in the broadcast system was the secret STATSI agent reporting unfavorably on his activities. During our tour of West Berlin, Dr. Ahnert came over and from the platform that allowed people to look into East Berlin described each point of interest on the other side of the wall to the amazement of other tourists on the platform. To drive from West Germany to West Berlin, required driving through East Germany, which at that time was a gruesome introduction to a fully militarized police state. The drive from West Berlin back to West Germany was equally stressful due to their ridiculously low speed limit that reminded us that only Third World nations and the USA have such ridiculous limits. Upon arriving home in the USA I was driving from the airport to the farm when Carolyn asked "what's wrong with the car" to which I replied nothing, honey, it's the speed limit which at that time was 55 mph.

Dr. Ahnert's spirit of entrepreneurship and innate love of freedom was proof to us that no regime can control men's minds and once the wall came down he demonstrated his innate worthiness very successfully.

Capt. Glen Ballou. Sailor, engineer, author, and good friend, Glen and his wife, Debbie, date back to my Altec years. Glen was the Pratt and Whitney, Sikorsky Aircraft, and United Technologies Corporate audio and visual expert who used Altec equipment in the design of board rooms and trade shows around the world.

Glen Ballou

Glen fully appreciated what we were trying to accomplish in founding Syn-Aud-Con and gave his full unrestricted support to us by becoming our first New England representative and filling our early classes. Glen always enjoyed visiting us at our home in the Cleveland National Forest, California and at our farm in Indiana, where between skiing and horse backing cemented our friendship in spite of painful adventures.

When I was asked to take over the original *Audio Cyclopedia* by the publisher (I had been a friend of the original creator, Howard Tremaine), I took Glen to the publisher as the best equipped professional in audio to carry out the job. Glen more than fulfilled the requirement.

Glen is “Capt. Glen” because he personifies the true “blue water” sailor with his 46 foot Beneteau sloop, 20,000 “blue water” miles of experience, and his 100 Ton Master Coast Guard license.

Glen has a rapid detection system for talent having known personally both Igor and Sergei Sikorsky. He enjoys many hobbies, as did Sergei Sikorsky—boating, flying, scuba diving, and classical music. Sergei was also an ardent shooter which Glen only tolerates as an observer.

We first met back in the mid-1960s and his worth has stood the test of over 50 years.

Richard Clarke. Richard Clarke, whose custom sound system in an automobile decimated all contestants in the production of quality sound inside the car, and his later exploit included the loudest sound ever produced inside a step van. Richard purchased the special in-the-ear microphones and used them to make exceptionally realistic recordings of Penske's Mercedes team cars at the Indianapolis motor Speedway including up close activities in the pits. Later ventures were the establishment of a DNA company and the purchase of the Western Electric plant that had built Nike missiles. When he took over that plant he found unused vacuum tubes used in the Atlantic cable and other priceless Western Electric sound equipment, see the letter he sent at the end of this chapter. Easily one of the most focused entrepreneurs it has ever been my privilege to know.

The Meaning of Communication. The future of audio and acoustics stands on the shoulders of the giants we have discussed, and numerous ones that we have inadvertently overlooked. The discoverers of new and better ways to generate, distribute, and control sound will be measured consciously or unconsciously by their predecessor's standards. Fad and fundamentals will be judged eventually. Age councils that "the ancients are stealing our inventions." The uncovering of an idea new to you is as thrilling as it was to the first person to do so.

The history of audio and acoustics is the saga of the mathematical understanding of fundamental physical laws. Hearing and seeing are illusionary, restricted by the inadequacy of our physical senses. The science and art of audio and acoustics are essential to our understanding of history inasmuch as art is metaphysical (above the physical). Also art precedes science.

That the human brain processes music and art in a different hemisphere from speech and mathematics suggests the difference between information, that can be mathematically defined and communication that cannot. A message is the flawless transmission of a text. Drama, music, and great oratory cannot be flawlessly transmitted by known physical systems. For example, the spatial integrity of a great orchestra in a remarkable acoustic space is today even with our astounding technological strides only realizable by attending the live performance.

The complexity of the auditory senses defies efforts to record or transmit it faithfully.

The perception of good audio will often flow from the listener's past experience, i.e., wow and flutter really annoys musicians whereas harmonic distortion, clipping, etc., grate on an engineer's ear–mind system.

I have not written about today's highly hyped products as their history belongs to the survivors of the early 21st century. It can be hoped that someday physicists and top engineers will for some magic reason return to the development of holographic audio systems that approach fidelity.

Telecommunication technology, fiber optics, lasers, satellites, etc. have obtained worldwide audiences for both trash and treasure.

The devilish power that telecommunications has provided demagogues is frightening, but shared communication has revealed to a much larger audience the prosperity of certain ideas over others, and one can hope that the metaphysics behind progress will penetrate a majority of the minds out there.

That the audio industry's history has barely begun is evidenced every time one attends a live performance. We will, one day, look back on the neglect of the metaphysical element, perhaps after we have uncovered the parameters at present easily heard but unmeasurable by our present sciences. History awaits the ability to generate *the* sound field rather than *a* sound field. When a computer is finally offered to us that is capable of such generation, the question it must answer is: ***"How does it feel?"***

1.21 Sound System Engineering

The 4th edition of *Sound System Engineering* offers a unique opportunity to put three mentors on your bookshelf who participated in the best of the 20th century's mental transform of audio and acoustics into the exponentially expanding techniques of the 21st century. The book relates personal recollections of the Bell Telephone Laboratories work in the 1930s up to the current attempt to control audio systems with your mind, not only gives you all the clues necessary to your own individual search of the past but the starting point for creative work in the present century. Anyone who can truthfully say he already knows the material in every chapter should be writing a book the three of us would be eager to read in as much as each one of us is still an ardent student of the world of audio and acoustics. What it took us a collective 150 years of experience and study we have shared in this volume. Now is your chance to find out if you can absorb it in a shorter period of time. Years ago Dick Heyser gave a paper he said took him 10 years to formulate, and he felt a certain pain giving it away in a single paper. Those of us, his listeners, then discovered yes!, he had given it away in a 10 minute paper; but 10 years later we were still sorting it out. But what he had given us was the desire to reach the plateau he was on and thereby climb out of the valley we were in. May the readers of this volume experience a similar urge and be the leaders in their century.

Don

Great talking to you and Carolyn today. Please find enclosed a gift for you and Gene. I promise that you will not have anything in your tech collection that is more rare. Until I found the stash in the WE building there was only one of these known to exist. The rest are at the bottom of the ocean. Even the Smithsonian only has a picture of the HQ175. My partner David Navone sold a couple on e-bay for the $1K + range and I have included 2 packs of papers from his research on the tubes history. When I found them I had no clue what they were but I knew they were unusual. They were locked in one of the huge vaults with foot thick doors. The papers on the design and production of the tubes and repeater parts is some of the most fascinating reading I have ever done. For its time this tube was really leading edge technology—quite possibly the most exotic and costly vacuum tube ever built.

Next week I will send you a picture of the Western Electric plant. Imagine a 1.5 million sqft building full of Western Electric paging and sound reinforcement stuff. Lots of ceiling speakers, large tube amps etc. This place has phone trunk lines over a foot in diameter. The GSA said the govt had the property on the books for 82 million dollars. Just think, most of those dollars were spent when a dollar was worth something!! Last project was the Reagan “star wars” project. They also produced the Nike 1 and 2 missiles, the guidance systems for the Minuteman missiles, tanks and airplanes during WW2. Most of the facility was top secret. After I bought it the city fire dept requested an inspection. They had never been inside the fence. The fire chief told a story that occurred years ago when lightning struck one of the towers and started a fire. Several trucks from the city responded but armed guards at the gate met them and turned them away. They told the firemen that their own dept would "take care of it". Now I have my own fire dept complete with chemical trucks and all kinds of paraphernalia. At its peak this facility employed 4000 people and was closed due to military budget cuts by Clinton in 93. It was owned by Redstone Arsenal and operated under contract to Western Electric.

If you ever visit NC I would love to show you around the 35 acre WE facility as well as my newest venture. A few years ago I started a forensic DNA lab. For one last yahoo while I was still young I wanted a business that had three important attributes. I wanted something with: 1-unlimited open ended future 2-good stable profit margin 3-AND THAT COULD NOT BE OUTSOURCED TO CHINA. After much time and thought I felt DNA filled the bill. When we start sending court evidence to China for processing I figure it's so close to the end that only the liberals will be left. We do tests for all kinds of criminal cases for the govt, court system, lawyers etc. We are the independent lab that did the tests for the Duke Lacrosse case that you surely heard about. There are only 13 privately owned ASCLAD accredited DNA labs in the world and we are one of them.

Much thanks for your positive influence and guidance over the years. Enjoy the tube and tech reading !!

RC

Bibliography

Lynn Olson, *The Soul of Sound*, revised and updated in 2005—a superb, accurate, and insightful picture of the greats and ingrates of consumer audio.

American Institute of Electrical Engineers, April 18, 1893. Ms. Mary Ann Hoffman of the IEEE History Center at Rutgers University is in the process of rescuing these seminar papers and getting them to restorers. She graciously provided me with a copy of the original.

Eugene Patronis, pictures of 1930s Western Electric theater loudspeaker.

William T. McQuaide, Audio Engineering Society, picture of Lincoln Walsh.

36 years of SynAudCon Newsletters for pictures, dates, and equipment.

Michael P. Frank, *Physical Limits of Computing*, University of Florida.

The ownership of some 750 audio and acoustic technical volumes now owned and preserved by Mary Gruska, who sensed their metaphysical value and purchased "the infinite with the finite."

Lawrence Lessing, *Man of High Fidelity, Edwin Howard Armstong*, Lippincott, 1956.

Tom Osborne, *Tom Osborne's Story in His Own Words*, a letter to Dave Packard explaining his development of the HP 9100 and the HP 35.

The vast resource of the Internet.

Chapter 2

Subjective Methods for Evaluating Sound Quality

by Ken Pohlmann

2.1 Introduction

At their core, virtually all endeavors in audio engineering share the same goal: To provide optimal sound quality for the conditions. Similarly, formally or not, every audio engineer listens to the audio signal to determine of it meets the design criteria. After the complexities of development and design, the verification of sound quality may appear to be a relatively slight task. But, of course, the problem is far from trivial. In fact, it is difficult to define sound quality and more difficult still to assess it.

It can be argued that the meaning of “good sound” is not well understood. How do we determine that a device sounds good? Who makes that determination? How do we annotate the fact that a $10 loudspeaker might sound relatively good, but its sound is obviously inferior to a $100 loudspeaker? When we must choose between sound qualities, which one is preferable? For example, is a “warm” sound better than a “crisp” sound? What does warm sound like, what does crisp sound like? Is good sound the same for hip-hop music as for classical music? Is it the same for a Chinese listener as for a British listener? Questions such as these are central to the topic of subjective evaluation of sound quality.

2.2 Subjective Evaluation versus Objective Measurements

Objective measurements provide reliable benchmarks for assessing performance outcomes. In many engineering endeavors, objective measurements completely quantify the performance of a device. However, the audio output of most audio devices is ultimately delivered to the human ear. It is difficult or impossible for an objective measure to determine if an audio signal will be pleasing to the ear. It can be argued that objective audio measurements can provide only limited modeling of sound quality. This dilemma forms the basis for subjective evaluations designed to test the aesthetic quality of audio signals.

There is a divide between subjective assessments of sound quality, and objective measurements. Measurements are vitally important, but with audio devices and acoustics, the ear is the final arbiter. In many cases, human subjects can perform irreplaceable analysis. For example, loudness is a human sensory factor, not a literal measure. In a loudness investigation, listeners are presented with various sounds, and each listener is asked to compare the loudness of sound A with that of B. The data is subjected to statistical analysis, and the dependence of loudness upon physical measurements of sound level is assessed. If the test is conducted properly and sufficient listeners are involved, the results are reliable. In this way, we discover, for example, that there is no linear relationship between sound level and loudness, pitch and frequency, or timbre and sound quality.

It is desirable to correlate subjective impressions of listeners with objective design parameters. This allows designers to know where audio fidelity limitations exist and thus know where improvements can be made. Likewise, for example, this knowledge would allow the bit rates of a codec to be lowered while knowing the effects on fidelity. The correlation between the listener's impressions and the objective means to measure the phenomenon is a difficult problem. Correlations are not always known. One way to correlate subjective impressions with objective data is with research—in particular, through critical listening. Over time, it is possible that patterns will emerge that will provide correlation. While correlation is desirable, critical listening plays an important role without it.

There is no instrument to directly measure qualities such as warmth or brilliance. However, in some cases, subjective terms can be related to objective measurements. For example, consider the term "definition" which is generally taken to mean clearness or distinctness in an acoustic space. Arbitrarily, it can be assessed by taking the energy in an echogram during the first 50 to 80ms and comparing it to the energy of the entire echogram. This compares the direct sound and early reflections, which are integrated by the ear, to the entire reverberant sound. This is a straightforward measurement of an impulsive sound from a pistol or other source. This is an example of how a subjective evaluation can be correlated to an objective measure.

In any subjective test, the terminology must be rigorously defined and understood by the administrator and the listening subjects. This provides language for listeners to use in their evaluations, it brings uniformity to the language used by a broad audience of listeners, and it provides a starting point in the objective qualification of subjective comments. Any confusion in terminology would compromise the integrity of the communication between the listener and the administrator. Consider the following descriptive words which are often applied to concert-hall acoustics: warmth, bassiness, definition, reverberance, fullness of tone, liveness, sonority, clarity, brilliance, resonance, blend, intimacy. These terms must be defined using vocabulary that is familiar to the listener.

2.3 Critical Listening

From a subjective standpoint, the optimal way to evaluate an audio signal is to carefully listen to it, using a large number of listeners. This kind of critical listening, when properly analyzed, is the preferred method for subjective sound-quality evaluation. Listening must be blind (sources not known) and use expert listeners, and appropriate statistical analysis must be performed to provide statistically confident results. Critical listening can be very time-consuming and in that respect - costly. In some cases, artifacts are discerned only after repeated listening trials.

Any particular audio device exhibits its own potential anomalies that should be accounted for in a critical listening test. By identifying various common defects, listeners can more closely evaluate performance. The defects must be defined, and in some cases, are specified in the listening test. It is usually helpful to design a listening test according to the type of audio device being tested. For example, the audio fidelity of devices can be considered in four categories:

Large Impairments. Sound-quality differences are readily audible to even untrained listeners. For example, two identical loudspeaker systems, one with normal tweeters and the other with tweeters disabled, would constitute a large impairment.

Medium Impairments. Sound-quality differences are audible to untrained listeners but may require more than casual listening. The ability to readily switch back and forth and directly compare two sources makes these impairments apparent. For example, stereo loudspeakers with a midrange driver wired out of phase would constitute a medium impairment.

Small Impairments. Sound-quality differences are audible to many listeners, however some training and practice may be necessary. For example, the fidelity difference between a music file coded at 128kbps, and 256 kbps, would reveal small impairments in the 128 kbps file. Impairments may be unique and not familiar to the listener, so they are more difficult to detect and take longer to detect.

Micro Impairments. Sound-quality differences are subtle and require patient listening by trained listeners over time. In many cases, the differences are not audible under normal listening conditions, with music played at normal levels. It may be necessary to amplify the music, or use test signals such as low-level sine tones and dithered silence. For example, slightly audible distortion on a –90dBFS 1kHz dithered sine wave would constitute a micro impairment.

When listening to large impairments such as from loudspeakers, sound-quality evaluations can rely on familiar objective measurements and subjective terms to describe differences and find causes for defects. However, when comparing higher-fidelity devices such as high-quality digital-to-analog converters, smaller impairments are considerably more difficult to quantify. It is important to have methodologies to identify, categorize, and describe these subtle differences.

Developing these methodologies requires training of critical listeners, ongoing listening evaluations, and discussions among listeners and designers to close the loop. Further, it is necessary to systematically find thresholds of audibility of various defects. This search introduces known defects and then determines the subjective audibility of the defects and thresholds of audibility.

When differences between devices are small, one approach is to study the residue (difference signal) between them. The analog outputs could be applied to a high-quality analog-to-digital converter; one signal is inverted; the signals are precisely time-aligned with bit accuracy; the signals are added (subtracted because of inversion). Then the residue signal may be studied. For example, if a device adds noise, only the noise is heard in the residue. Two identical signals yield a null file. The process of preparing a residue signal is shown in Fig.2-1.

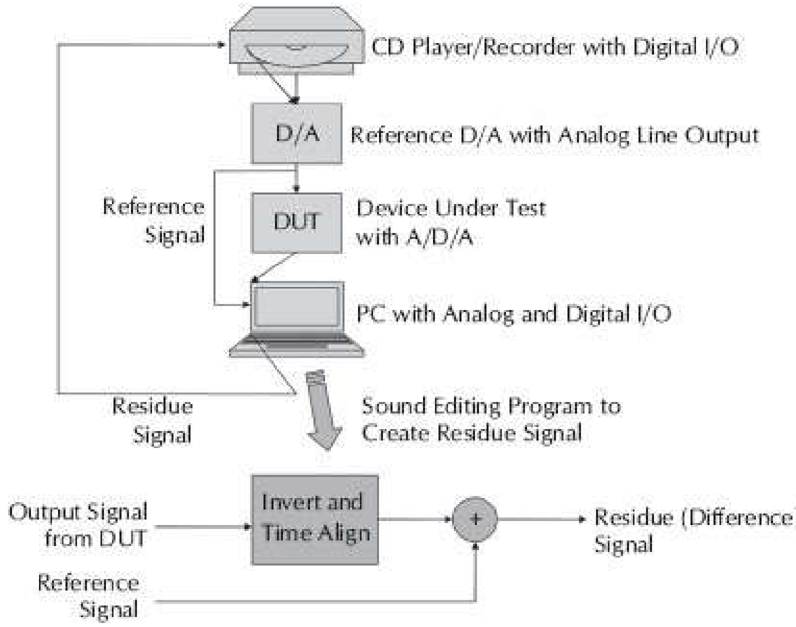


Figure 2-1. Method used to create a residue signal.

Expert listeners are preferred over nonexpert listeners because experts are more familiar with peculiar and subtle artifacts. An expert listener more reliably detects details and impairments that are not noticed by casual listeners. It has been estimated that statistically one experienced listener can replace seven inexperienced ones. Listeners in any test should be trained in the testing procedure and in particular should listen to artifacts that the device under test might exhibit. For example, listeners might start their training with very low bit-rate examples or left-minus-right signals, or residue signals with exposed artifacts so they become familiar with a codec’s signature. When a reference is used, it must be of the highest quality possible. For example, a 16-bit recording may not be an adequate reference when testing high-quality devices because some devices could outperform the reference.

Some listening tests may be conducted using high-quality headphones; external conditions such as loudspeaker characteristics, room acoustics, and speaker and listener position are eliminated. Headphones can allow critical evaluation of subtle audible details. In some cases, open-air headphones (with flat diffuse-field response) are recommended because they provide a more "natural" playback sound; however, a quiet room is needed. Closed-ear headphones can be used to minimize audibility of ambient noise. Other listening tests use loudspeaker playback. For example, this is necessary for multichannel evaluations. When loudspeaker playback is used, room acoustics plays an important role in the evaluation. A proper listening room must provide suitable acoustics and also provide a low-noise environment including isolation from external noise. In some cases, rooms are designed and constructed according to standard reference criteria. Listening room standards are described in Section 2.5.

To yield useful conclusions, the results of a listening test must be subjected to accepted statistical analysis. For example, the ANOVA variance model is often used. Care must be taken to generate a valid analysis that has appropriate statistical significance. The number of listeners, the number of trials, the confidence interval, and other variables can all dramatically affect the validity of the conclusions. In many cases, several listening tests, designed from different perspectives, must be employed and analyzed to fully determine the quality of a device. Statistical analysis is described in Section 2.10.

2.4 Listening Test Methodologies and Standards

Numerous listening-test methodologies and standards have been developed. They can be followed closely or used as practical guidelines to develop other testing. In any case, a listening test must consider source material, signal path, evaluators, methodology, and analysis. Typically the assumption is made that ideal sound quality is equal to transparency. A signal under evaluation is transparent when there is no audible difference between it and the reference. Below this threshold, we can measure the perceptual distance from transparency and describe impairments. Audio quality is thus the subjective distance between an evaluated signal, and a known reference. Any difference is an impairment.

An ideal listening test should meet at least five criteria:

1. Be consistently repeatable with different times, places, and listeners.
2. Consider only the selected audible parameter(s).
3. Reveal the magnitude of differences with a score.
4. Report listener comments.
5. Provide statistically valid results.

Some listening tests can only ascertain whether a device is perceptually transparent; that is, whether expert listeners can tell a difference between the original and the processed file, using test signals and a variety of music. In an ABX test, the listener is presented with the known A and B sources, and an unknown X source that can be either A or B; the assignment is made pseudo-randomly for each trial. The listener must identify whether X has been assigned to A or B. The test answers the question of whether the listener can hear a difference between A and B. ABX testing cannot be used to conclude that there is no difference; rather, it can show that a difference is heard. Short music examples (perhaps 15 to 20 seconds) can be auditioned repeatedly to identify artifacts. It is useful to analyze ABX test subjects individually, and report the number of subjects who heard a difference.

Other listening tests may be used to estimate the margin, or how much the signal can be degraded before transparency is lost. Other tests are designed to gauge relative transparency. This is a more difficult task. If two devices both exhibit audible noise and artifacts, only human subjectivity can determine which device is preferable. Moreover, different listeners may have different preferences in this choice of the lesser of two evils. For example, one listener might be more troubled by bandwidth reduction while another is more annoyed by noise.

Critical listening tests must use double-blind methods in which neither the tester nor the listener knows the identities of the selections. For example, in an “A-B-C triple-stimulus, hidden-reference, double-blind” test the listener is presented with a known A unprocessed reference signal, and two B and C unknown signals. Each stimulus is a recording of perhaps 10 to 15 seconds in duration. One of the unknown signals is identical to the known reference and the other is the coded signal under test. The assignment is made randomly and changes for each trial. The listener must assign a score to both unknown signals, rating them against the known reference. The listener can audition any of the stimuli, with repeated hearings. Trials are repeated, and different stimuli are used. Headphones or loudspeakers can be used. The playback volume level should be fixed in a particular test for more consistent results. The scale shown in Fig. 2-2 can be used for scoring. This five-point impairment scale was devised by the International Radio Consultative Committee (CCIR) and is often used for subjective evaluation. Panels of expert listeners rate the impairments they hear on a 41-point continuous scale in categories from 5.0 (transparent) to 1.0 (very annoying impairments).

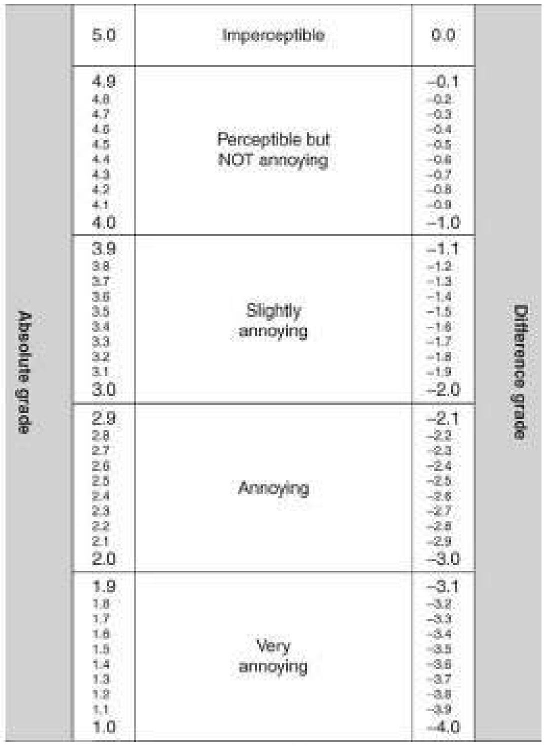


Figure 2-2. Five-point CCIR impairment scale.

The signal selected by the listener as the hidden reference is given a default score of 5.0. Subtracting the score given to the actual hidden reference from the score given to the impaired coded signal yields the subjective difference grade (SDG). For example, original, uncompressed material may receive an averaged score of 4.8 on the scale. If a device obtains an average score of 4.8, the SDG is 0 and the device is said to be transparent (subject to statistical analysis). A lower SDG score (for example, -2.6) assesses how far from transparency a device is. Numerous statistical analysis techniques can be used. Perhaps 50 listeners are needed for good statistical results.

MUSHRA (MUltiple Stimulus with Hidden Reference and Anchors) is an evaluation method used when known impairments exist. This method uses a hidden reference and one or more hidden anchors; an anchor is a stimulus with a known audible limitation.

For example, one of the anchors is a lowpass-coded signal. A continuous scale with five divisions is used to grade the stimuli: excellent, good, fair, poor, and bad. MUSHRA is specified in ITU-R BS.1534. Other issues in sound evaluation are described in ITU-T P.800, P.810, and P.830; ITU-R BS.562-3, BS.644-1, BS.1284, BS.1285, and BS.1286, among others.

In particular, subjective listening tests can be conducted using the ITU-R Recommendation BS.1116-1. This methodology addresses selection of audio materials, performance of playback system, listening environment, assessment of listener expertise, grading scale and methods of data analysis.

2.5 Standard Listening Rooms

Although headphones allow very accurate listening, devices such as loudspeakers must, of course, be auditioned in a room. Any room greatly affects the characteristics of the sound arriving at the listener's ears. To normalize the effects of room acoustics, various standard room designs have been developed.

In addition to listening-test methodology, the ITU-R Recommendation BS.1116-1 standard also describes a reference listening room. The BS.1116-1 specification recommends a floor area of 20 to 60m^2 for monaural and stereo playback and an area of 30 to 70m^2 for multichannel playback. For distribution of low-frequency standing waves, the standard recommends that room dimension ratios meet these three criterion:

$$1.1\left(\frac{w}{h}\right) \leq \frac{l}{h} \leq 4.5\left(\frac{w}{h}\right) - 4 \qquad (2\text{-}1)$$

$$\frac{w}{h} < 3 \tag{2-2}$$

$$\frac{l}{h} < 3 \tag{2-3}$$

where,
l, w, h are the room's length, width and height.

The ⅓ octave sound pressure level, over a range of 50 to 16,000Hz, measured at the listening position with pink noise is defined by a standard response contour. Average room reverberation time is specified to be:

$$T_{ave} = 0.25\left(\frac{V}{V_0}\right)^{1/3} \tag{2-4}$$

where,
T_{ave} is the average reverberation time,
V is the listening room volume,
V_o is a reference volume of $100m^3$.

This reverberation time is further specified to be relatively constant in the frequency range of 200 to 4,000Hz, and to follow allowed variations between 63 and 8,000 Hz. Early boundary reflections in the range of 1,000 to 8,000Hz that arrive at the listening position within 15ms must be attenuated by at least 10dB relative to direct sound from the loudspeakers. It is recommended that the background noise level does not exceed ISO noise rating of NR10, with NR15 as a maximum limit.

The IEC 60268-13 specification (originally IEC 268-13) describes a residential-type listening room for loudspeaker evaluation. The specification is similar to the room described in the BS.1116-1 specification. The 60268-13 specification recommends a floor area of 25 to 40m^2 for monaural and stereo playback and an area of 30 to 45m^2 for multichannel playback. To spatially distribute low-frequency standing waves in the room, the specification recommends three criteria for room-dimension ratios:

$$\frac{w}{h} \leq \frac{l}{h} \leq 4.5\left(\frac{w}{h}\right) - 4 \qquad (2\text{-}5)$$

$$\frac{w}{h} < 3 \qquad (2\text{-}6)$$

$$\frac{l}{h} < 3 \qquad (2\text{-}7)$$

where,
l, w, h are the room's length, width and height.

The reverberation time (measured according to the ISO 3382 standard in ⅓ octave bands with the room unoccupied) is specified to fall within a range of 0.3 to 0.6s in the frequency range of 200 to 4,000Hz. Alternatively, average reverberation time should be 0.4s and fall within a frequency contour given in the standard. The ambient noise level should not exceed NR15 (20–25dBA).

The EBU 3276 standard specifies a listening room with floor area greater than 40m^2 and volume less than 300m^3. Room dimension ratios and reverberation time follow the BS.1116-1 specification. In addition, to mitigate unwanted effects of modal resonances, dimension ratios should differ by more than ±5%. Room response measured as a ⅓ octave response with pink noise follows a standard contour.

2.6 Source Material

To reveal artifacts it is important to use audio materials that stress the device under test. Moreover, because different devices respond differently, a variety of materials is needed, including materials that specifically stress each device. In addition, tracks should be well recorded to avoid introducing artifacts of their own. Generally, music with transient, complex tones, rich in harmonic content around the ear's most sensitive region, 1 to 5kHz, is useful. Particularly challenging examples include glockenspiel, castanets, triangle, harpsichord, tambourine, speech, trumpet, and bass guitar.

Selection of test tracks is usually a highly personal choice. Indeed, when possible, listeners should be encouraged to employ their own music selections. Familiarity with the test tracks is an important aspect of critical listening. It is only after listening to a track over many playback systems, in many rooms, that one gains an understanding of the exact content of its recorded sound. Only then can the track be used as a reference to test a device. Table 2-1 provides examples of audio tracks that have been useful in listening tests.

Table 2-1. Examples of Music Tracks Used for Subjective Listening

Pink Floyd, *Dark Side of the Moon*, "Speak to Me"
Pink Floyd, *Delicate Sound of Thunder*, "On the Turning Away"
KT Tunstall, *Eye to the Telescope*, "Black Horse and the Cherry Tree"
KT Tunstall, *Eye to the Telescope*, "Suddenly I See"
Five For Fighting, *The Battle For Everything*, "100 Years"

Jennifer Warnes, *Famous Blue Raincoat*, “First We Take Manhattan”
Dire Straits, *The Very Best of Dire Straits*, “Sultans of Swing”
Dire Straits, *On Every Street*, “Calling Elvis”
Dire Straits, *Brothers in Arms*, “Brothers in Arms”
Madonna, *Ray of Light*, “Candy Perfume Girl”
Santana, *Supernatural*, “Migra”
Blur, *13*, “Tender”
Gomez, *Liquid Skin*, “Hangover”
The Mavericks, *Trampoline*, “Dream River”
Alison Krauss, *New Favorite*, “New Favorite”
Steely Dan, *Gaucho*, “Babylon Sisters”

It is sometimes useful to point out particular aspects of a music track to a listener. For example, on the Jennifer Warnes track “First We Take Manhattan,” a listening guide might note:

- Crispness of instruments.
- Transient response of drum kit.
- Enunciation and intimacy of vocal line.
- Natural sibilance.
- Smooth lead guitar, note echo effect.
- Inner voices such as organ and woodblock.
- Reverberation on voice.
- Bright mix without overclarity.
- Natural compression.
- Mix balance at low levels.
- Overall high fidelity.

On the Pink Floyd track “On the Turning Away,” some musical cues are:

- Smooth bass response.

- Clarity and presence in vocals.
- Audibility of reverberation, echo, and doubling effects on voice.
- Clean acoustic guitar with detail of instrument.
- Clean keyboards.
- Crisp tambourine.
- Strong kick drum.
- Toms big but not boomy.
- Prominent lead and rhythm guitar.

When providing listening guidance such as this, as noted, it is important that listeners understand the meaning of the terms being employed. Differences in usage will clearly invalidate the results of any listening session.

As an alternative to music content, it is sometimes useful to employ nonmusic content. This content is selected because of specific acoustic characteristics. Because these tracks are usually less complex than music, and perhaps of shorter duration, they may be easier to use. Some examples include: reverberation tails, sine waves, correlated and uncorrelated pink noise, male and female speech, castanets, and sampled piano chromatic scale (consistent level across the audio band).

2.7 Evaluation of Perceptual Codecs

Perceptual codecs offer unique opportunities and challenges to subjective testing. Perceptual codecs are designed to take advantage of weaknesses in the human auditory system, and to avoid detection of their degraded transparency. At very low bit rates, a codec cannot be designed to be transparent, so listening tests are needed to optimize sound quality by minimizing audible artifacts.

When a codec is not transparent, artifacts such as changes in timbre, bursts of noise, granular ambient sound, shifting in stereo imaging, and spatially unmasked noise can be used to identify the "signature" of the device. Bandwidth reduction is also readily apparent, but a constant bandwidth reduction is less noticeable than a continually changing bandwidth. Changes in high-frequency content in successive transform blocks can create audible artifacts.

Speech is often a difficult codec test signal because its coding requires high resolution in both time and frequency. With low bit rates or long transform windows, coded speech can assume a strangely reverberant quality. A stereo or surround recording of audience applause sometimes reveals spatial coding errors. Subband codecs can have unmasked quantization noise that appears as a burst of noise in a processing block. In transform codecs, errors are reconstructed as basis functions (for example, a windowed cosine) of the codec's transform. A codec with a long block length can exhibit a pre-echo burst of noise just before a transient, or there might be a tinkling sound or a softened attack. Transform codec artifacts tend to be more audible at high frequencies. Changes in high-frequency bit allocation can result in a swirling sound due to changes in high-frequency timbre.

Within a codec type, for example, with MP3 codecs, the MPEG standard dictates that all compliant decoders should perform identically and sound the same. It is encoders that more likely introduce sonic differences. However, some decoders may not properly implement the MPEG standard and thus are not compliant. For example, they may not support intensity stereo coding or variable rate bitstream decoding. MP3 encoders, for example, can differ significantly in audio performance depending on the psychoacoustic model, tuning of the nested iteration loops, and strategy for switching between long and short windows. Another factor is the joint-stereo coding method and how it is optimized for a particular number of channels, audio bandwidth and bit rate. Many codecs have a range of optimal bit rates; quality does not improve significantly above those rates, and quality can decrease dramatically below them. An understanding of the design properties of devices such as codecs can help focus the scope of a listening test, thus improving its effectiveness.

2.8 Evaluation of Speech Codecs

The performance of speech coding systems such as used in cell phones can be assessed through subjective listening tests. For example, the mean opinion score (MOS) is often used. On this five-point scale, a score of 4.0 or higher denotes high quality or near-transparent performance. Scores between 3.5 and 4.0 are often permissible for speech communications systems. Scores below 3.0 denote low quality. For example, a signal may be highly intelligible, but the speech may sound synthetic or otherwise unnatural. A circuit-merit (CM) quality scale can be used to rate subjective performance. The average of CM scores obtained from listeners provides a mean opinion score. The ranking of CM scores is shown in Table 2-2.

Table 2-2 Ranking of CM scores for subjective performance of speech codecs

CM5 Excellent	Speech perfectly understandable
CM4 Good	Speech easily understandable, some noise
CM3 Fair	Speech understandable with a slight effort, occasional repetitions needed

CM2	Poor	Speech understandable only with considerable effort, frequent repetitions needed
CM1	Unsatisfactory	Speech not understandable

The diagnostic rhyme test (DRT) is a standardized subjective method used to evaluate speech intelligibility. In this ANSI test, listeners are presented with word pairs. They must choose which word they perceive. The words are different only in their leading consonants and pairs are selected so that six attributes of speech intelligibility are evaluated in their present or absent state. The diagnostic acceptability measure (DAM) test is also used to evaluate speech codec performance.

Historically, telephone systems provided a narrow bandwidth from approximately 300 to 3,400Hz. It is generally agreed that the lower cutoff frequency, although it causes a “thin” sound because of missing lower frequencies and thus decreases sound quality, does not greatly diminish speech intelligibility. It has been reported that the upper cutoff frequency of 3,400Hz allows 97% of speech sounds to be understood. Recognition of sentences can achieve 99% intelligibility because of assistance by context. Newer wideband telephone speech systems provide response from 50Hz to 7kHz. Although the improvement in measured intelligibility is slight, these systems provide a more natural speech sound with clearer sound quality.

2.9 Evaluation of Speech Intelligibility in Rooms

Speech intelligibility is the highest design priority for any acoustic space intended for spoken word. This is the case in many places of worship, auditoriums, and drama theaters. In rooms where amplification is not used, the acoustical elements of the room design must be carefully considered to provide a high level of speech intelligibility. Sound systems are often used to overcome acoustical limitations, and to provide intelligibility in even very large spaces.

Speech intelligibility in a room is often estimated using subject-based measures, that is, by using live experimentation. A talker reads from a list of words and phrases, and listeners in the room write down what they hear. The list includes examples of important speech sounds. A test may use between 200 and 1,000 words. For example, some English words used in subject-based intelligibility testing include: aisle, barb, barge, bark, baste, bead, beige, boil, choke, chore, cod, and coil.

The higher the percentage of correctly understood words and phrases, the higher the speech intelligibility. In some cases, listening difficulty is measured. When the level of speech is the same as the noise level, intelligibility can be high, but listeners can still have difficulty in understanding what is being said, and considerable attention is required. When the speech level is raised by 5 or 10dB over the noise, intelligibility is not greatly improved, but listeners report much less difficulty in hearing.

2.10 Listening Test Statistical Evaluation

To be meaningful, interpretation of listening test results must be carefully considered. For example, in an ABX test, if a listener correctly identifies the reference in 12 out of 16 trials, has an audible difference been noted? Statistic analysis is used to interpret the results. Because the test is a sampling, we define our results in terms of probability. The larger the sampling, the more reliable the result. A central concern is the significance of the results. If the results are significant, they are due to audible differences. Otherwise they are due to chance. In an ABX test, a correct score 8 of 16 times indicates that the listener has not heard differences; the score could be arrived at by guessing. A score of 12/16 might indicate an audible difference, but could also be due to chance. To assess this, we can define a null hypothesis H_0 that holds that the result is due to chance, and an alternate hypothesis H_1 that holds it is due to an audible difference. The significance level α is the probability that the score is due to chance. The criterion of significance α' is the chosen threshold of α that will be accepted. If α is less than or equal to α' then we accept that the probability is high enough to accept the hypothesis that the score is due to an audible difference. The selection of α' is arbitrary but a value of 0.05 is often used. This can be analyzed with binomial distribution using the equation:

$$z = \frac{c - 0.5 - np_1}{[np_1(1 - p_1)]^{1/2}} \qquad (2\text{-}8)$$

where,
z is the standard normal deviate,
c is the number of correct responses,
n is the sample size,
p_1 is the proportion of correct responses in a population due to chance alone ($p_1 = 0.5$ in an ABX test).

With a score of 12/16, $z = 1.75$. Binomial distribution thus yields a significance level of 0.038. The probability of getting a score as high as 12/16 from chance alone (and not from audible differences) is 3.8%. In other words, there is a 3.8% chance that the listener did not hear a difference. However, since α is less than α' ($0.038 < 0.05$) we conclude that the result is significant and there is an audible difference, at least according to how we have selected our criterion of significance. If α' is selected to be 0.01, then the same score of 12/16 is not significant and we would conclude that the score is due to chance.

We can define parameters that characterize the risk that we are wrong in accepting a hypothesis. A Type 1 error risk (also often noted as α') is the risk of rejecting the null hypothesis when it is actually true. Its value is determined by the criterion of significance; if $\alpha' = 0.05$ then we will be wrong 5% of the time in assuming significant results. Type 2 error risk β defines the risk of accepting the null hypothesis when it is false. Type 2 risk is based on the sample size, value of α', the value of a chance score, and effect size or the smallest score that is meaningful. These values can be used to calculate sample size using the equation:

$$n = \left\{ \frac{z_1[p_1(1-p_1)]^{1/2} + z_2[p_2(1-p_2)]^{1/2}}{p_2 - p_1} \right\}^2 \qquad (2\text{-}9)$$

where,

n is the sample size,

p_1 is the proportion of correct responses in a population due to chance alone ($p_1 = 0.5$ in an ABX test),

p_2 is the effect size: hypothesized proportion of correct responses in a population due to audible differences,

z_1 is the binomial distribution value corresponding to Type 1 error risk,

z_2 is the binomial distribution value corresponding to Type 2 error risk.

For example, in an ABX test, if Type 1 risk is 0.05, Type 2 risk is 0.10, and effect size is 0.70, then the sample size should be 50 trials. The smaller the sample size, that is, the number of trials, the greater the error risks. For example, if 32 trials are conducted, $\alpha' = 0.05$, and effect size is 0.70. To achieve a statistically significance result, a score of 22/32 is needed.

Binomial distribution analysis provides good results when a large number of samples are available. Other types of statistical analysis such as signal detection theory can also be applied to ABX and other subjective testing. Statistical analysis can appear impressive, but its results cannot validate a test that is inherently flawed.

2.11 Example of a Testing Procedure and Evaluation Form

A listening test, prepared by the author, uses paired comparisons to evaluate small impairments. The test compares a device under test (DUT) to a reference, or one DUT against another DUT. In this example, headphones are used for playback. An example of the evaluation signal path is shown in Fig. 2-3. This test accomplishes three things:

1. It determines if a listener can hear a difference between the source pairs; if so, the listener is vetted.
2. It asks the vetted listener for his or her subjective comments.
3. It also asks the vetted listener to score the sources using a

numerical scale.

The test is in two parts. The evaluation form presented to listeners is shown in Fig. 2-4. Part One is an ABX test; it determines whether the listener can hear a difference between two sources. The test is double-blind, using an ABX box. For example, if a listener is correct 12 out of 16 times, we are 95% confident that they heard a difference (the listener is said to be vetted). We can improve the confidence by increasing the threshold (for example, 13 or 14), or by increasing the number of trials (for example, 20, 50, 100). If a listener cannot hear any difference, they are excused. Their opinions are not useful.

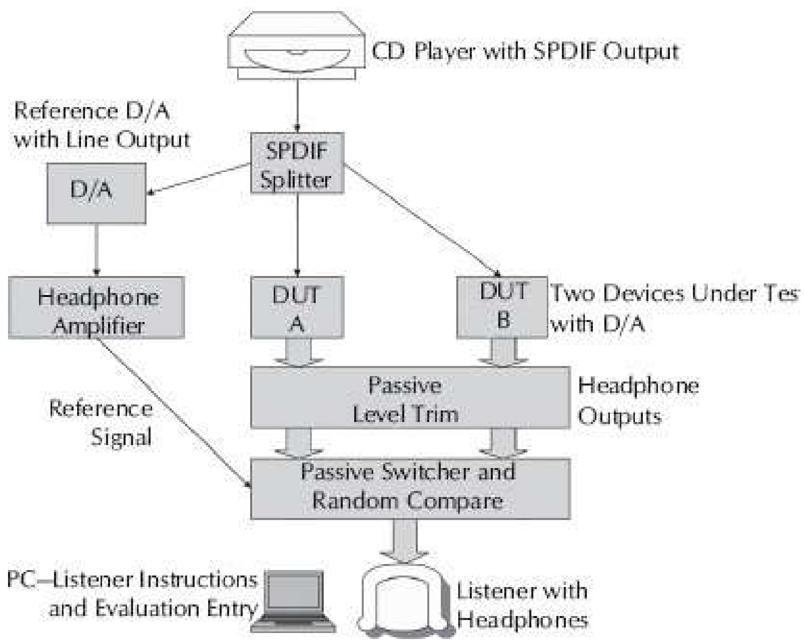


Figure 2-3. Signal path used for a subjective listening evaluation.

Part Two is a non-blind evaluation. The vetted listener writes comments describing their subjective evaluation. The Form points out multiple areas to listen to (tonal balance, dynamic range, etc). The listener is also asked to provide a numerical score.

The criteria in Part Two cover a wide range of listening parameters. They are selected for headphone listening and omit some traditional parameters associated with the effects of loudspeakers in rooms. The criteria are comprehensive, but in many cases the listener may focus on a few criteria (such as low-level detail or audible artifacts) and not have any meaningful comments on other areas.

In Part Two, the paired comparison can be a DUT and a reference, or two DUTs. In the first case, the listener does not have to write comments about the reference. In the second case, he or she must write comments about both DUTs. The reason for commenting is to help designers understand what the listener is hearing, and possibly objecting to.

In Part Two, either the reference or DUT A is arbitrarily given a score of 0. The other source is scored relative to that on a ±10 point scale for each criteria. This allows for DUTs that sound worse, or better than the reference. As noted, two DUTs can be scored at once; this is essentially a time saver. If possible, it is recommended that each DUT be evaluated and scored against the reference. This provides a more stable score.

In some other tests, essentially, Part One and Part Two are merged. The listener must evaluate and score blindly. This is not recommended. It places more burden on the listener and may detract from the quality of the comments and scores. It also leads to practical problems. If it is revealed that a listener did not hear a difference on a trial, then those comments can be omitted. But if the listener did hear a difference on a trial, can the comments be trusted? What if, on that particular trial, the listener was correct because of chance? It is better to have two parts, in which it is determined, a priori, within statistical error, that the listener is hearing a difference and thus their comments are valid.

In this test, the listener uses a file containing music and non-music tracks. The listener is free to select any tracks for the ABX test and for evaluating any of the criteria. Some other tests assign specific tracks to specific criteria. But listeners are very individualistic in how they listen. Assigning specific tracks unnecessarily limits their ability to hear details. However, when writing their comments, listeners should be asked to provide comments on which tracks they are listening to, in case this is important to subsequent analysis.

Ideally, a preference listening evaluation such as this would use 10 to 20 expert listeners; this consensus should provide accurate results. In practice, this is difficult to do. A still-valid approach is to use 1 or 2 expert listeners in a double-blind test. To facilitate this, the test removes listener bias, verifies that audible differences are being heard, and then trusts in the scoring of a few expert listeners.

After preliminary listening, a compilation file using these (and other) tracks will be prepared. The form supplied to listeners also provides definitions of some common subjective listening terms. As noted, unless the meanings of the evaluation terms are well understood, subjective results will not be meaningful.

Figure 2-4. Example of a Form Supplied to Listeners Performing Evaluation in a Small Impairments Listening Test.

This section is an example of an evaluation form supplied to listeners performing a subjective evaluation.

Listener's Name:
Audio Source:
Audio Source:
Headphones Used:
Today's Date:

Listening Test – Part One

Instructions to listener: Use the ABX box to compare two audio sources, using headphones. One source is A, and the other source is B. In each trial, A or B is randomly assigned to X. Your task is to listen to A, B, and X and determine whether A is assigned to X, or if B is assigned to X. You may listen to A, B and X as many times as you want. You may use any music tracks that you want. You may write down any comments as well as which music track is most helpful if it helps you complete the test; these comments will not be graded. Then write down whether X is A or B for that trial. Please complete 16 trials using this form:

Trial Number	Comments	Music Notes	Is X assigned to A or B?
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
Your Score for this Test: / 16			

The ABX box will now reveal the correct answers in the Answer mode. If you scored correctly 12 or more correct answers, congratulations! You are probably hearing a difference between A and B. You may proceed to Part Two. If you scored less than 12, you may not be hearing a difference. You can reset the ABX box and try again, listening more closely to small details. The ABX box will automatically create new assignments.

Listening Test – Part Two

Instructions to listener: Use the ABX box in the Answer mode to switch between A and B. There is only one trial, the assignments of A and B do not change, and you will see which source you are listening to. In Part Two, you are not being scored. Rather, you are scoring the sound quality of sources A and B. Your task is to listen to A and B as many times as you want. You may use any music tracks that you want.

Source A may be a device under test, or a known reference. You are comparing this to source B. Please listen to the different Sound Criteria, and write comments and notate music tracks for both sources. The definitions of the Sound Criterion terms are listed below. Source A or the reference always has a 0 (zero or neutral) score. Compare source B to this 0 benchmark and then score the sound quality of B. Your grade will be an integer between −10 (much worse) and +10 (much better). For example, if the Fidelity of B is just slightly better than A, then its Fidelity score will be +1. Please use this scoring scale:

−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
much worse				somewhat worse						equal			somewhat better							much better

Before beginning the evaluation, it is important that you understand the meanings of the sound quality criteria terms (such as tonal balance, dynamic range, etc). The terms are defined on the form. These terms might have a meaning that is different from the meaning you are accustomed to. It is important that you use the terms as they are defined here.

SOURCE A or REFERENCE				SOURCE B	
Sound Criteria	Comments and Music Notes	Score 0		Comments and Music Notes	Score ±10
Tonal Balance		0			
Dynamic Range		0			
Bass Clarity		0			
Midrange Clarity		0			
Treble Clarity		0			
Imaging		0			
Sound-stage		0			
Low-Level Detail		0			
Audible Artifacts		0			
Total Fidelity		0			
				Score for Source B:	

Sound Criteria Definitions

Tonal Balance. (spectral balance, octave balance). A playback device that is tonally balanced has a consistent frequency response that ideally is flat across the audio band. No frequency area is dominant, or lacking. Sound that has poor tonal balance may have peaks or dips in frequency response, and this may exaggerate or obscure certain notes or change timbre. Bass frequencies should blend naturally with higher frequencies. Correct tonal balance also requires good low-frequency and high-frequency extension, at the bottom and top of the spectrum.

Dynamic Range. Sound should have natural travel across the loudness spectrum from soft to loud, without strain. At loud levels, the sound should not tighten up or sound compressed with a decrease in recorded dynamic range. There should be no spectral shifts at different loudness levels. There should be no dynamic pumping, such as a quick decrease in mid frequencies when a bass note is attacked. At maximum loudness, when playing a loud musical passage, the sound should still be clear, unclipped and free of excess distortion at all frequencies. At loud levels, listen for low frequencies that become harder or sharper in timbre; for example, the sound of a kick drum can change. Listen for dynamic spectral shift, any differences in timbre or spectral balance that change from soft volume to loud volume. At loud levels, this may be caused by compression in one or several frequency ranges. At soft levels, the sound should maintain a musically appropriate spectral balance. This is not related to loudness compensation; loudness compensation is related to playback volume, not to loudness and dynamics of the musical material. Maximum loudness should be adequately loud for satisfactory playback.

Bass Clarity. (bass clearness, definition, distinctness). Sound has good clarity when you can hear musical details in the bass region. Instruments such as drums and bass guitar should have a natural sound quality that is solid and tight without softening. Transients have a clean attack with accurate bass punch. When bass sound has poor clarity, it may be muddy, boomy, boxy, weak, or thin.

Midrange Clarity. (midrange clearness, definition, distinctness). Sound has good clarity when you can hear musical details in the midrange region. Voices and Instruments should have a natural and smooth sound quality. Voices should be intelligible. Transients should be clean with a tight attack. When midrange sound has poor clarity, voices may sound sibilant, chesty or nasal. Instruments may sound muddy or noisy. You may hear changes to the spectrum such as coloration, and ringing, or a honking sound. There may be blurring of midrange frequencies caused by intermodulation of low frequencies. There may be intermodulation when many voices, such as a choir, are sung. There may be a harsh glare sound at upper midrange frequencies. Listen for any time distortion such as comb filtering that can cause changes in timbre or a hollow sound.

Treble Clarity. (treble clearness, definition, distinctness). Sound has good clarity when you can hear musical details in the treble region. Instruments should have a natural and clean sound with a sense of "air" around them. Transients should be crisp, clean, and delineated. When treble sound is poor, instruments may sound noisy, sharp, too bright, too dull, or shrill. There may be a harsh glare sound at high frequencies.

Imaging. When stereo is played, there should be a consistent spread of images across the stereo panorama. The panorama should reach continuously from one channel to the other without gaps. The images should usually be stable and accurately localized across the panorama, and not too diffuse. Images should be distinct and distinctly separated as individual voices; listen for the "black space between instruments." There should be accurate front-to-back depth to a recording; to do this, the system must accurately reproduce the ratio of direct (first arriving) sound, to reverberant sound. Poor imaging may result in localization of an instrument in more than one place, or high and low frequencies coming from different places. Localization should not change with changes in pitch, loudness or timbre. Movement of a listener's head or change in seating should not greatly affect imaging. Sound localized to a subwoofer is another example of poor imaging. A monaural sound would have a center-image, but no others, so its image breadth is very poor.

Soundstage (spaciousness, openness, ambience). A stereo soundstage should usually be open, with width and breadth and front/back depth to it. It has fullness; it symmetrically fills the musical space and has the spatial presence and natural room acoustics geometry of music performed on a real stage. Liveness and presence should envelop you spatially giving a "you are there" feeling. Rather than listening to a reverberant environment, you are in the environment. Ambience should be diffuse and coming from all directions, and not localized.

Ambience should have spectral balance; for example, reverberation should not be too thin, or too boomy. The space is consistent with the space of the recording; you should be able to discern differences in the size and acoustical response of different recorded environments. A soundstage that is not spacious may be closed and narrow, it may be flat and lack presence. A poor soundstage may have odd phase artifacts with a phasey sound with inconsistent directional effects, and spectral changes resulting from small changes in listener head position.

Low-Level Detail. Very soft sounds should be cleanly reproduced. Natural ambient sounds, noise, and room tone should all be audible and not obscured or masked by other introduced sounds. Reverberation tails should cleanly and smoothly decay to silence. The device should not add any other soft sounds, nonlinearities, granulation noise or other noises. Listen carefully for any noise that is dependent on the audio signal.

Audible Artifacts. (Other extraneous sounds). A playback device should be clean and only convey the sounds in the original recording. Ideally, there should be no added sounds at all, especially sounds that are independent of the audio signal. Extraneous sounds might include noise floor, other noises, hum, buzz, clicks, pops and other artifacts. Make sure that some odd sound you hear isn't on the original recording. There should be no added sounds or timbre changes especially at high frequencies.

Total Fidelity. This criterion summarizes your impression and preference. A device with good fidelity is transparent, neither adding nor subtracting from the original recorded sound. The sound should have no perceptual relationship to the playback device. This measures the overall accuracy and pleasantness of the sound.

2.12 Conclusion

At some future time, it is possible that objective measurements will be developed that will comprehensively evaluate the aesthetic performance of audio signals. Such measurements would model the processing performed by the human ear, eliminating the need for human listeners. Parameters such as clarity, fullness, imaging, and soundstage would be quantified, or perhaps discarded in favor of better metrics. Perhaps the system could further allow engineers to tune the system to provide unique aesthetic results such as a warm sound. However, until that objective measuring system is developed, subjective listening will provide valuable input in the development of audio devices. Although time consuming, it is appropriate that human listeners shape the signals that will be appreciated by their peers.

Bibliography

L. Allard, M. Rossi, M.A. Zufferey, “Assessment of the Pertinence of Perceived Attributes for Listening Tests,” *Audio Engineering Society* Preprint 5089; Convention 108; February 2000.

R.J. Beaton, J.G. Beerends, M. Keyhl, W.C. Treurniet, “Objective Perceptual Measurement of Audio Quality,” *Audio Engineering Society* Paper RED-012; Conference: Collected Papers on Digital Audio Bit-Rate Reduction; June 1996.

R.J. Beaton, P. Wong, “A Disk-Based System for the Subjective Assessment of High-Quality Audio,” *Audio Engineering Society* Preprint 3497; Convention 94; March 1993.

S. Bech, “The Influence of Room Acoustics on Reproduced Sound, Part 1: Selection and Training of Subjects for Listening Tests,” *Audio Engineering Society* Preprint 2850; Convention 87; October 1989.

S. Bech, “Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data,” *Audio Engineering Society* Paper 8-014; Conference: AES 8th International Conference: The Sound of Audio; May 1990.

S. Bech, “Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment,” *Journal of the Audio Engineering Society*, Vol., 40 No. 7/8 pp. 590·610; July 1992.

M. Burkhard, K. Genuit, “Merging Subjective and Objective Acoustical Measurements,” *Audio Engineering Society* Paper 11-015; Conference: AES 11th International Conference: Test & Measurement; May 1992.

H. Burstein, “Approximation Formulas for Error Risk and Sample Size in ABX Testing,” *Journal of the Audio Engineering Society*, Vol. 36, No.11, pp. 879–883; November 1988.

D. Clark, “High-Resolution Subjective Testing Using a Double-Blind Comparator,” *Journal of the Audio Engineering Society*, Vol. 30, No. 5, pp. 330–338; May 1982.

D. Clark, Listening Technology for Automotive Sound Systems, *Audio Engineering Society* Preprint 5723; Convention 114; March 2003.

C. Colomes, C. Schmidmer, T. Thiede, W.C. Treurniet, “Perceptual Quality Assessment for Digital Audio: PEAQ: The New ITU Standard for Objective Measurement of the Perceived Audio Quality,” *Audio Engineering Society* Paper 17-033; Conference: AES 17th International Conference: High-Quality Audio Coding; September 1999.

D. Daniels, “The Futility of Quantifying Subjective Audio Measurements,” *Audio Engineering Society* Preprint 4379; Convention 101; November 1996.

F.A. Everest, “Instruction in Critical Listening,” *Audio Engineering Society* Preprint 1890; Convention 72; October 1982.

F.A. Everest, and K. Pohlmann, *Master Handbook of Acoustics*, Fifth Edition., New York, NY: McGraw-Hill, 2009.

A. Furmann, E. Hojan, M. Niewiarowicz, P. Perz, “On the Correlation between the

Subjective Evaluation of Sound and the Objective Evaluation of Acoustic Parameters for a Selected Source," *Journal of the Audio Engineering Society*, Vol. 38, No. 11, pp. 837–844; November 1990.

A. Gabrielsson, B. Lindström, "Perceived Sound Quality of High Fidelity Loudspeakers," *Audio Engineering Society* Preprint 2010; Convention 74; October 1983.

A. Gabrielsson, B. Lindström, "Perceived Sound Quality of High-Fidelity Loudspeakers," *Journal of the Audio Engineering Society*, Vol.33, No. 1/2, pp. 33–53; January 1985.

C. Grewin, "Can Objective Methods Replace Subjective Assessments?," *Audio Engineering Society* Preprint 4067; Convention 99; October 1995.

C. Grewin, S. Bergman, O. Kejving, "A Listening Test System for Evaluation of Audio Equipment," *Audio Engineering Society* Preprint 2335; Convention 80; March 1986.

C. Grewin, T. Rydén, "Subjective Assessments on Low Bit-Rate Audio Codecs," *Audio Engineering Society* Paper 10-013; Conference: AES 10th International Conference: Images of Audio; September 1991.

T. Grusec, S. Lyman, A Sensitive Method for the Subjective Evaluation of Audio Codecs, *Audio Engineering Society* Paper 11-026; Conference: AES 11th International Conference: Test & Measurement; May 1992.

T. Grusec, L. Thibault, G. Soulodre, "Subjective Evaluation of High-Quality Audio Coding Systems: Methods and Results in the Two-Channel Case," *Audio Engineering Society* Preprint 4065; Convention 99; October 1995.

R. Harley, "The Role of Critical Listening in Evaluating Audio Equipment Quality," *Audio Engineering Society* Preprint 3176; Convention 91; October 1991.

F.K. Harvey, M.R. Schroeder, "Subjective Evaluation of Factors Affecting 2-Channel Stereophony," *Audio Engineering Society* Preprint 136; Convention 12; October 1960.

K.E. Heber, "Test Signals for Subjective Vehicle Audio Evaluation," *Audio Engineering Society* Preprint 4111; Convention 99; October 1995.

T. Hirvonen, M. Vaalgamaa, J. Backman, M. Karjalainen, "Listening Test Methodology for Headphone Evaluation," *Audio Engineering Society* Preprint 5736; Convention 114; March 2003.

W. Hoeg, "Listening Conditions for Subjective Assessment of Sound Quality: The Status of International Standardization," *Audio Engineering Society* Preprint 3788; Convention 96; February 1994.

A. Komly, "Assessing the Performance of Low-Bit-Rate Audio Codecs: An Account of the Work of ITU-R Task Group 10/2," *Audio Engineering Society* Paper RED-010; Conference: Collected Papers on Digital Audio Bit-Rate Reduction; June 1996.

L. Leventhal, "How Conventional Statistical Analyses Can Prevent Finding Audible Differences in Listening Tests," *Audio Engineering Society* Preprint 2275; Convention 79; October 198.

M. Link, "Internet Audio Quality and the Mushra Test," *Audio Engineering Society* Paper 17; Conference: 17th UK Conference; April 2002.

S.P. Lipshitz, J. Vanderkooy, “The Great Debate: Subjective Evaluation,” *Journal of the Audio Engineering Society*, Vol. 29, No. 7/8, pp. 482–491; July 1981.

S.E. Olive, “A Method for Training Listeners and Selecting Program Material for Listening Tests,” *Audio Engineering Society* Preprint 3893; Convention 97; November 1994.

J.M. Risch, “A User-Friendly Methodology for Subjective Listening Tests,” *Audio Engineering Society* Preprint 3178; Convention 91; October 1991.

E.H. Rothauser, G. Urbanek, “Some Problems in Subjective Testing,” *Audio Engineering Society* Preprint 488; Convention 31; October 1966.

S. Shlien, G. Soulodre, “Measuring the Characteristics of “Expert” Listeners,” *Audio Engineering Society* Preprint 4339; Convention 101; November 1996.

G. Soulodre, M. Lavoie, “Subjective Evaluation of Large and Small Impairments in Audio Codecs,” *Audio Engineering Society* Paper 17-032; Conference: AES 17th International Conference: High-Quality Audio Coding; September 1999.

T. Sporer, “Evaluating Small Impairments with the Mean Opinion Scale: Reliable or Just a Guess?,” *Audio Engineering Society* Preprint 4396; Convention 101; November 1996.

H. Staffeldt, “Correlation Between Subjective and Objective Data for Quality Loudspeakers,” *JAES* Volume 22 Number 6 pp. 402·415; July 1974.

T. Thiede, et al, “PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality,” *Journal of the Audio Engineering Society*, Vol. 48, No. 1/2, pp. 3–29; January 2000.

F.E. Toole, “Subjective Measurements of Loudspeaker Sound Quality and Listener Performance,” *Journal of the Audio Engineering Society*, Vol. 33, No. 1/2, pp. 2–32; January 1985.

F.E. Toole, “Listening Tests: Turning Opinion into Fact,” *Audio Engineering Society* Preprint 1766; Convention 69; May 1981.

F.E. Toole, “Subjective Evaluation: Identifying and Controlling the Variables,” *Audio Engineering Society* Paper 8-013; Conference: AES 8th International Conference: The Sound of Audio; May 1990.

Toole, Floyd E., “Subjective Measurements of Loudspeaker Sound Quality,” *Audio Engineering Society* Preprint 1900; Convention 72; October 1982.

F.E. Toole, S.E. Olive, “Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things,” *Audio Engineering Society* Preprint 3894; Convention 97; November 1994.

W.C. Treurniet, G.A. Soulodre, “Evaluation of the ITU-R Objective Audio Quality Measurement Method,” *Journal of the Audio Engineering Society*, Vol. 48, No. 3, pp. 164–173; March 2000.

P. Wilson, “A Repeatable Technique for Listening Tests,” *Audio Engineering Society* Preprint 376; Convention 17; October 1965.

W. Woszczyk, S. Bech, V. Hansen, “Interaction Between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes,” *Audio Engineering Society*

Preprint 4133; Convention 99; October 1995.

Chapter 3

Psychoacoustics

by Dr. Peter Xinya Zhang

3.1 Psychoacoustics and Subjective Quantities

Unlike other senses, it is surprising how limited our vocabulary is when talking about hearing.[1] Especially in the audio industry, we do not often discriminate between subjective and objective quantities. For instance, the quantities of frequency, level, spectrum, etc. are all objective, in a sense that they can be measured with a meter or an electronic device; whereas the concepts of pitch, loudness, timbre, etc. are subjective, and they are auditory perceptions in our heads. Psychoacoustics investigates these subjective quantities (i.e., our perception of hearing), and their relationship with the objective quantities in acoustics. *Psychoacoustics* got its name from a field within psychology, i.e., recognition science, which deals with all kinds of human perceptions, and it is an interdisciplinary field of many areas, including acoustics, psychology, behavioral science, engineering, computer science, physics, biology, medical science, audiology, speech science, etc.

Although there are clear and strong relationships between certain subjective and objective quantities, e.g., pitch versus frequency, other objective quantities also have influences. For example, changes in sound level can affect pitch perception. Furthermore, because no two persons are identical, when dealing with perceptions as in psychoacoustics, there are large individual differences, which can be critical in areas such as sound localization.[2] In psychoacoustics, researchers have to consider both average performances among population and individual variations. Therefore, psychophysical experiments and statistical methods are widely used in this field.

Compared with other fields in acoustics, psychoacoustics is relatively new, and has been developing greatly. Although many of the effects have been known for some time (e.g., Hass effect[3]), new discoveries have been found continuously. To account for these effects, models have been proposed. New experimental findings might invalidate or modify old models, or make certain models more or less popular. This process is just one representation of how we develop our knowledge. For the purpose of this handbook, we will focus on summarizing the known psychoacoustic effects rather than discussing the developing models.

3.2 Ear Anatomy and Function

Before discussing various psychoacoustic effects, it is necessary to introduce the physiological bases of those effects, namely the structure and function of our auditory system. The human ear is commonly considered in three parts: the outer ear, the middle ear, and the inner ear. The sound is gathered (and as we shall see later, modified) by the external ear called the *pinna* and directed down the *ear canal* (auditory meatus). This canal is terminated by the *tympanic membrane* (eardrum). These parts constitute the outer ear, as shown in Figs. 3-1 and 3-2. The other side of the eardrum faces the middle ear. The middle ear is air filled, and pressure equalization takes place through the *eustachian tube* opening into the *pharynx* so normal atmospheric pressure is maintained on both sides of the eardrum. Fastened to the eardrum is one of the three ossicles, the *malleus*, which, in turn, is connected to the *incus* and *stapes*. Through the rocking action of these three tiny bones the vibrations of the eardrum are transmitted to the *oval window* of the *cochlea* with admirable efficiency. The sound pressure in the liquid of the cochlea is increased some 30–40dB over the air pressure acting on the eardrum through the mechanical action of this remarkable middle ear system. The clear liquid filling the cochlea is incompressible, like water. The *round window* is a relatively flexible pressure release allowing sound energy to be transmitted to the fluid of the cochlea via the oval window. In the inner ear the traveling waves set up on the basilar membrane by vibrations of the oval window stimulate *hair cells* that send nerve impulses to the brain.

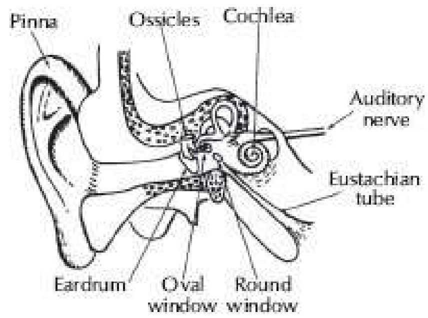


Figure 3-1. A cross-section of the human ear showing the relationship of the various parts.

3.2.1 Pinna

The *pinna*, or the human *auricle*, is the most lateral (i.e., outside) portion of our auditory system. The beauty of these flaps on either side of our head may be questioned, but not the importance of the acoustical function they serve. Fig. 3-3 shows an illustration of various parts of the human pinna. The entrance to the ear canal, or *concha*, is most important acoustically for filtering because it contains the largest air volume in a pinna. Let us assume for the moment that we have no pinnae, just holes in the head, which is actually a simplest model for human hearing, called the *spherical head model*. Cupping our hands around the holes would make sounds louder as more sound energy is directed into the opening. How much does the pinna help in directing sound energy into the ear canal? We can get some idea of this by measuring the sound pressure at the opening of the ear canal with and without the hand behind the ear. Wiener and Ross[4] did this and found a gain of 3 to 5dB at most frequencies, but a peak of about 20dB in the vicinity of 1500Hz. Fig. 3-4 shows the transfer function measured by Shaw,[5] and the curves numbered 3 and 4 are for concha and pinna flange, respectively. The irregular and asymmetric shape of a pinna is not just for aesthetic reasons. In section 3.11, we will see that it is actually important for our ability to localize sounds and to aid in spatial-filtering of unwanted conversations.

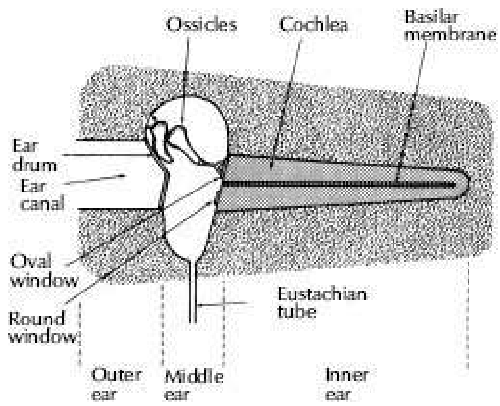


Figure 3-2. Highly idealized portrayal of the outer ear, middle ear, and inner ear.

3.2.2 Temporal Bones

On each of the left and right sides of our skull, behind the pinna, there is a thin, fanlike bone—namely, the *temporal bone*—covering the entire human ear, except for the pinna. This bone can be further divided into four portions, i.e., the *squamous, mastoid, tympanic* and *petrous portions*. The obvious function for the temporal bone is to protect our auditory system. Other than cochlear implant patients, whose temporal bone has to be partly removed during a surgery, people might not pay much attention to it, especially regarding acoustics. However the sound energy that propagates through the bone into our inner ear, as opposed to through the ear canal and middle ear, is actually fairly significant. For patients with conductive hearing loss, e.g., damage of middle ear, there are currently commercially available devices, which look something like headphones and are placed on the temporal bone. People with normal hearing can test it by plugging their ears while wearing the device. Although the timbres sound quite different from normal hearing, the filtered speech is clear enough to understand. Also because of this bone conduction, along with other effects such as acoustic reflex, which will be discussed in section 3.2.4.1, one hears his or her own voice differently from how other people hear the voice. While not receiving much attention in everyday life, it might be sometimes very important. For example, an experienced voice teacher often asks a student singer to record his or her own singing and playback with an audio system. The recording will sound unnatural to the singer but will be a more accurate representation of what the audience hears.

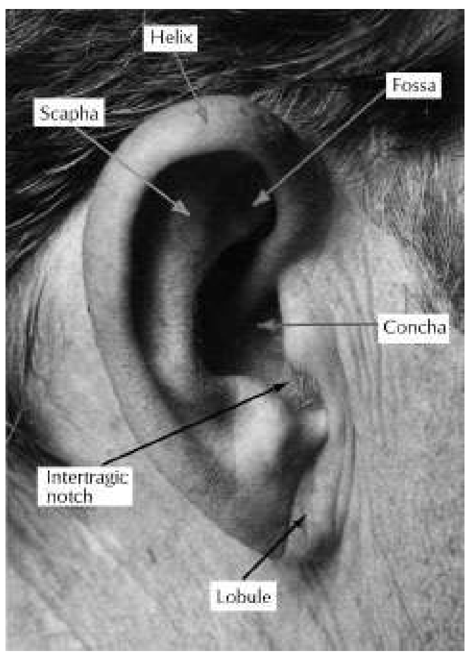


Figure 3-3. The human outer ear, the pinna, with identification of some of the folds, cavities, and ridges that have significant acoustical effect.

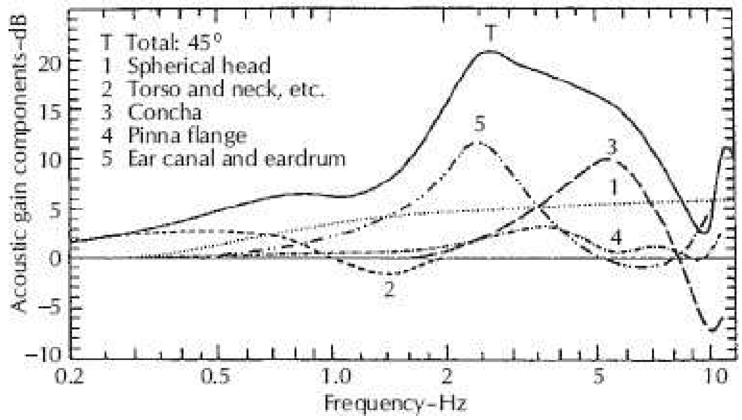


Figure 3-4. The average pressure gain contributed by the different components of the outer ear in humans. The sound source is in the horizontal plane, 45° from straight ahead.[5]

3.2.3 Ear Canal

The ear canal has a diameter about 5 to 9mm and is about 2.5cm long. It is open to the outside environment at the concha, and is closed at the tympanic membrane. Acoustically, it can be considered as a closed pipe whose cross-sectional shape and area vary along its length. Although being bent and irregular in shape, the ear canal does demonstrate the modal characteristic of a closed pipe. It has a fundamental frequency of about 3kHz, corresponding to a quarter wavelength close to the length of the ear canal. Because of this resonant frequency, our hearing is most sensitive to a frequency band around 3kHz, which is, not just by coincidence, the most important frequency band of human speech. On Fig. 3-4, the number 5 curve shows the effect of the ear canal, taking the eardrum into account as well. As can be seen, there is an approximately 11dB of gain at around 2.5kHz. After combining all the effects of head, torso and neck, pinna, ear canal and eardrum, the total transfer function is the curve marked with a letter T on Fig. 3-4. It is relatively broadly tuned between 2 and 7kHz, with as much as 20dB of gain. Unfortunately, because of this resonance, in very loud noisy environments with broadband sound, hearing damage usually first happens around 4kHz.

3.2.4 Middle Ear

The outer ear, including the pinna and the ear canal, ends at the eardrum. It is an air environment with low impedance. On the other hand, the inner ear, where the sensory cells are, is a fluid environment with high impedance. When sound (or any wave) travels from one medium to another, if the impedances of the two media do not match, much of the energy would be reflected at the surface, without propagating into the second medium. For the same reason, we use microphones to record in the air and hydrophones to record under water. To make our auditory system efficient, the most important function of the middle ear is to match the impedances of outer and inner ears. Without the middle ear, we would suffer a hearing loss of about 30 dB (by mechanical analysis[6] and experiments on cats[7]).

A healthy middle ear (without middle ear infection) is an air-filled space. When swallowing, the eustachian tube is open to balance the air pressure inside the middle ear and that of the outside world. Most of the time, however, the middle ear is sealed from the outside environment. The main components of the middle ear are the three ossicles, which are the smallest bones in our body: the malleus, incus, and stapes. These ossicles form an ossicular chain, which is firmly fixed on the eardrum and the oval window on each side. Through mostly three types of rocking actions—namely piston motion, lever motion, and buckling motion[8]—the acoustic energy is transferred into the inner ear effectively. The middle ear can be damaged temporarily by middle ear infection, or permanently by genetic disease. Fortunately, with current technology, doctors can rebuild the ossicles with titanium, the result being a total recovering of hearing.[9] Alternatively one can use devices that rely on bone conduction.[10]

Acoustic Reflex

There are two muscles in the middle ear: the *tensor tympani* that is attached to the malleus, and the *stapedius muscle* that is attached to the stapes. Unlike other muscles in our bodies, these muscles form an angle with respect to the bone, instead of along the bone, which makes them very ineffective for motion. Actually the function of these muscles is for changing the stiffness of the ossicular chain. When we hear a very loud sound, i.e., at least 75dB higher than the hearing threshold, when we talk or sing, when the head is touched, or when the body moves,[11] these middle ear muscles will contract to increase the stiffness of the ossicular chain, which makes it less effective, so that our inner ear is protected from exposure to the loud sound. However, because this process involves a higher stage of signal processing, and because of the filtering features, this protection works only for slow onset and low-frequency sound (up to 1.2kHz)[12] and is not effective for noises such as an impulse or noise with high frequencies (e.g., most of the music recordings today).

3.2.5 Inner Ear

The inner ear, or the *labyrinth*, is composed of two systems: the *vestibular system*, which is critical to our sense of balance, and the auditory system, which is used for hearing. The two systems share fluid, which is separated from the air-filled space in the middle ear by the oval window and the round window. The auditory portion of the inner ear is the snail-shaped cochlea. It is a mechanical-to-electrical transducer and a frequency-selective analyzer, sending coded nerve impulses to the brain. This is represented crudely in Fig. 3-5. A rough sketch of the cross section of the cochlea is shown in Fig. 3-6. The cochlea, throughout its length (about 35 mm if stretched out straight), is divided by *Reissner's membrane* and the *basilar membrane* into three separate compartments—namely, the *scala vestibuli*, the *scala media*, and the *scala tympani*. The *scala vestibuli* and the *scala tympani* share the same fluid, *perilymph*, through a small hole, the *helicotrema*, at the apex; while the *scala media* contains another fluid, *endolymph*, which contains higher density of potassium ions facilitating the function of the hair cells. The *basilar membrane* supports the *Organ of Corti*, which contains the hair cells that convert the relative motion between the *basilar membrane* and the *tectorial membrane* into nerve pulses to the auditory nerve.

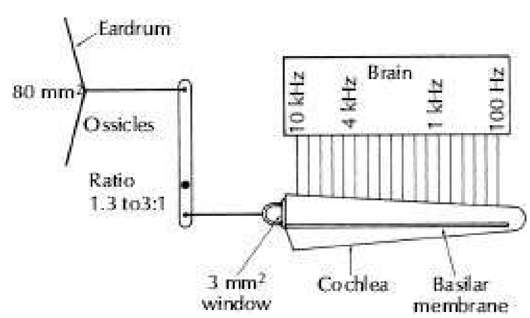


Figure 3-5. The mechanical system of the middle ear.

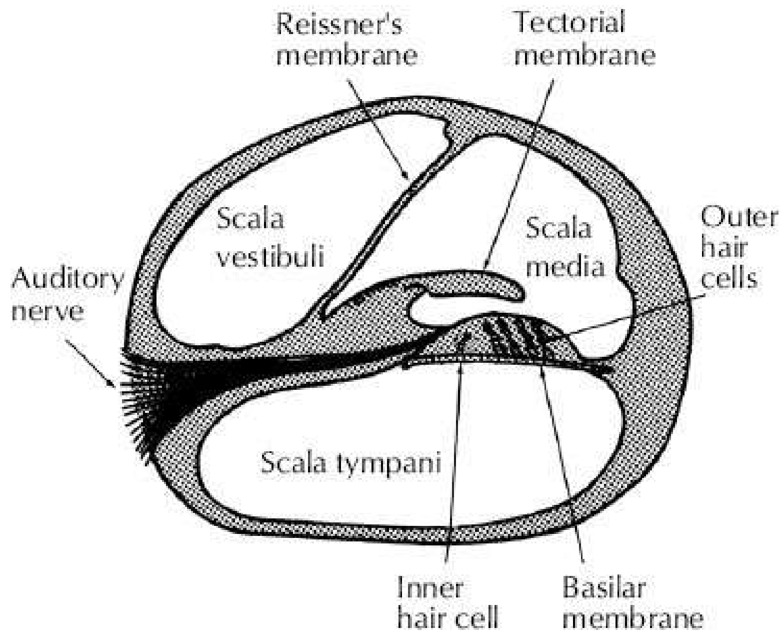


Figure 3-6. Cross-sectional sketch of the cochlea.

When an incident sound arrives at the inner ear, the vibration of the stapes is transported into the scala vestibuli through the oval window. Because the cochlear fluid is incompressible, the round window connected to the scala tympani vibrates accordingly. Thus, the vibration starts from the base of the cochlea, travels along the scala vestimbuli, all the way to the apex, and then through the helicotrema into the scala tympani, back to the base, and eventually ends at the round window. This establishes a traveling wave on the basilar membrane for frequency analysis. Each location at the basilar membrane is most sensitive to a particular frequency, i.e., the characteristic frequency, although it also responds to a relatively broad frequency band at smaller amplitude. The basilar membrane is narrower (0.04mm) and stiffer near the base, and wider (0.5mm) and looser near the apex. (By contrast, when observed from outside, the cochlea is wider at the base and smaller at the apex.) Therefore, the characteristic frequency decreases gradually and monotonically from the base to the apex, as indicated in Fig. 3-5. The traveling-wave phenomenon illustrated in Figs. 3-7 and 3-8 shows the vibration patterns, i.e., amplitude versus location, for incident pure tones of different frequencies. An interesting point in Fig. 3-8 is that the vibration pattern is asymmetric, with a gradual tail close to the base (for high frequencies) and a steep edge close to the apex (for low frequencies). Because of this asymmetry, it is easier for the low frequencies to mask the high frequencies than vice versa.

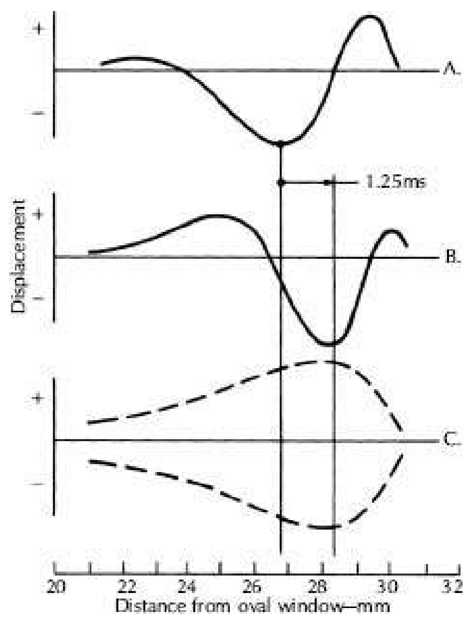


Figure 3-7. A traveling wave on the basilar membrane of the inner ear. (After von Bekesy, Reference 13.) The exaggerated amplitude of the basilar membrane for a 200Hz wave traveling from left to right is shown at A. The same wave 1.25ms later is shown at B. These traveling 200Hz waves all fall within the envelope at C.

Within the Organ of Corti on the basilar membrane, there are a row of *inner hair cells* (IHC), and three to five rows of *outer hair cells* (OHC), depending on location. There are about 1500 IHCs and about 3500 OHCs. Each hair cell contains stereociliae, i.e., hairs, that vibrate corresponding to the mechanical vibration in the fluid around them. Because each location on the basilar membrane is most sensitive to its own characteristic frequency, the hair cells at the location also respond most to the characteristic frequency. The IHCs are sensory cells, like microphones, which convert mechanical vibration into electrical signal, i.e., neural firings. The OHCs, on the other hand, change their shapes according to vibration of the basilar membrane and control signals received from efferent nerves. Their function is to give an extra gain or attenuation, so that the output of the IHC is tuned to the characteristic frequency much more sharply than the IHC itself. Fig. 3-9 shows the tuning curve (output level vs. frequency) for a particular location on the basilar membrane with and without functioning OHCs. The tuning curve is much broader with poor frequency selectivity when the OHCs do not function. The OHCs make our auditory system an active device, instead of a passive microphone. Because the OHCs are active and consume a lot of energy and nutrition, they are usually damaged first due to loud sound or ototoxic medicines (i.e., medicine that is harmful to the auditory system). Not only does this kind of hearing loss make our hearing less sensitive, it also makes our hearing less sharp. Thus, as is easily confirmed for people with hearing loss, simply adding an extra gain with hearing aids would not totally solve the problem.

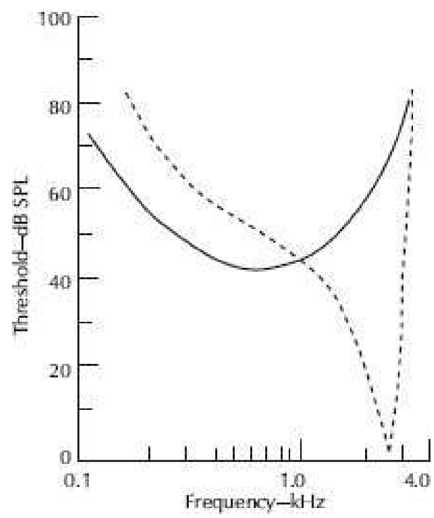


Figure 3-9. Tuning curve with (dashed) and without (solid) functioning outer hair cells. (Liberman and Dodds.[14]

3.3 Frequency Selectivity

3.3.1 Frequency Tuning

As discussed in section 3.2.5, the inner hair cells are sharply tuned to the characteristic frequencies with help from the outer hair cells. This tuning character is also conserved by the auditory neurons connecting to the inner hair cells. However, this tuning feature varies with level. Fig. 3-10 shows a characteristic diagram of tuning curves from a particular location on the basilar membrane at various levels. As can be seen in this graph, as level increases, the tuning curve becomes broader, indicating less frequency selectivity. Thus, in order to hear music more sharply, one should play back at a relatively low level. Moreover, above 60dB, as level increases, the characteristic frequency decreases. Therefore when one hears a tone at a high level, a neuron that is normally tuned at a higher characteristic frequency is now best tuned to the tone. Because eventually the brain perceives pitch based on neuron input, at high levels, without knowing that the characteristic frequency has decreased, the brain hears the pitch to be sharp.

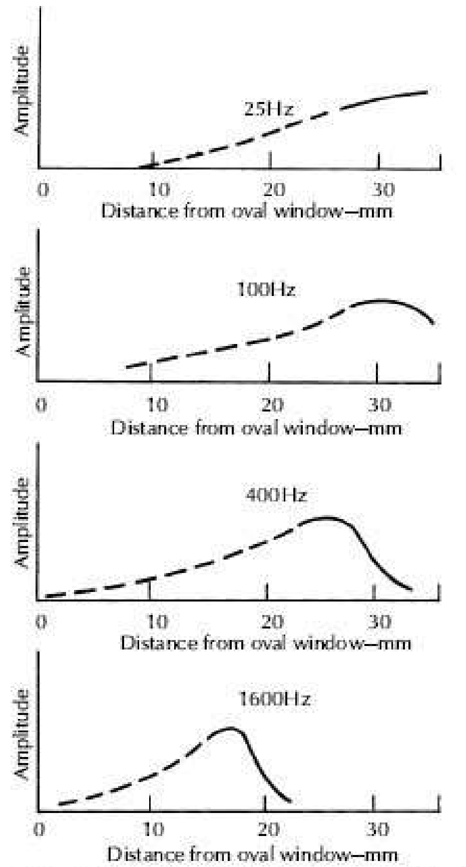


Figure 3-8. An illustration of vibration patterns of the hair cells on the basilar membrane for various incident pure tones. There is a localized peak response for each audible frequency.[13]

Armed with this knowledge, one would think that someone who was engaged in critical listening—a recording engineer, for example—would choose to listen at moderate to low levels. Why then do so many audio professionals choose to monitor at very high levels? There could be many reasons. Loud levels may be more exciting. It may simply be a matter of habit. For instance, an audio engineer normally turns the volume to his or her customary level fairly accurately. Moreover, because frequency selectivity is different at different levels, an audio engineer might choose to make a recording while listening at a “realistic” or “performance” level rather than monitoring at a level that is demonstratedly more accurate. Finally, of course, there are some audio professionals who have lost some hearing already, and in order to pick up certain frequency bands they keep on boosting the level, which unfortunately further damages their hearing.

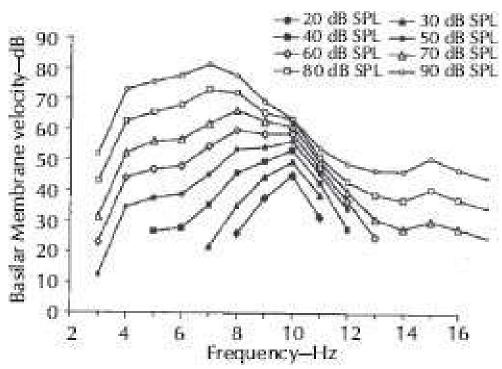


Figure 3-10. Response curve at various levels at a particular location of the basilar membrane of a chinchilla.[15,16]

3.3.2 Masking and Its Application in Audio Encoding

Suppose a listener can barely hear a given acoustic signal under quiet conditions. When the signal is playing in presence of another sound (called “a *masker*”), the signal has to be stronger so that the listener can hear it.[17] The masker does not have to include the frequency components of the original signal for the masking effect to take place, and a masked signal can already be heard when it is still weaker than the masker.[18]

Masking can happen when a signal and a masker are played simultaneously (simultaneous masking), but it can also happen when a masker starts and ends before a signal is played. This is known as forward masking. Forward masking can happen even when the signal starts more than 100ms after the masker stops.[19]

The masking effect has been widely used in psychoacoustical research. For example, Fig. 3-10 shows the response curve for a chinchilla. For safety reasons, performing such experiments on human subjects is not permitted. However, with masking effect, one can vary the level of a masker, measure the threshold (i.e., the minimum sound that the listener can hear), and create a diagram of a psychophysical tuning curve that reveals similar features.

Besides scientific research, masking effects are also widely used in areas such as audio encoding. Now, with distribution of digital recordings, it is desirable to reduce the sizes of audio files. There are *lossless encoders*, an algorithm to encode the audio file into a smaller file such that the original audio can be completely reconstructed with another algorithm (decoder). However, the file sizes of the lossless encoders are still relatively large. To further reduce the size, some less important information has to be eliminated. For example, one might eliminate high frequencies, which is not too bad for speech communication. However, for music, some important quality might be lost. Fortunately, because of the masking effect, one can eliminate some weak sounds that are masked so that listeners hardly notice the difference. This technique has been widely used in audio encoders, such as MP3.

3.3.3 Auditory Filters and Critical Bands

Experiments show that our ability to detect a signal depends on the bandwidth of the signal. Fletcher (1940)[18] found that, when playing a tone in the presence of a bandpass masker, as the masker bandwidth was increased while keeping the overall level of the masker unchanged, the threshold increased as bandwidth increased up to a certain limit, beyond which the threshold remained constant. One can easily confirm that, when listening to a bandpass noise with broadening bandwidth and constant overall level, the loudness is unchanged, until a certain bandwidth is reached, and beyond that bandwidth the loudness increases as bandwidth increases, although the reading of an SPL meter is constant. An explanation to account for these effects is the concept of auditory filters. Fletcher proposed that, instead of directly listening to each hair cell, we hear through a set of auditory filters, whose center frequencies can vary or overlap, and whose bandwidth is varying according to the center frequency. These bands are referred to as *critical bands* (CB). Since then, the shape and bandwidth of the auditory filters have been carefully studied. Because the shape of the auditory filters is not simply rectangular, it is more convenient to use the *equivalent rectangular bandwidth* (ERB), which is the bandwidth of a rectangular filter that gives the same transmission power as the actual auditory filter. Recent study by Glasberg and Moore (1990) gives a formula for ERB for young listeners with normal hearing under moderate sound pressure levels[20]

$$\mathrm{ERB} = 24.7 \times (4.37F + 1) \qquad (3\text{-}1)$$

where,
the center frequency of the filter F is in kHz,
ERB is in Hz.

Sometimes, it is more convenient to use an ERB number as in Eq. 3-2,[20] similar to the Bark scale proposed by Zwicker et al.[21]

$$\text{ERB number} = 21.4\log_{10}(4.37F + 1) \qquad (3\text{-}2)$$

where,
the center frequency of the filter F is in kHz.

Table 3-1 shows the ERB and Bark scale as a function of the center frequency of the auditory filter. The Bark scale is also listed as a percentage of center frequency, which can then be compared to filters commonly used in acoustical measurements: octave (70.7%), half octave (34.8%), one-third octave (23.2%), and one-sixth octave (11.6%) filters. The ERB is shown in Fig. 3-11 as a function of frequency. One-third octave filters which are popular in audio and have been widely used in acoustical measurements ultimately have their roots in the study of human auditory response. However, as Fig. 3-11 shows, the ERB is wider than 1/3-octave for frequencies below 200Hz; is smaller than 1/3-octave for frequencies above 200Hz; and, above 1kHz, it approaches 1/6-octave.

Table 3-1. Critical Bandwidths of the Human Ear

Critical Band No	Center Frequency Hz	Bark Scale (Hz)	%	Equivalent Rectangular Band (*ERB*), Hz
1	50	100	200	33
2	150	100	67	43
3	250	100	40	52
4	350	100	29	62
5	450	110	24	72
6	570	120	21	84
7	700	140	20	97
8	840	150	18	111
9	1000	160	16	130
10	1170	190	16	150
11	1370	210	15	170
12	1600	240	15	200
13	1850	280	15	220
14	2150	320	15	260
15	2500	380	15	300
16	2900	450	16	350
17	3400	550	16	420
18	4000	700	18	500
19	4800	900	19	620
20	5800	1100	19	780
21	7000	1300	19	990
22	8500	1800	21	1300
23	10500	2500	24	1700
24	13500	3500	26	2400

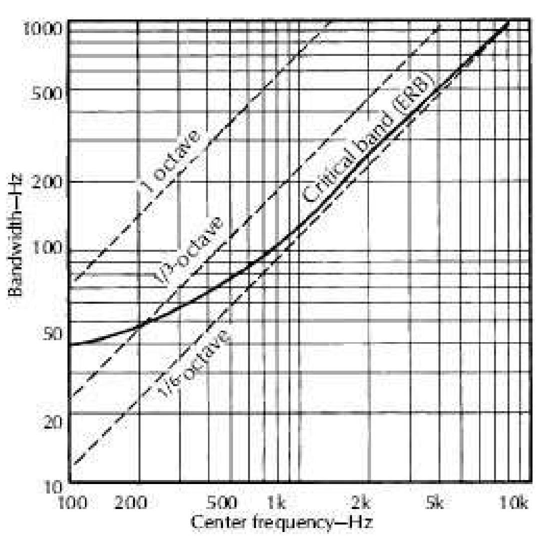


Figure 3-11. A plot of critical bandwidths (calculated ERBs) of the human auditory system compared to constant percentage bandwidths of filter sets commonly used in acoustical measurements.

3.4 Nonlinearity of the Ear

When a set of frequencies are input into a linear system, the output will contain only the same set of frequencies, although the relative amplitudes and phases can be adjusted due to filtering. However, for a nonlinear system, the output will include new frequencies that are not present in the input. Because our auditory system has developed mechanisms such as the active processes in the inner ear, it is nonlinear. There are two types of non-linearity—namely, *harmonic distortion* and *combination tones*. Harmonic distortion can be easily achieved by simply distorting a sine-tone. The added new components are harmonics of the original signal. A combination tone happens when there are at least two frequencies in the input. The output might include combination tones according to

$$f_c = |n \times f_1 \pm m \times f_2| \qquad (3\text{-}3)$$

where,

f_c is the frequency of a combination tone,

f_1 and f_2 are the two input frequencies, and n and m are any integer numbers.

For example, when two tones at 600 and 700Hz are input, the output might have frequencies such as 100Hz (= 700 – 600Hz), 500Hz (= 2 × 600 – 700Hz), and 400Hz (= 3 × 600 – 2 × 700Hz), etc.

Because the harmonic distortion does not change the perception of pitch, it would not be surprising if we are less tolerant of the combination tones.

Furthermore, because the auditory system is active, even in a completely quiet environment, the inner ear might generate tones. These *otoacoustic emissions*[22] are a sign of a healthy and functioning inner ear, and quite different from the *tinnitus* resulting from exposure to dangerously high sound pressure levels.

3.5 Perception of Phase

The complete description of a given sound includes both an amplitude spectrum and a phase spectrum. People normally pay a lot of attention to the amplitude spectrum, while caring less for the phase spectrum. Yet academic researchers, hi-fi enthusiasts, and audio engineers all have asked, "Is the ear able to detect phase differences?" About the middle of the 19th century, G. S. Ohm wrote, "Aural perception depends only on the amplitude spectrum of a sound and is independent of the phase angles of the various components contained in the spectrum." Many apparent confirmations of Ohm's law of acoustics have later been traced to crude measuring techniques and equipment, although the law works for simple sounds: when the tone consists of only a few low-order harmonics, the ear is phase-deaf.

Actually, the phase spectrum sometimes can be very important for the perception of timbre. For example, an impulse and white noise sound quite different, but they have identical amplitude spectrum. The only difference occurs in the phase spectrum.

Another common example is speech: if one scrambles the relative phases in the spectrum of a speech signal, it will not be intelligible. Now, with experimental evidence, we can confirm that our ear is capable of detecting phase information. For example, the neural firing of the auditory nerve happens at a certain phase, which is called the *phase-locking*, up to about 5kHz.[23] The phase-locking is important for pitch perception. In the brainstem, the information from left and right ears is integrated, and the *interaural phase difference* can be detected, which is important for spatial hearing. These phenomena will be discussed in more detail in sections 3.9 and 3.11.

3.6 Auditory Area and Thresholds

The auditory area depicted in Fig. 3-12 describes, in a technical sense, the limits of our aural perception. This area is bounded at low sound levels by our threshold of hearing. The softest sounds that can be heard fall on the threshold of hearing curve. Above this line the air molecule movement is sufficient to elicit a response. If, at any given frequency, the sound pressure level is increased sufficiently, a point is reached at which a tickling sensation is felt in the ears. If the level is increased substantially above this threshold of feeling, it becomes painful. These are the lower and upper boundaries of the auditory area. There are also frequency limitations below about 20Hz and above about 16kHz, limitations that (like the two thresholds) vary considerably from individual to individual. We are less concerned here about specific numbers than we are about principles. On the auditory area of Fig. 3-12, all the sounds of life are played out—low frequency or high, very soft or very intense. Speech does not utilize the entire auditory area. Its dynamic range and frequency range are quite limited. Music has both a greater dynamic range than speech and a greater frequency range. But even music does not utilize the entire auditory area.

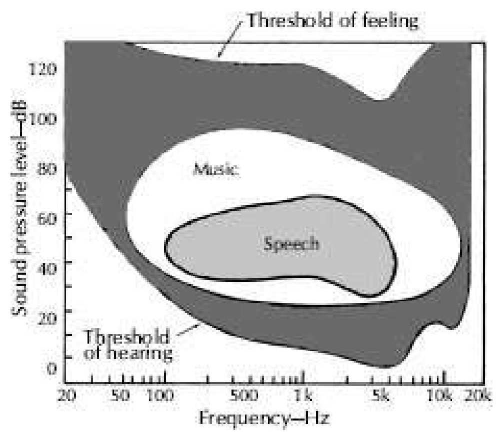


Figure 3-12. All sounds perceived by humans of average hearing acuity fall within the auditory area. This area is defined by the threshold of hearing and the threshold of feeling (pain) and by the low and high frequency limits of hearing. Music and speech do not utilize the entire auditory area available, but music has the greater dynamic range (vertical) and frequency demands (horizontal).

3.7 Hearing Over Time

If our ear was like an ideal Fourier analyzer, in order to translate a waveform into a spectrum, the ear would have to integrate over the entire time domain, which is not practical and, of course, not the case. Actually, our ear only integrates over a limited time window (i.e., a filter on the time axis), and thus we can hear changes of pitch, timbre, and dynamics over time, which can be shown on a *spectrogram* instead of a simple spectrum. Mathematically, it is a spectral-temporal analysis instead of a Fourier analysis. Experiments on gap detection between tones at different frequencies indicate that our temporal resolution is on the order of 100ms,[24] which is a good estimate of the time window of our auditory system. For many perspectives (e.g., perceptions on loudness, pitch, timbre), our auditory system integrates acoustical information within this time window.

3.8 Loudness

Unlike level or intensity, which are physical or objective quantities, loudness is a listener's subjective perception. As the example in section 3.3, even if the SPL meter reads the same level, a sound with a wider bandwidth might sound much louder than a sound with a smaller bandwidth. Even for a pure tone, although loudness follows somewhat with level, it is actually a quite complicated function, depending on frequency. A tone at 40dB SPL is not necessarily twice as loud as another sound at 20dB SPL. Furthermore, loudness also varies among listeners. For example, a listener who has lost some sensitivity in a certain critical band will perceive any signal in that band to be at a lower level relative to someone with normal hearing.

Although there is no meter to directly measure a subjective quantity such as loudness, psychophysical scaling can be used to investigate loudness across subjects. Subjects can be given matching tasks, where they are asked to adjust the level of signals until they match, or comparative tasks, where they are asked to compare two signals and give a number for loudness.

3.8.1 Equal Loudness Contours and Loudness Level

By conducting experiments using pure tones with a large population, Fletcher and Munson at Bell Labs (1933) derived equal loudness contours, also known as the *Fletcher-Munson curves*. Fig. 3-13 shows the equal loudness contours later refined by Robinson and Dadson, which have been recognized as an international standard. On the figure, the points on each curve correspond to pure tones that give the same loudness to an average listener. For example, a pure tone at 50Hz at 60dB SPL is on the same curve as a tone at 1kHz at 30dB. This means that these two tones have identical loudness to an average listener. Obviously, the level for the 50Hz tone is 30dB higher than the level of the 60Hz tone, which means that we are much less sensitive to the 50Hz tone. Based on the equal loudness contours, *loudness level*, in *phons*, is introduced. It is always referenced to a pure tone at 1kHz. The loudness level of a pure tone (at any frequency) is defined as the level of a 1kHz tone that has identical loudness to the given tone for an average listener. For the above example, the loudness of the 50Hz pure tone is 30phons, which means it is as loud as a 30dB pure tone at 1kHz. The lowest curve marked with “minimum audible” is the hearing threshold. Although many normal listeners can hear tones weaker than this threshold at some frequencies, on average, it is a good estimate of a minimum audible limit. The tones louder than the curve of 120 phons will cause pain and hearing damage.

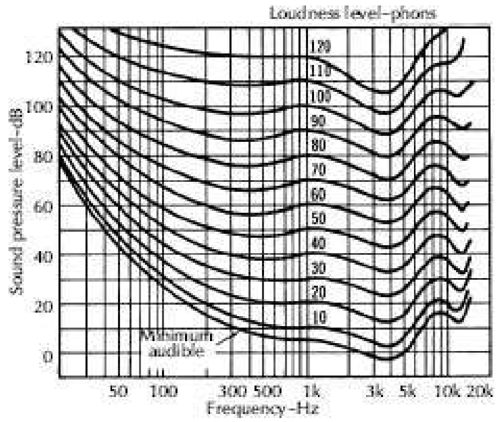


Figure 3-13. Equal loudness contours for pure tones in a frontal sound field for humans of average hearing acuity determined by Robinson and Dadson. The loudness levels in phons correspond to the sound pressure levels at 1000Hz. (ISO Recommendation 226.)

The equal loudness contours also show that human hearing is most sensitive around 4kHz (which is where hearing damage due to loud sounds first happens), less sensitive to high frequencies, and much less sensitive to very low frequencies (which is why a subwoofer has to be very powerful to produce strong bass, the price of which is the masking of mid- and high-frequencies and potential hearing damage). A study of this family of curves tells us why treble and bass frequencies seem to be missing or down in level when favorite recordings are played back at low levels.[25]

One might notice that for high frequencies above 10kHz, the curves are nonmonotonic for low levels. This is due to the second resonant mode of the ear canal. Moreover, at low frequencies below 100Hz, the curves are close to each other, and the change of a few dB can give you the feeling of more than 10 dB of dynamic change at 1kHz. Furthermore, the curves are much flatter at high levels, which unfortunately encouraged many to listen to reproduced music at abnormally high levels, again causing hearing damage. Actually, even if one wanted to have flat or linear hearing, listening at abnormally high levels might not be wise, because the frequency selectivity of our auditory system will be much poorer, leading to much greater interaction between various frequencies. Of course, one limitation of listening at a lower level is that, if some frequency components fall below the hearing threshold, then they are not audible. This problem is especially important for people who have already lost some acuity at a certain frequency, where his or her hearing threshold is much higher than normal. However, in order to avoid further damage of hearing, and in order to avoid unnecessary masking effect, one still might consider listening at moderate levels.

The loudness level considers the frequency response of our auditory system, and therefore is a better scale than the sound pressure level to account for loudness. However, just like the sound pressure level is not a scale for loudness, the loudness level does not directly represent loudness, either. It simply references the sound pressure level of pure tones at other frequencies to that of a 1kHz pure tone. Moreover, the equal loudness contours were achieved with pure tones only, without consideration of the interaction between frequency components, e.g., the compression within each auditory filter. One should be aware of this limit when dealing with broadband signals, such as music.

3.8.2 Level Measurements with A-, B-, and C-Weightings

Although psychoacoustical experiments give better results on loudness, practically, level measurement is more convenient. Because the equal loudness contours are flatter at high levels, in order to make level measurements somewhat representing our loudness perception, it is necessary to weight frequencies differently for measurements at different levels. Fig. 3-14 shows the three widely used weighting functions.[26] The A-weighting level is similar to our hearing at 40dB, and is used at low levels; the B-weighting level represents our hearing at about 70dB; and the C-weighting level is more flat, representing our hearing at 100dB, and thus is used at high levels. For concerns on hearing loss, the A-weighting level is a good indicator, although hearing loss often happens at high levels.

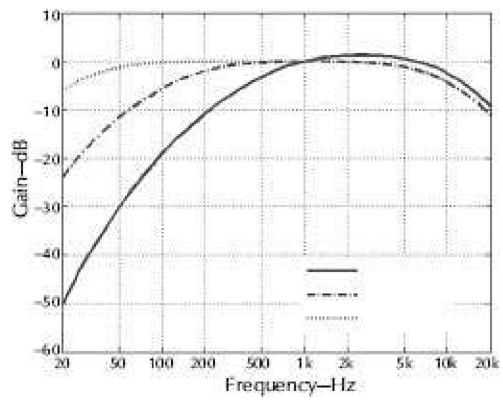


Figure 3-14. Levels with A-, B-, and C-weightings.[26]

3.8.3 Loudness in Sones

Our perception of loudness is definitely a compressed function (less sensitive for higher levels), giving us both sensitivity for weak sounds and large dynamic range for loud sounds. However, unlike the logarithmic scale (dB) that is widely used in sound pressure level, experimental evidence shows that loudness is actually a power law function of intensity and pressure as shown in Eq. 3-4.

$$\begin{aligned} \text{Loudness} &= k \times I^{\alpha} \\ &= k' \times p^{2\alpha} \end{aligned} \tag{3-4}$$

where,

k and k' are constants accounting for individuality of listeners,
I is the sound intensity,
p is the sound pressure,
α varies with level and frequency.

The unit for loudness is *sones*. By definition, one sone is the loudness of a 1kHz tone at a loudness level of 40phons, the only point where phons and SPL meet. If another sound sounds twice as loud as the 1kHz tone at 40phons, it is classified as 2sones, etc. The loudness of pure tones in sones is compared with the *SPL* in dB in Fig. 3-15. The figure shows that above 40dB, the curve is a straight line, corresponding to an exponent of about 0.3 for sound intensity and an exponent of 0.6 for sound pressure as in Eq. 3-4. The exponent is much greater for levels below 40dB, and for frequencies below 200Hz (which can be confirmed by the fact that the equal loudness contours are compact for frequencies below 200Hz on Fig. 3-13).

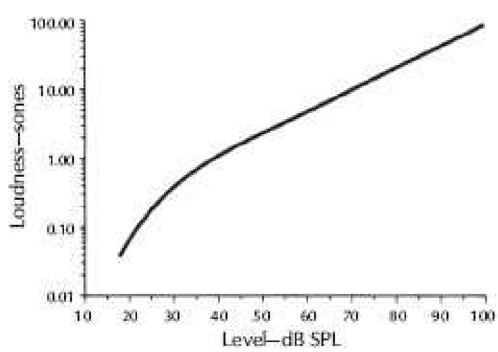


Figure 3-15. Comparison between loudness in sones and loudness level in phons for a 1 kHz tone.[15,27]

One should note that Eq. 3-4 holds for not only pure tones, but also bandpass signals within an auditory filter (critical band). The exponent of 0.3 (<1) indicates compression within the filter.

However, for a broadband signal that is wider than one critical bandwidth, Eq. 3-4 holds for each critical band, and the total loudness is simply the sum of loudness in each band (with no compression across critical bands).

3.8.4 Loudness versus Bandwidth

Due to less compression across critical bands, broadband sounds, such as white noise, seem to be much louder than pure tones or narrowband noise of the same sound pressure level. In fact, as in the example in section 3.3.3, increasing the bandwidth does not increase loudness until the critical bandwidth is exceeded. Beyond that point multiple critical bands are excited, and the loudness increases markedly with increase in bandwidth because of less compression across critical bands. For this reason, the computation of loudness for a wide band sound must be based on spectral distribution of energy. Filters no narrower than critical bands are required and 1/3-octave filters are commonly used.

3.8.5 Loudness of Bursts

Life is filled with burst-type sounds: snaps, pops, crackles, bangs, bumps, and rattles. For tone bursts or noise bursts with duration greater than 100ms, loudness is independent of burst duration. The effect on loudness for bursts shorter than 200ms is shown in Fig. 3-16. This curve shows how much higher the level of short bursts of noise and pure tones must be to sound as loud as continuous noise or pure tones. Bursts longer than 200ms are perceived to be as loud as continuous noise or tones of the same level. For the shorter bursts, the level must be increased to maintain the same loudness as for the longer bursts. Noise and tonal bursts are similar in the level of increase required to maintain the same loudness. Fig. 3-16 indicates that the ear has a time constant of about 200ms, confirming the time window on the order of 100ms, as discussed in section 3.7. This means that band levels should be measured with RMS detectors having integration times of about 200ms. This corresponds to the FAST setting on a sound level meter while the SLOW setting corresponds to an integration time of 500ms.

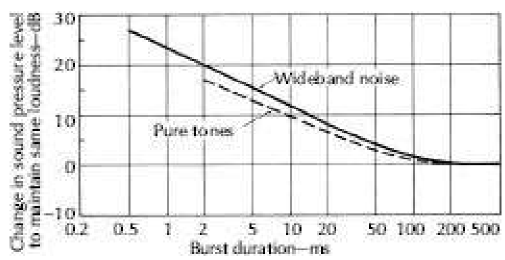


Figure 3-16. Short bursts of sound must be increased in level to sound as loud as longer bursts.

3.8.6 Perception of Dynamic Changes

How sensitive is our hearing to dynamic changes? In other words, how much intensity or level change will lead to a perception of change of loudness? To discuss this kind of problem, we need the concept of just-noticeable difference (JND), which is defined as the minimum change that can be detected. Weber's Law states that the JND in intensity, in general and not necessarily for hearing, is proportional to the overall intensity. If Weber's Law holds, the Weber fraction in dB as defined in Eq. 3-5 would be a constant, independent of the overall intensity and the overall level.

$$\text{Weber fraction in dB} = 10\log_{10}\left(\frac{\Delta I}{I}\right) = constant \qquad (3\text{-}5)$$

where,
I is the intensity,
ΔI is the JND of intensity.

Please note that the Weber fraction in dB is *not* the JND of SPL (ΔL), which can be calculated according to Eq. 3-6.

$$\Delta L = 10\log_{10}\left(1 + \frac{\Delta I}{I}\right) \qquad (3\text{-}6)$$

If ΔI is much smaller than I, Eq. 3-6 is approximately

$$\Delta L = 4.35 \times \left(1 + \frac{\Delta I}{I}\right) \qquad (3\text{-}7)$$

Fig. 3-17 shows the measurement of the Weber fraction for broadband signals up to 110dB SPL.[28] Above 30dB, the Weber fraction in dB is indeed a constant of about -10dB, corresponding to a JND (ΔL) of 0.4dB. However, for weak sounds below 30dB, the Weber fraction in dB is higher, and can be as high as 0dB, corresponding to a JND (ΔL) of 3 dB. In other words, our hearing is less sensitive (in level) for dynamic changes of sounds weaker than 30 dB. Interestingly, when measuring with pure tones, it was found that the Weber fraction is slightly different from the broadband signals.[29] This phenomenon is known as the *near-miss to Weber's Law*. Fig. 3-17 includes a more recent measurement for pure tones,[30] which demonstrates that the Weber fraction gradually decreases up to 85 dB SPL and can be lower than -12 dB, corresponding to a JND (ΔL) less than 0.3dB. The near-miss to Weber's Law for pure tones is believed to be associated with the broad excitation patterns across frequency at high levels.[31]

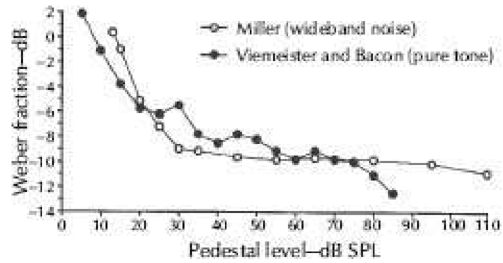


Figure 3-17. Just-noticeable difference (JND) for a broadband noise and for a 1 kHz tone.[15,28,30]

3.9 Pitch

Pitch seems to be a very clear concept, and yet it is very hard to give an accurate definition. The definition by the American National Standards Institute (ANSI) is as follows: “Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high.”[32] Like loudness, pitch is a subjective quantity. The ANSI standard also states: “Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus.”[32] William Hartmann has argued that an earlier ANSI definition (1960)[17] is better because it added that “low-to-high” refers to a property characterizing melody and not timbre.[33]

Roughly speaking, the sound we perceive as having a clear pitch is either a pure tone at a certain frequency, or a complex tone with a certain fundamental and a series of harmonics whose frequencies are multiples of the fundamental frequency. For example, when a violin is playing a tone of concert A (440Hz), the spectrum includes not only the frequency of 440Hz but also the frequencies of 880 (2 × 440)Hz, 1320 (3 × 440)Hz, and 1760 (4 × 440)Hz, etc.

A sound has pitch if it can be matched with a pure tone, i.e., a listener must be able to adjust the frequency of a pure tone to produce an identical pitch as the given sound. An opposite example is as follows. When one hits a small drum, it might sound higher than a bigger drum. However, normally one cannot match the sound that a drum produces with a pure tone. The exception, of course, would be a tympani or a steel drum. Therefore, the sound that most drums make does not result in the perception of pitch. Another attribute of pitch is that, if a sound has pitch, one can use it to make a melody. One could use a frequency generator to produce a pure tone at a frequency of 10kHz, and one could match it with another tone by listening to the beats. However, it would not be perceived as a tone, and it could not be used as part of a melody; therefore it would not be thought of as having pitch.[34] This will be discussed further in section 3.9.3.

3.9.1 The Unit of Pitch

The unit of *mel* is proposed as a measure of the subjective quantity of pitch.[35] It is always referenced to a pure tone at 1kHz at 40dB above a listener's threshold, which is defined as 1000mels. If another sound produces a pitch that sounds two times as high as this reference, it is considered to be 2000mels, etc. Fig. 3-18 shows the relationship between pitch in mels and frequency in Hz. The frequency axis in Fig. 3-18 is in logarithmic scale. However, the curve is not a straight line, indicating that our pitch perception is not an ideal logarithmic scale with respect to frequency in Hz. This relationship is probably more important for melodic intervals (when notes are played sequentially) than for chords (when notes are played simultaneously). In a chord, in order to produce a clean harmony, the notes have to coincide with the harmonics of the root note; otherwise, beats will occur, sounding out of tune. In the music and audio industry, it is much more convenient to use frequency in Hz or the unit of *cents* based on the objective quantity of frequency.

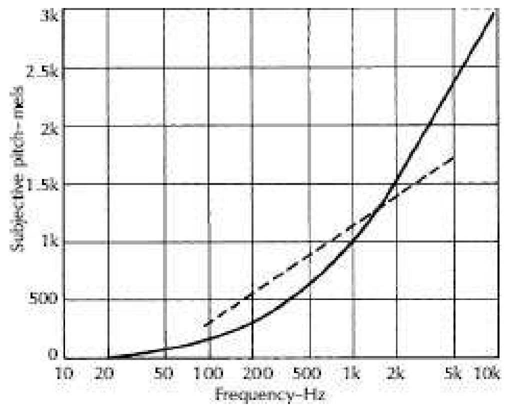


Figure 3-18. The relationship between frequency, a purely physical parameter, and pitch, a subjective reaction to the physical stimulus.[35]

Because our hearing is approximately a logarithmic scale on frequency, e.g., doubling frequency transposes a musical note to an octave higher, musical intervals between two tones can be described objectively in the unit of cents as defined by Eq. 3-8.

$$\begin{aligned}\text{Musical interval in cents} &= 1200\log_2\left(\frac{f_2}{f_1}\right)\\ &= \frac{1200}{\log_{10}2}\times\log_{10}\left(\frac{f_2}{f_1}\right) \qquad (3\text{-}8)\\ &= 3986\times\log_{10}\left(\frac{f_2}{f_1}\right)\end{aligned}$$

where,
f_1 and f_2 are the fundamental frequencies of the two tones.

Thus, a semitone on a piano (equal temperament) is 100 cents, and an interval of an octave is 1200cents. Using the unit of cents, one can easily describe the differences among various temperaments (e.g., equal temperament, Pythagorean scale, Just-tuning, etc.).

3.9.2 Perception of Pure and Complex Tones

How does our brain perceive pitch? The basilar membrane in the inner ear functions as a frequency analyzer: pure tones at various frequencies will excite specific locations on the basilar membrane. This would seem to suggest that the location of the maximum excitation on the basilar membrane determines the pitch. Actually, the process is much more complicated: besides the place coding, there is also temporal coding, which takes advantage of the temporal synchrony among neural spikes. The temporal coding is necessary for perceiving the pitch of complex tones, the virtual pitch with missing fundamentals, etc.[36,37] The theories based on place coding and temporal coding have been proposed to explain the origin of perception of pure and complex pitches. For either the place theory or the temporal theory, there is experimental evidence supporting and disfavoring it. As we develop our knowledge, we will probably understand more about when each coding takes place.

When hearing a complex tone, we are tolerant of harmonics slightly mistuned.[38] For instance, if three frequencies of 800, 1000, and 1200Hz (i.e., the 4th, 5th and 6th harmonics of a fundamental of 200Hz) are combined and presented to a listener, the pitch perceived is 200Hz. If all of them are shifted upward by 30Hz, i.e., 830, 1030, and 1230Hz, the fundamental, theoretically, is now 10Hz, and those three components are now the 83rd, 103rd, and 123rd harmonics of the fundamental of 10Hz. However, when playing this mistuned complex tone, listeners can hear a clear pitch at 206Hz, which matches the middle frequency, i.e., 1030Hz, the 4th harmonic of the fundamental. Although the other two frequencies are slightly mistuned (in opposite directions) as harmonics, the pitch is very strong.

It is worth mentioning that pitch recognition is often a synthetic process. When two harmonics of the same fundamental are presented to each ear independently, it is sometimes possible to hear the pitch at the fundamental frequency.[39]

3.9.3 Phase-Locking and the Range of Pitch Sensation

What is the range of the fundamental frequency that produces a pitch? Is it the same as the audible range from 20Hz to 20kHz? The lowest key in a piano is 27.5Hz, not too far from the lowest limit. However, for the high limit, it is only about 5kHz. Because the pitch perception requires temporal coding, the auditory neurons have to fire at a certain phase of each cycle, which is called *phase-locking*. Unfortunately, the auditory system is not able to phase-lock to frequencies above 5kHz.[40] This is why the highest note on a piccolo, which is the highest pitch in an orchestra, is 4.5kHz, slightly lower than 5kHz. Notes with fundamentals higher than 5kHz are not perceived as having pitch and cannot be used for musical melodies, though it should be noted that the highest note on a pipe organ is 8.4kHz. One can easily confirm this statement by transposing a familiar melody by octaves: when the fundamental is above 5kHz, although one can hear something changing, the melody cannot be recognized any more.

3.9.4 Frequency Difference Limen

The frequency difference limen is another way of saying “the just-noticeable difference in frequency.” It is the smallest frequency difference that a listener can discriminate. Experiments with pure tones of duration of 500ms show that, for levels higher than 10dB above threshold, between 200Hz and 5kHz, the frequency difference limen is less than 0.5% of the given frequency, corresponding to 9cents[41] (about 1/10 of a semitone).

3.9.5 Dependence of Pitch on Level

Pitch can be affected by level, however, the influence is not universal across frequency. At frequencies below 1kHz, the pitch decreases as level increases; whereas at frequencies above 3kHz, the pitch increases with increasing level; and at frequencies between 1 and 3kHz, varying the level has little effect on pitch. This is known as *Stevens rule*. Terhardt et al. summarized several studies of level dependence of pure tones and came up with the following equation for an average listener.[42]

$$100 \times \frac{p-f}{f} = 0.02 \times (L-60) \times \left(\frac{f}{1000} - 2\right) \qquad (3\text{-}9)$$

where,

f is the frequency in Hz of a pure tone at a sound pressure level of *L* in dB,

p is the frequency of a pure tone at 60dB SPL that matches the pitch of the given tone *f*.

3.9.6 Perfect Pitch

Some people, especially some musicians, develop perfect pitch, also known as *absolute pitch*. They can identify the pitch of a musical tone without help from an external reference like a tuner, i.e., they have established an absolute scale of pitch in their heads. Some of them describe the feeling of certain note analogous to a certain color. Some believe that one can establish a sensation of perfect pitch if he or she has a lot of experience listening to music in certain keys (which is normally due to musical training) before the age of 4. It is fair to state that having perfect pitch is not a requirement for a fine musician. Other than the advantage of tuning musical instruments or singing without a tuning device, there is no evidence that a person with perfect pitch would sing more accurately in tune. With the help of a tuner or accompaniment, a good musician without perfect pitch would do just as well. There is, however, a disadvantage due to an age effect. For senior persons (especially those above the age of 65), the pitch scales are often shifted so that they hear music that is being played normally to be out of tune. This might not be noticeable for a person without perfect pitch. However, for a senior musician with perfect pitch, he or she might find it annoying when perceiving everyone else in the orchestra playing out of tune. He or she has to live with it, otherwise he or she might be the only one in the orchestra playing out of tune.

3.9.7 Other Pitch Effects

The pitch of a complex tone is given a name corresponding to the fundamental frequency. However, the fundamental of a complex tone can be missing or masked with a narrowband noise, while still producing a clear pitch.[43] The pitch produced without fundamental is called *virtual pitch*, and it is evidence favoring the temporal theory over the place theory. The waveform of a virtual pitch has the same period as a normal complex tone including the fundamental at the same frequency.

When listening to a broadband signal with certain interaural phase relationships, although listening with one ear does not produce a pitch, when listening with both ears, one can hear a pitch on top of the background noise. Such pitches are called "binaural pitches."[44,45]

3.10 Timbre

Timbre is our perception of sound color. It is that subjective dimension that allows us to distinguish between the sound of a violin and a piano playing the same note. The definition by the American Standards Association states that the timbre is "that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar," and "timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus."[17] Timbre enables instrumental recognition. However, paradoxically the timbre of different notes on any given instrument can be quite different. There are two aspects of timbre. The tone color refers to the steady-state spectrum, e.g., tones with strong high harmonics have a "bright" timbre. On the other hand, the transient refers to the onset and offset, i.e., percussive instruments (e.g., piano) vs. slow onsets (e.g., bowed string or brass). The timbre of a sound produced in a concert hall may even vary with listener position because of the effects of air absorption and because of the frequency-dependent absorption characteristics of room surfaces.

It is worth noting that, in order to more completely describe timbre, both amplitude and phase spectra are necessary. As the example in section 3.5 shows, although a white noise and an impulse have identical amplitude spectra, they sound quite differently due to the difference in the phase spectra. Sometimes the onset and offset of a tone might be important for timbre (e.g., the decay of a piano tone). Both the steady-state and the transient characteristics of a timbre appear in a spectrogram (i.e., the spectrum developing with time), as shown in Fig. 3-19.

3.11 Binaural and Spatial Hearing

What is the advantage of having two ears? One obvious advantage is a backup: if one ear is somehow damaged, there is another one to use, a similar reason to having two kidneys. This explanation is definitely incomplete. In hearing, having two ears gives us many more advantages. Because of having two ears, we can localize sound sources, discriminate sounds originated from different locations, hear conversations much more clearly, and be more immune to background noises.

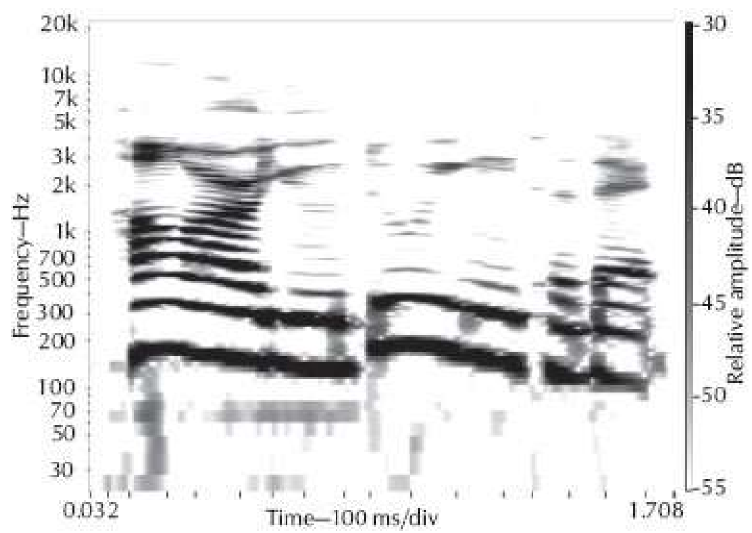


Figure 3-19. An example of a spectrogram: male voice of "How are you doing today?" The vertical axis is frequency, the horizontal axis is time, and the darkness of a point represents the level of a

particular frequency component at a given time.

3.11.1 Localization Cues for Left and Right

When a sound source is on the left with respect to a listener, it is closer to the left ear than to the right ear. Therefore the sound level is greater at the left ear than that at the right ear, leading to an *interaural level difference* (*ILD*). Sometimes, people also use the term "*interaural intensity difference* (IID)" to describe the same quantity. Moreover, because the sound wave reaches the left ear earlier than the right ear, there is an *interaural time difference* (*ITD*) between the two ears. However, the auditory neurons do not directly compare ITD, and instead, they compare the *interaural phase difference* (*IPD*). For a pure tone, *ITD* and *IPD* are linearly related. The *ITD* and *IPD* are also referred to as the *interaural temporal difference*. In summary, for localization of left and right, there are two cues, i.e., the *ILD* and *ITD* cues.

Adjusting either the *ILD* or the *ITD* cues can affect sound localization of left and right. In reality, both of those cues vary. There are limits for both cues. In order to better localize using the *ILD* cues, the interaural differences should be greater. Because of diffraction around the head, at frequencies below 1kHz, the *ILD* cues do not discriminate well among different azimuths. Therefore, *ILD* cues are utilized at high frequencies, when the head shadow has a big effect blocking the contralateral ear (the ear not pointing at the source). On the other hand, there is a limit for the *ITD* cues as well. At frequencies above 700Hz, the *IPD* of a source at extreme left or right would exceed 180°. For a pure tone, this would lead to confusion: a tone far to the right might sound to the left, as shown in Fig. 3-20. Since we care most about the sound sources in front of us, this limit of 700Hz can be extended upward a little bit. Furthermore, with complex signal with broad bandwidth, we can also use the time delay (or phase difference) of the low-frequency modulation. In general, the frequency of 1.2kHz (or a frequency range between 1 and 1.5kHz) is a good estimate for a boundary, below which *ITD* cues are important, and above which *ILD* cues are dominant.

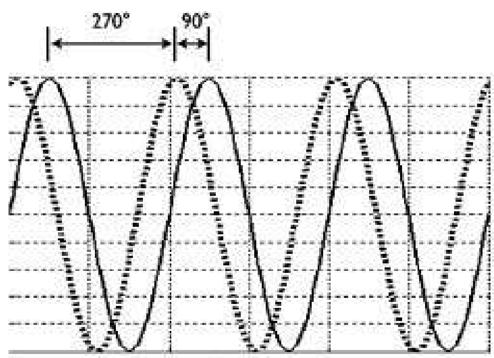


Figure 3-20. Confusion of interaural phase difference (*IPD*) cues at high frequencies. The dashed curve for the left ear is lagging the solid curve for the right ear by 2 70°, but it is heard as the left ear is leading the right ear by 90°.

In recording, adjusting the *ILD* cues is easily achieved by panning between the left and right channels. Although adjusting *ITD* cues also move the sound image through headphones, when listening through loudspeakers, the *ILD* cues are more reliable than *ITD* cues with respect to the loudspeaker positions.

3.11.2 Localization on Sagittal Planes

Consider two sound sources, one directly in front of and one directly behind the head. Due to symmetry, the *ILD* and *ITD* are both zero for those sources. Thus, it would seem that, by using only

ILD and *ITD* cues, a listener would not be able to discriminate front and back sources. If we consider the head to be a sphere with two holes at the ear-positions (the spherical head model), the far-field sources producing a given *ILD* and *ITD* all locate on the surface of a cone as shown in Fig. 3-21. This cone is called the *cone of confusion*.[46] If only *ILD* and *ITD* cues are available for a listener with a spherical head, he or she would not be able to discriminate far-field sound sources on a cone of confusion. Of course, the shape of a real head with pinnae is different from the spherical head, which changes the shape of the cone of confusion, but the general conclusion still holds. For a near-field source, when both *ILD* and *ITD* cues are available, the listener can further limit the confusion into a certain cross-section of the cone (i.e., the dark “donut” in Fig. 3-21). That is the best one can do with *ILD* and *ITD* cues. However, in reality, most people can easily localize sound sources in front, in the back, and above the head, etc, even with eyes closed. We can localize sources in a *sagittal plane* (a vertical plane separating the body into, not necessarily equal, left and right parts) with contribution of the asymmetrical shape of our pinnae, head, and torso of our upper body. The pinnae are asymmetrical when looked at from any direction. The primary role of the pinna is to filter or create spectral cues that are virtually unique for every angle of incidence. Different locations on the cone of confusion will be filtered differently, producing spectral cues unique to each location.

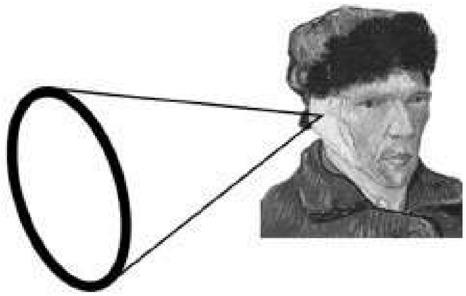


Figure 3-21. Cone of confusion for a spherical head with two holes at the ear positions. If only *ITD* cues are available, the listener cannot discriminate positions on the surface of the cone of confusion, corresponding to a given ITD. If *ILD* cues are also available, due to the diffraction of the head, the listener can further limit the confusion range into a circle (the dark "donut" on the figure).

The common way of describing the spectral cue for localization is the *head-related transfer function* (*HRTF*), an example of which is shown in Fig. 3-22. It is the transfer function (gain versus frequency), illustrating the filtering feature of the torso, head, and outer ear, for each location in space (or, more often, for each angle of incidence). Nowadays, with probe microphones inserted close to the eardrum, *HRTF* can be measured with high accuracy. Once it is obtained for a given listener, when listening to a recording made in an anechoic chamber processed with the proper *HRTF* (i.e. convolved with the head-related impulse response corresponding to the proper *HRTF*), one can "cheat" the auditory system and make the listener believe that the recording is being played from the location corresponding to the *HRTF*.

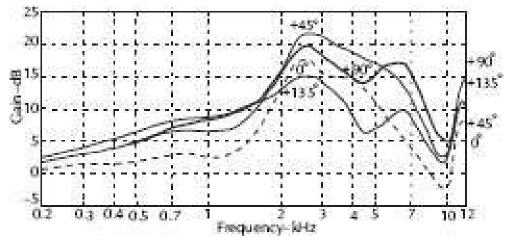


Figure 3-22. An example of head-related transfer functions. The source is in the horizontal plane (defined by both ears and the nose) and the angles are measured from the forward direction of the listener (straight ahead). The source is on the same side of the head as the measured ear (ipsilat-eral).[5]

There are two challenges when using spectral cues. The first is discriminating between the filtering feature and the spectrum of the source. For instance, if one hears a notch around 9kHz, it might be due to an *HRTF*, or the original source spectrum might have a notch around 9kHz. Unfortunately there is no simple way to discriminate between them. However, for a familiar sound (voice, instruments, etc.) with a spectrum known to the auditory system, it is easier to figure out the *HRTF*s and thus easier to localize the source than an unknown sound. If one has trouble discriminating sounds along the cone of confusion, one can use the cues of head motion. For example, suppose a listener turns his or her head to the left. If the source moves to the right, the source is in front; whereas if the source moves farther to the left, it must be in the back. The second challenge is the individuality of *HRTF*s. No two people share the same pinna and head shape, and we have learned our own pinnae and head size/shape over years of experience. If one listens to sounds processed with the *HRTF*s of someone else, although the left–right localization will be good, there will be a lot of front–back confusion,[47] unless the listener's head and ears happen to be similar in size and shape to those whose *HRTF* is measured.[48] The human binaural system is remarkably adaptive. Experiments with ear molds[49] show that, if a subject listens exclusively through another set of ears, although there is originally a lot of front–back confusion, in about 3 weeks, the subject will learn the new ears and localize almost as well as with their original ears. Instead of forgetting either the new or the old ears, the subject actually memorizes both sets of ears, and becomes in a sense "bi-pinneal", and is able to switch between the two sets of ears.

3.11.3 Externalization

Many listeners prefer listening to music through loudspeakers instead of through headphones. One of the reasons is that when listening through headphones, the pinnae are effectively bypassed, and the auditory system is not receiving any of the cues that the pinnae produce. Over headphones, the instruments and singers' voices are all perceived or localized inside the head. When listening through loudspeakers, although the localization cues are not perfect, the sounds are externalized if not localized, somewhat more naturally. If, however, music playing through the headphones includes the *HRTF*s of the listener, he or she should be able to externalize the sound perfectly.[50] Algorithms are available to simulate 3D sound sources at any location in free field and in a regular room with reverberation. The simulation is accurate to up to 16kHz, and listeners cannot discriminate between the real source and the virtual (simulated) sound.[51,52] An inconvenience nevertheless is that the system has to be calibrated to each listener and each room. In 1985, Jones et al.[53] devised a test for stereo imagery utilizing a reverberator developed at the Northwestern University Computer Music Studio. The reverberator utilized

*HRTF*s to create very compelling simulations of 3D space and moving sound sources within 3D space. The test by Jones et al.,[53] called LEDR (Listening Environment Diagnostic Recording) NU™, contained sound examples that moved in very specific sound paths. When played over loudspeaker systems that were free from phase or temporal distortions and in environments free from early reflections, the paths were perceived as they were intended. In the presence of early reflection or misaligned crossovers or drivers, the paths are audibly corrupted.

3.11.4 Precedence (Hass) Effect

When two clicks are presented simultaneously to a listener, one on the left and one on the right, the listener would perceive a click in front, i.e., average the localization cues of the two clicks. However, if one of the clicks is delayed (up to 5ms) compared to the other, the listener still perceives them as one fused click but will localize the fused image with cues of the first click only and ignore the localization cue of the later one. For delays longer than 5ms, the listener will hear two distinct clicks instead of one fused click. For speech, music, or other complex signals, this upper limit can be increased to about 40 ms. This phenomenon that the auditory system localizes on the first arrival is called *precedence effect*, or *Haas effect*.[54,55]

The precedence effect has very practical uses in audio. For example, in a large church, it may not be possible or practical to cover the entire church from one loudspeaker location. One solution is to place a primary loudspeaker in the front of the church, with secondary loudspeakers along the side walls. Because of the precedence effect, if the signal to the loudspeakers along the walls is delayed so that the direct sound from the front arrives at a listener first, the listener will localize the front loudspeaker as the source of the sound, even though most of the content will actually be coming from the loudspeaker to the side, which is much closer to the listener, and may even be producing a level at the listener's head that is as much as 10dB greater. When such systems are correctly set up, it will sound as though the secondary loudspeakers are not even turned on. Actually turning them off demonstrates exactly how important they are, as without them the sound is unacceptable, and speech may even be unintelligible.

3.11.5 Franssen Effect

The *Franssen Effect*[56] can be a very impressive demonstration in a live room. A pure tone is played through two loudspeakers at two different locations. One loudspeaker plays the tone first and is immediately faded, while the same pure tone is boosted at the other loudspeaker, so that the overall level is not changed significantly, as shown in Fig. 3-23. Although the original loudspeaker is not playing at all, most of the audience will still believe that the sound is coming from the first loudspeaker. This effect can last for a couple of minutes. One can make this demonstration more effectively by disconnecting the cable to the first loudspeaker, and the audience will still localize the sound to that loudspeaker. The Franssen effect reveals the level of our auditory memory of source locations in a live room.

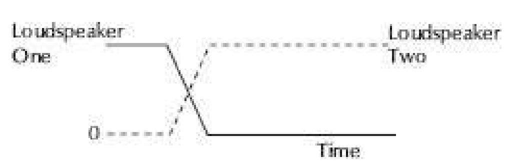


Figure 3-23. Franssen effect.[56] The figure shows the level of two

loudspeakers at two difference locations in a live room. Loudspeaker One plays a pure tone first, and is immediately faded. Meanwhile, the same tone played by Loudspeaker Two is boosted, so that the overall level in the room is not changed significantly. After Loudspeaker One stops playing, listeners will still perceive the sound originated from Loudspeaker One, up to a couple of minutes.

3.11.6 Cocktail Party Effect and Improvement of Signal Detection

In a noisy environment, such as a cocktail party, many people are talking simultaneously. However, most people have the ability to listen to one conversation at a time, while ignoring other conversations going on around them. One can even do this without turning his or her head to the attended talker. As we mentioned earlier, one benefit of binaural hearing is the ability to spatially filter. Because the talkers are spatially separated, our auditory system can filter out unwanted sound spatially. People with hearing difficulties usually suffer greatly in a noisy environment because they are unable to pick up an individual's conversation out of the background.

The binaural masking level difference is a particularly strong effect when background noise is in phase across the two ears and the signal is binaurally 180° out of phase. For narrowband signals the reduction in masking by the background can be as much as 25dB. So, in general, binaural hearing not only gives us localization ability, but also improves our ability to detect an acoustic signal, especially in a noisy or reverberant environment.

3.11.7 Distance Perception

Distance cues are fairly difficult to replicate. In free field conditions, the sound pressure level will decrease 6dB with every doubling of the distance between a point source and an observer. Thus reducing the level should make us feel the source is farther away. In practice, however, we tend to underestimate the distance: the level has to be attenuated by 20dB in order to give us the perception of a doubled distance.[57] Of course if we do not know how loud the original source is, we do not have an absolute scale based on level.

When a sound source is very far, because the air absorbs high-frequencies more than the low-frequencies, the perceived sound would contain more low-frequency energy, with a darker timbre. This is why thunder far away is just rumble whereas thunder nearby has a crack to it. However, this is a very weak effect[58] and therefore is relatively insignificant for events nearby, which is mostly the case in everyday life.

A more compelling cue for replicating and perceiving distance is adjusting the ratio of the direct to reverberant sound. In real spaces, a sound nearby will not only be louder but also will have a relatively high direct-to-reverberant ratio. As the sound moves away, it gets quieter, and the direct-to-reverberant ratio reduces until critical distance is reached. At critical distance the direct and reverberant levels are equal. Moving a sound source beyond critical distance will not result in an increased sense of distance.

3.12 Auditory Scene Analysis

The aural environment that we live in includes sounds from multiple sources. Unlike vision, determining the contribution from each sound source can be a challenge, as they are all “mixed” into the acoustic waveform that is received by our ears. By complex processes of auditory scene analysis,[59] the auditory system is usually able to segregate sounds from different sources and group together those from the same source.

When multiple sources are present simultaneously, the frequency components with the same onset or those that form a harmonic series tend to be grouped together, a process that is known as simultaneous grouping. When different sounds are heard in sequence, the auditory system may group the sounds into separate auditory streams or the same auditory stream. This grouping is known as sequential grouping or auditory streaming. Sounds with similar spectra tend to be grouped together. For instance, when the notes from a melody are interleaved with the notes from another, and the frequencies from one melody are much higher than those from the other, a listener can hear the independent melodies. Segregation into separate streams occurs more often if the alternation of the tones is rapid.

It is common to experience a sound being interrupted by another sound. If the interruption is short, the auditory system can restore the part of the original sound that is masked (or even missing) by the interruption, and the listener hears a continuous sound. This phenomenon is known as perceptual restoration.[60] As an example, if a recording being played back is interrupted with frequent gaps, a listener would hear a discontinuous recording. But if the gaps are filled with broadband noise, the listener hears a continuous recording, even though it is accompanied by pulsed noise.[61]

3.13 Speech

3.13.1 Speech Organ

The vocal folds (the two pieces of muscular mucous membrane in the larynx, also known as the “vocal cords”) produce a broadband buzz sound, which is filtered by the vocal tract (including the throat, mouth, and nasal cavity) to produce various spectra, especially for vowels. The tongue plays an important role in changing the shape of the oral cavity and in producing certain consonants (e.g., /t/). For adults, the fundamental frequency of vocal folds in natural speech is between 110Hz (male voice) and 220Hz (female voice), and it varies among individuals. This fundamental frequency changes in singing to produce melodies, and in ordinary speech to convey meanings.

3.13.2 Vowels and Consonants

Language communication is very important for humans. There are many different languages in the world, and even more regional dialects. Although their vocabularies and grammars are all different, speech perception in all languages comes down to the perception of vowels and consonants.

Vowels are the phonemes (i.e., the smallest unit in speech) produced with an open vocal tract. Vowels are normally voiced, i.e., the focal folds vibrate while pronouncing vowels. (In whispering, vowels are unvoiced.) They are relatively long and constitute the greatest portion of speech duration. Voiced vowels are complex tones producing clear pitches, thus it is the vowels that make the melody in singing. On a spectrogram as in Fig. 3-19, the vowels show a clear harmonic structure. Besides pure vowels, there are diphthongs, which combine multiple vowels into one phoneme (e.g., /ei/). Table 3-2 shows the vowels in English. Other languages may have more or fewer vowels, and some languages (e.g., French) have nasal vowels that are not present in English.

Table 3-2. The Vowels in English

Pure Vowel			
Phoneme	Word	Phoneme	Word
/ɑː/	father	/iː/	seat
/ʌ/	cut	/ɪ/	bit
/æ/	ash	/ɔː/	talk
/ɛ/	bed	/ɒ/	top
/uː/	goose	/ɜː/	work
/ʊ/	put	/ə/	paper
Diphthong			
Phoneme	**Word**	**Phoneme**	**Word**
/aɪ/	like	/əʊ/	bone
/eɪ/	take	/ɪə/	ear
/ɔɪ/	boy	/ɛə/	care
/aʊ/	cow	/ʊə/	poor

Consonants are phonemes that are articulated by complete or partial closure of a vocal tract. They are short and have to combine with vowels to produce a syllable. (A syllable is an element consisting one or more phonemes, and one or more syllables make a word.) By contrast, a vowel can make a syllable by itself. By using consonants, with the same number of vowels, many more syllables can be produced, which makes communication much more efficient. On a spectrogram, consonants normally appear as broadband signals without clear harmonic structure. By forms of articulation, consonants can be categorized into plosives (blocking the airflow), fricatives (forcing air through a narrow channel), nasals (making use of the nasal cavity), liquids (/l/ and /r/), and glides (/j/ and /w/, also called "semivowels", which are similar to the vowels /i/ and /u/, but are short and do not function as vowels, e.g., the "y" in the word "yes" and the "w" in the word "wood"). Nasals, liquids, and glides are all voiced, while plosives and fricatives normally appear in pairs of voiced and unvoiced (e.g., /b/

and /p/, and /v/ and /f/). Table 3-3 shows the consonants in English. In English, normally unvoiced plosives (/p/, /t/, and /k/) are aspirated (produced with a strong burst of air), and voiced plosives (/b/, /d/, and /g/) are unaspirated (produced with little burst of air). But this may not be the case in other languages. For instance, in Latin languages (e.g., Spanish, French and Italian), with a few exceptions, both voiced and unvoiced plosives are unaspirated (similar to the "t" in the English word "stop"). In mandarin Chinese, on the other hand, all of the plosives are unvoiced, but they appear in pairs of aspirated and unaspirated.

3.13.3 Formants and Speech Intelligibility

On a spectrum, the peaks of the spectral envelope are defined as formants.[62] As shown in Fig. 3-24, the formants are labeled by the order of their frequencies, F1, F2, F3, etc., with F1 being the lowest. For vowels, normally three or more formants can be identified. The frequencies of these formants determine which vowel is heard, and the lowest two (F1 and F2) are most important. The consonants either modify the placement of adjacent vowels or add extra formants. Table 3-4 shows the lowest three formants of some vowels. Interestingly, experiments show that after a short training a subject can understand a synthetic speech replacing formants with sine-tones.[63]

The frequency band between 300Hz and 4kHz is important for speech intelligibility. A listener would have a hard time identifying the words if this frequency band was removed. Speech intelligibility is by itself an important topic, which is studied in a separate chapter in this handbook.

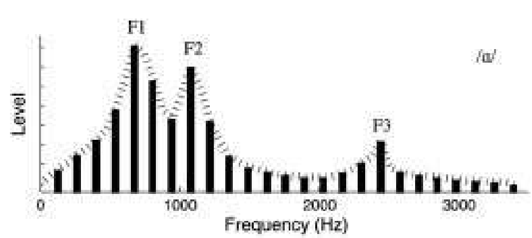
Table 3-3. The Consonants in English

Plosive			
Unvoiced		Voiced	
Phoneme	Word	Phoneme	Word
/p/	put	/b/	book
/t/	top	/d/	dog
/k/	kind	/g/	good

Fricative			
Unvoiced		Voiced	
Phoneme	Word	Phoneme	Word
/f/	find	/v/	voice
/θ/	thin	/ð/	this
/s/	see	/z/	zoo
/ʃ/	ship	/ʒ/	pleasure
/tʃ/	chair	/dʒ/	just

Liquid		Glide (semivowel)	
Phoneme	Word	Phoneme	Word
/l/	leg	/j/	yes
/r/	read	/w/	we

Nasal	
Phoneme	Word
/m/	man
/n/	no
/ŋ/	sing



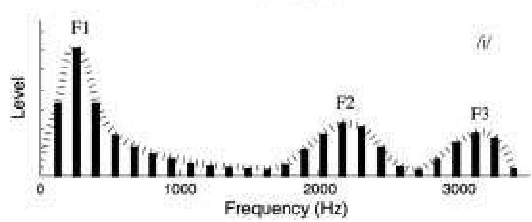


Figure 3-24. Spectra of the vowels /*a*/ and /i/.

Table 3-4. The Formant Frequencies (Hz) of Some Vowels Averaged for 76 Talkers[64]

Phoneme		/ɑ/	/ʌ/	/æ/	/ɛ/	/u/
Men	F1	730	640	660	530	300
	F2	1090	1190	1720	1840	870
	F3	2440	2390	2410	2480	2240
Women	F1	850	760	860	610	370
	F2	1220	1400	2050	2330	950
	F3	2810	2780	2850	2990	2670
Children	F1	1030	850	1010	690	430
	F2	1370	1590	2320	2610	1170
	F3	3170	3360	3320	3570	3260
Phoneme		/ʊ/	/i/	/ɪ/	/ɔ/	/ə/
Men	F1	440	270	390	570	490
	F2	1020	2290	1990	840	1350
	F3	2240	3010	2550	2410	1690
Women	F1	470	310	430	590	500
	F2	1160	2790	2480	920	1640
	F3	2680	3310	3070	2710	1960
Children	F1	560	370	530	680	560
	F2	1410	3200	2730	1060	1820
	F3	3310	3730	3600	3180	2160

3.13.4 Singing Voice

Especially in classical opera, singers have to perform without help from a sound reinforcement system. The energy from a trumpet can easily overwhelm a human voice. However, great singers can project to the last row of the audience while the whole orchestra is playing. This is achieved by emphasizing a strong formant around 3kHz (also called the “singer’s formant”), where the energy of orchestra is relatively lower.[65] This technique produces a "bright" voice, which is not only aesthetically beautiful, but also practically useful, making lyrics audible in presence of accompaniment.

3.13.5 Tonal Languages

While the tone, i.e., the fundamental frequency contour of vowels, in English may represent the speaker's emotion and mood, etc., it does not function semantically. However, in tonal languages such as Chinese and Thai, a syllable is considered to be different with different tones. In mandarin Chinese, besides a neutral tone, there are four tones, while Cantonese and Thai can have as many as nine tones.

3.13.6 Influence of Visual Cues

Because the acoustic signal one hears can be easily altered by environment noise, visual cues are found to be very helpful. Lip reading is an example. As shown in the McGurk effect,[66] when listeners see a mouth movement of /ɡɑː/ while presented with an audio of /bɑː/, they normally hear the syllables of /ɡɑː/ or /dɑː/, and very few would actually hear /bɑː/. This shows that visual cues can sometimes be as important as auditory cues, especially for the consonants that are difficulty to discriminate.

3.14 Remarks

Psychoacoustics is an interdisciplinary field of many areas. In order to quantitatively study human perception of sound, including subjective aspects, statistical analysis is needed after experiments with many subjects. The auditory system is very adaptive, and it develops as we grow up. Because each subject is different, individuality is an important factor to consider in summarizing meaningful conclusions. Moreover, psychoacoustics is a developing field. As new phenomena continue to be discovered, our understanding of hearing also improves accordingly. New models are being developed and existing models are modified or even rejected. In applications, many of the findings in psychoacoustics have been used in other related fields, and can be found in other chapters in this handbook.

Further Reading

Textbooks on psychoacoustics and auditory physiology are available for various audiences. One might find some of the following books to be helpful:

J. Blauert, "Spatial hearing: The psychophysics of human sound localization," Revised ed., MIT Press, Cambridge (1996).

J. D. Durrant and J. H. Lovrinic, "Bases of hearing science," 3rd ed, Williams and Wilkins, Baltimore, MD (1995).

H. Fastl and E. Zwicker, "Psychoacoustics: facts and models," 3rd ed, Springer, NY (2006).

W. M. Hartmann, "Signals, sound and sensation," Springer, NY (1998).

B. C. J. Moore, "An introduction to the psychology of hearing," 5th ed, Academic Press, London (2003).

J. O. Pickles, "An introduction to the physiology of hearing," 2nd ed, Academic Press, London (1988).

C. J. Plack, "The sense of hearing," Lawrence Erlbaum Associates, NJ (2005).

W. A. Yost, "Fundamentals of hearing: An introduction," 5th ed, Academic Press, NY (2006).

References

1. Douglas Jones, personal communications (2007).
2. J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, **106**, 1493-1510 (1999).

3. H. Haas, “Über den Einfluss eines Einfachechos an die Hörsamkeit von Sprache, *Acustica*,” **1**, 49-58 (1951).

4. F. M. Weiner and D. A. Ross, “Pressure distribution in the auditory canal in a progressive sound field,” *J. Acoust. Soc. Am.*, **18**, 401-408 (1946).

5. E. A. G. Shaw, “The external ear,” *Handbook of sensory physiology*, ed. W.D. Keidel and W.D. Neff, **vol. 5/1**, pp. 455–490, Springer, Berlin (1974).

6. J. D. Durrant and J. H. Lovrinic, “Bases of hearing science,” 3rd ed, p164, Williams and Wilkins, Baltimore, MD (1995).

7. V. Nedzelnitsky, “Sound pressures in the basal turn of the cat cochlea,” *J. Acoust. Soc. Am.*, **68**, 1676–1689 (1980).

8. J. O. Pickles, “An introduction to the physiology of hearing,” 2nd ed, pp. 16–20, Academic Press, London (1988).

9. M. Portmann and G. Lacher, “Present management of the reconstruction of the middle ear,” *Clinical Otolaryngology*, **12** (5), 389–395 (1987).

10. US Patent, “Communication device using bone conduction,” US Patent 6885753.

11. P. W. Carmel and A. Starr, “Acoustic and non-acoustic factors modifying middle-ear muscle activity in waking cats,” *J. Neurophysiol*, **26**, 598–616 (1963).

12. J. D. Durrant and J. H. Lovrinic, “Bases of hearing science,” 3rd ed, p172, Williams and Wilkins, Baltimore, MD (1995).

13. G. von Békésy, “Experiments in hearing,” McGraw-Hill, NY (1960).

14. M. C. Liberman and L. W. Dodds, “Single neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,” *Hear. Res.*, **16**, 55–74 (1984).

15. C. J. Plack, “The sense of hearing,” Lawrence Erlbaum, NJ (2005).

16. M. A. Ruggero, N. C. Rich, A. Recio, S.S. Narayan, and L. Robles, “Basilar-membrane responses to tones at the base of the chinchilla cochlea,” *J. Acoust. Soc. Am.*, **101**, 2151–2163 (1997).

17. American Standards Association, “Acoustical Terminology SI,” pp. 1–1960, American Standards Association, NY (1960).

18. H. Fletcher, “Auditory patterns,” *Rev. Mod. Phys.*, **12**, 47–65 (1940).

19. W. Jesteadt, S. P. Bacon, and J. R. Lehman, “Forward masking as a function of frequency, masker level, and signal delay,” *J. Acoust. Soc. Am.*, **71**, 950–962 (1982).

20. B. R. Glasberg and B. C. J. Moore, “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.*, **47**, 103–138 (1990).

21. E. Zwicker and E. Terhardt, “Analytical expressions for critical band rate and critical bandwidth as a function of frequency,” *J. Acoust. Soc. Am.*, **68**, 1523–1525 (1980).

22. D. T. Kemp, “Stimulated acoustic emissions from within the human auditory system,” *J. Acoust. Soc. Am.*, **64**, 1386–1391 (1978).

23. D. H. Johnson, “The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,” *J. Acoust. Soc. Am.*, **68**, 1115–1122 (1980).

24. C. Formby and T. G. Forrest, “Detection of silent temporal gaps in sinusoidal markers,” *J. Acoust. Soc. Am.*, **89**, 830–837 (1991).

25. F. E. Toole, “Loudness – Applications and Implications to Audio,” *dB The Sound Engineering Magazine*, Part I, vol. **7**, no. 5, pp.27–30, May 1973; Part II, **vol. 7**, no. 6, pp.25–28, June 1973.

26. ANSI S1.4–1983, ANSI S1.42–2001, IEC 61672–1:2002.

27. R. P. Hellman, “Growth of loudness at 1000 and 3000 Hz,” *J. Acoust. Soc. Am.*, **60**, 672–679 (1976).

28. G. A. Miller, “Sensitivity to changes in the intensity of white noise and its relation to masking and loudness,” *J. Acoust. Soc. Am.*, **19**, 609–619 (1947).

29. W. J. McGill and J. P. Goldberg, “A study of the near-miss involving Weber’s Law and pure-tone intensity discrimination,” *Percept. Psychophys.*, **4**, 105–109 (1968).

30. N. F. Viemeister and S. P. Bacon, “Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones,” *J. Acoust. Soc. Am.*, **84**, 172–178 (1988).

31. B. C. J. Moore and D. H. Raab, “Pure-tone intensity discrimination: some experiments relating to the ‘near-miss’ to Weber’s Law,” *J. Acoust. Soc. Am.*, **55**, 1049–1054 (1974).

32. American National Standards, “Acoustical Terminology SI,” American National Standards Institute, New York (1994).

33. W. M. Hartmann, “Signals, sound and sensation,” Springer, NY (1998).

34. F. Attneave and R. K. Olson, “Pitch as a medium: a new approach to psychophysical scaling,” *Am. J. Psychol.*, **84**, 147–166 (1971).

35. S. S. Stevens, J. Volkman, and E. B. Newman, “A scale for the measurement of the psychological magnitude of pitch,” *J. Acoust. Soc.*

Am., **8**, 185–190 (1937).

36. J. F. Schouten, “The residue and the mechanism of hearing,” *Proc. Kon. Akad. Wetenschap.*, **43**, 991–999 (1940).

37. J. F. Schouten, “The residue revisited,” *Frequency Analysis and Periodicity Detection in Hearing*, ed. R. Plomp and G. Smoorenburg, A. W. Sijthoff, Leiden, pp. 41–58 (1970).

38. B. C. J. Moore, B. R. Glasberg, and R. W. Peters, “Relative dominance of individual partials in determining the pitch of complex tones,” *J. Acoust. Soc. Am.*, **77**, 1853–1859 (1985).

39. A. J. M. Houtsma and J. L. Goldstein, “The central origin of the pitch of pure tones: evidence from musical interval recognition,” *J. Acoust. Soc. Am.*, **51**, 520–529 (1972).

40. D. H. Johnson, “The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones,” *J. Acoust. Soc. Am.*, **68**, 1115–1122 (1980).

41. C. C. Wier, W. Jesteadt, and D. M. Green, “Frequency discrimination as a function of frequency and sensation level,” *J. Acoust. Soc. Am.*, **61**, 178–184 (1977).

42. E. Terhardt, G. Stoll, and M. Seewann, “Algorithm for extraction of pitch and pitch salience from complex tonal signals,” *J. Acoust. Soc. Am.*, **71**, 679–688 (1982).

43. J. C. R. Licklider, “Auditory frequency analysis,” Information Theory, ed. C. Cherry, pp. 253–268, Academic, NY (1956).

44. E. M. Cramer and W. H. Huggins, “Creation of pitch through binaural interaction,” *J. Acoust. Soc. Am.*, **30**, 413–417 (1958).

45. F. A. Bilsen and J. Raatgever, “On the dichotic pitch of simultaneously

presented interaurally delayed white noises. Implications for binaural theory," *J. Acoust. Soc. Am.*, **108**, 272–284 (2000).

46. A. W. Mills, "Auditory localization," *Foundations of Modern Auditory Theory*, **Vol. II**, ed. J. V. Tobias, Academic Press, NY (1972).

47. F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: stimulus synthesis," *J. Acoust. Soc. Am.*, **85**, 858–867 (1989).

48. J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, **106**, 1480–1492 (1999).

49. M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.*, **53**, 400–408 (1973).

50. W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, **99**, 3678–3688 (1996).

51. P. X. Zhang and W. M. Hartmann, "Transaural virtual reality applied to front-back localization," *J. Acoust. Soc. Am.*, **117**, 2485–2486 (2005).

52. P. X. Zhang, "Detection and localization of sounds: virtual tones and virtual reality," Thesis, Michigan State University, East Lansing (2006).

53. D. R. Jones, W. L. Martens, and G. S. Kendall, "Optimizing control rooms for stereo imagery," *J. Acoust. Soc. Am.*, **78**, S9 (1985).

54. H. Haas, "The Influence of a Single Echo on the Audibility of Speech," *Acustica*, **vol. 1**, no. 2, 1951, in German. (An English translation by Dr. Ingr.K. P. R. Ehrenberg appears in the *Journal of the Audio Engineering Society*, **vol. 20**, no. 2, pp.146–159, March 1972.)

55. R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The

precedence effect," *J. Acoust. Soc. Am.*, **106**, 1633–1654 (1999).

56. W. M. Hartmann and B. Rakerd, "Localization of sound in rooms IV: The Franssen effect," *J. Acoust. Soc. Am.*, **86**, 1366–1373 (1999).

57. J. Blauert, "Spatial hearing: The psychophysics of human sound localization," pp. 122–123, MIT Press, Cambridge (1997).

58. U. Ingard, "A review of the influence of meteorological conditions on sound propagation," *J. Acoust. Soc. Am.*, **25**, 405–411 (1953).

59. A. S. Bregman, "Auditory scene analysis: the perceptual organization of sound," MIT Press, Cambridge (1990).

60. R. M. Warren, "Auditory perception: a new analysis and synthesis," Cambridge University Press, Cambridge (1999).

61. T. Houtgast, "Psychophysical experiments on 'tuning curves' and 'two-tone inhibition'," *Acoustica*, **29**, 168–179 (1973).

62. G. Fant, "Acoustic Theory of Speech Production," Mouton, The Hague (1960).

63. R. E. Remez, P. E. Rubin, D. B. Pisoni, and T. D. Carell, "Speech perception without traditional speech cues," *Science*, **212**, 947–950 (1981).

64. G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.*, **24**, 175–184 (1952).

65. J. Sundberg, "Articulatory interpretation of the 'singing formant'," *J. Acoust. Soc. Am.*, **55**, 838–844 (1974).

66. H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, **264**, 746–748 (1976).

Chapter 4

Hearing Physiology—Disorders—Conservation

by S. Benjamin Kanters

4.1 Introduction

As practitioners in the audio arts and sciences, we endeavor to understand sound, its behavior in space, analog and digital signals, and the technologies of transducers and signal processors. Our livelihoods rely on our skills to capture and deliver sound with the highest quality, and within the limits dictated by the medium or the client.

To make many, and sometimes all, of our decisions, we listen. We listen critically to determine whether a microphone choice is appropriate, equalization or compression is true to the voice or instrument, or if the voice or instrument is true to the composition. Acoustical and sound system designers will employ high-level math and digital analysis tools to design a system or room. But the final arbiter will be if it "sounds good."

For all that we know and continually strive to learn, we perhaps know the least about the one tool that is, arguably, the most important of any in our toolboxes, our hearing. It is incumbent upon those of us who listen professionally (including musicians) to better understand how our hearing mechanism actually works, and how to take care of our ears and our hearing acuity.

This chapter will build upon the Chapter 3 *Psychoacoustics* introduction to the hearing mechanism as the means to understand the mechanics of hearing loss and other disorders, then discuss strategies and current technologies to support hearing conservation and good hearing health.

4.2 The Hearing Mechanism—addendum to Chapter 3 Psychoacoustics

4.2.1 Hair Cells

Fig. 4-1 illustrates a close-up of the arrangement of inner and outer hair cells in the organ of corti. Note how the outer hair cells rest upon “supporting cells” (Dieter’s Cells). Also, unlike the inner hair cells, the tallest stereocilia of the outer cells are embedded in the gelatinous tectorial membrane.

As noted in section 3.2.5, there is one row of inner hair cells (about 3,500) and three rows of outer hair cells (about 11,500). The inner hair cells are the actual “listening cells,” similar to the rods and cones in the retina of the eye. These cells’ output to the auditory nerve is directly proportional to the deflection of the stereocilia. The auditory centers of the brain are receiving amplitude values from the inner hair cells for the different “pass-bands,” as determined by the basilar membrane; all-told, a function that could be equated with a realtime analyzer.

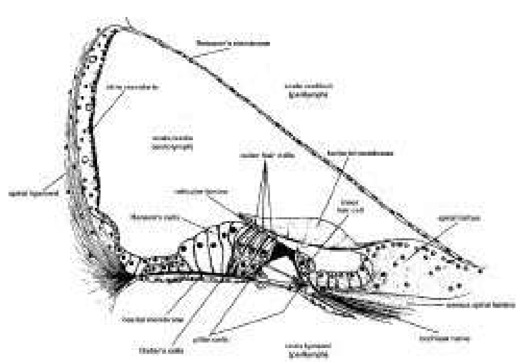


Figure 4-1. Scala media, the organ of Corti and hair cells.

Courtesy William Yost.

The outer hair cells have no sensory function. Instead they are central to our 120dB of dynamic range by providing mechanical amplification to low-level sound/stimulus. They are described in Chapter 3 *Psychoacoustics* as changing shape according to the vibrations of the basilar membrane. Often described as a “somatic motility” the cells’ flexing action increase displacement of the organ of corti, and thus, increasing stimulation of the inner hair cells. This is commonly referred to as “the cochlear amplifier,” providing gain as needed; more gain for quiet sounds, less for moderate, and none at all for loud. Regulation of this gain function is still being studied. One theory is that “inbound” nerve fibers provide controlling feedback from the brain. A more recent theory suggests that the motor function is regulated within each individual cell.

We enjoy approximately 120dB of dynamic range, from the threshold of hearing (20μPa) to the threshold of pain. Without outer hair cells, our threshold of hearing would be about 60dB SPL. The outer hair cells provide the gain necessary to hear the lower 60dB. From 0 to about 20dB SPL, the cells provide about 65dB of gain. Then from 20 up to about 100dB SPL, progressively less gain to the point that the cells essentially “shut off” and the action of the organ of corti is totally passive. This cochlear amplifier function could be characterized as an upward compressor.

4.2.2 Summary of the Hearing Mechanism

The outer ear structures direct sound energy to the tympanic membrane with added resonances and reflections that amplify phoneme and consonant frequencies and provide localization cues.

The middle ear provides a mechanical impedance-matching transformer, allowing efficient transfer of energy from acoustic to hydraulic. In the cochlea, the basilar membrane takes the displacement force (waveform) and breaks the signal down into frequency bands. Band-specific displacement energy is then delivered to the organ of corti, where the inner hair cells generate the neural signals that contain the amplitude information for each frequency band. Outer hair cells provide the necessary gain to low-level sound, giving us a total dynamic sensitivity of 120dB.

4.3 Measurement

4.3.1 Sound Level Measurement

To understand hearing disorders, one must first understand standards for measuring sound and hearing. Given the importance of *knowing* if one is in a noise-hazardous environment, a sound level meter (dB SPL) should be a critical part of every audio engineer's toolbox. We should be as sensitive to volume levels as many of us are to pitch and timbre. Reasonably accurate meters are now available for as little as $30. An example of such a meter is shown in Fig. 4-2. These are not accurate enough for research, but perfectly appropriate for general reference measurement. Another option is one of the many sound level meter applications available for many smart phones. Tests indicate that these apps are relatively accurate up to around 95 to 100dB SPL. However at higher sound levels, the microphones and/or input circuitry distort. The result is erroneous readings. That said, at those levels, the safe exposure limit is between 15 and 30 minutes and one should be thinking about using protection or just getting out of that environment.

These varieties of sound level meters are fine for measurement of sound levels at a single point in time, but ineffective for noise hazard assessment, since many measurements must be taken over the time period of exposure.

Figure 4-2. dBA meter.

4.3.2 Time-Based Measurement

Accurate noise hazard assessment must take into account not only the noise intensity but also the duration of noise exposure (see section 4.3.5 Safe Exposure Standards). Additionally, since many situations involve varying intensity (music is a perfect example) a mathematical average of the exposure over the exposure period must also be calculated to accurately assess the degree of the noise hazard. Fig. 4-3 is an example of a noise dosimeter.

Figure 4-3. Dosimeter. Courtesy Casella.

Once started, the dosimeter stores level readings at regular intervals (every 10 to 60s). After the measurement run, modern meters utilize a USB connection to export the readings to a spreadsheet table. In addition, many also have software to calculate one or all of the following time-based measurements: *LAeq, LAavg* and exposure "dose." *LAeq* (equivalent continuous sound level) and *LAav*g are both the average sound level over the period of measurement, *LAeq* is derived from highspeed response readings where *LAavg* is derived from slow-response measurements. Noise "dose" is a calculation that expresses the exposure as a percentage of a safe, single-day's exposure, as noted in either the NIOSH or OSHA scales below.

Fig. 4-4 illustrates a dosimetric study of a 3-hour NASCAR race. The graph shows the minute by minute dBA readings (note how noise level drops during all yellow-flag periods) as well as *LAeq, LAavg* and dose percentages based upon both NIOSH and OSHA standards, Fig. 4-6.

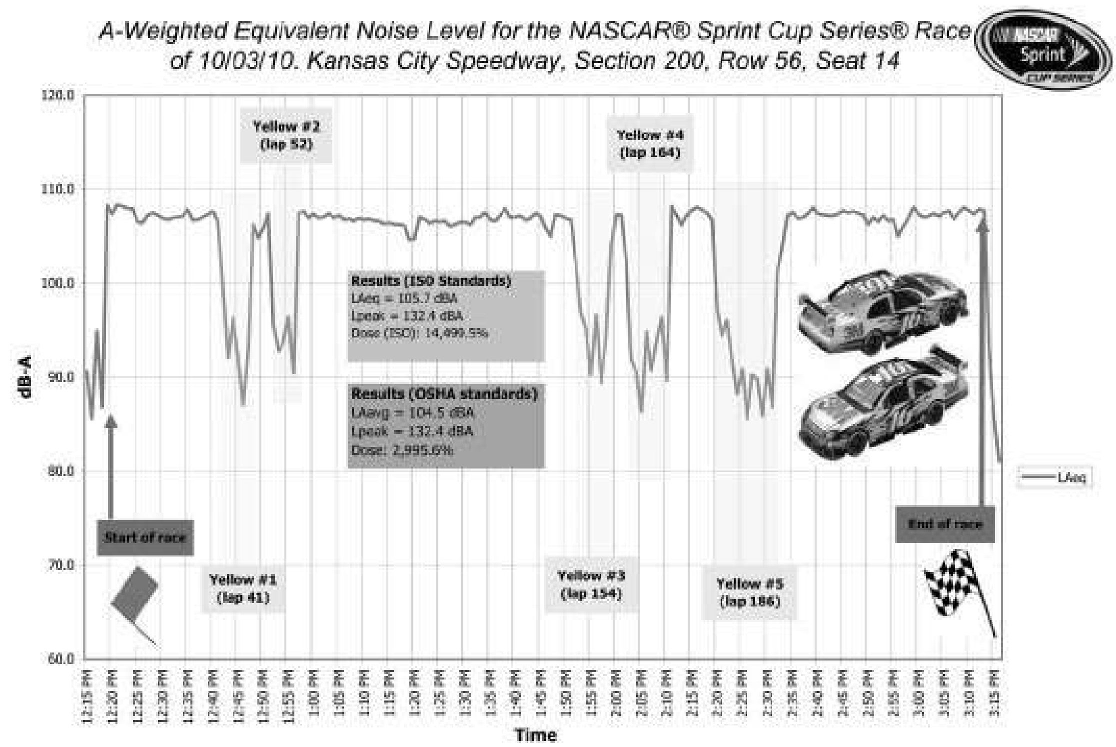


Figure 4-4. Dosimetric study of an auto race. Courtesy Dominique Cheenne, PhD.

4.3.3 dB Scale Weighting

Most sound level meters offer switchable A- and C-weighting. A detailed discussion of A, B and C weighting is given in Chapter 3 Psychoacoustics section 3.8.2. In the arena of hearing-based measurement, the A-scale is the standard, and has been adopted as the measurement for assessing noise exposure by NIOSH and OSHA. Thus, the term "dBA" is used when referring to SPL measurements with A-weighting.

4.3.4 Hearing Measurement

Fig. 4-5 is an example of a blank audiogram chart. The results of a hearing test are recorded here, on a dBHL (Hearing Level) scale. Playing the noted frequencies at different levels, the audiologist marks the chart ("X" for left ear, "O" for right), showing the patient's hearing thresholds at each of those frequencies. Readings at 0dB HL indicate that the subject can hear down to the clinical threshold of hearing. Readings down to 25dB HL are considered nominal. The audiogram, then, is a frequency response graph of one's hearing. Admittedly, the bandwidth is limited from the perspective of audio professionals. This is due to two factors:

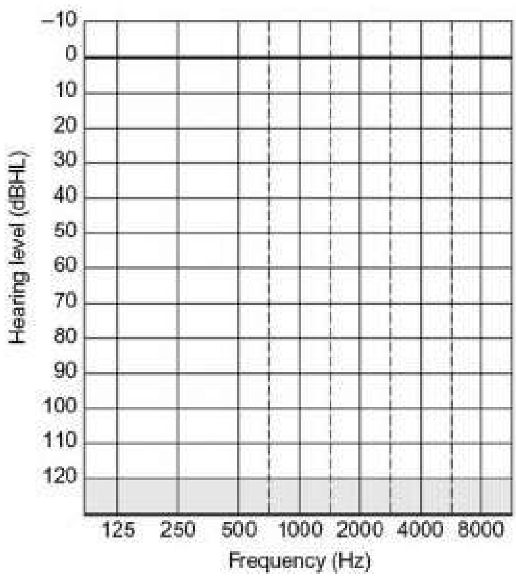


Figure 4-5. Blank Audiogram.

1. Traditional audiology focuses on speech communications.
2. There are no pure-tone sensitivity norms established for frequencies down to 20Hz or up to 20kHz.

Currently there are new testing protocols and technologies under development to test the full bandwidth of human hearing with a high degree of confidence.

4.3.5 Safe Exposure Standards

Fig. 4-6 shows safe exposure times for different levels, as determined through researched conducted by the National Institute of Occupational Safety and Health (NIOSH) and regulated by the Occupational Safety and Health Administration (OSHA). These numbers are the result of a study of 1172 workers from a cross section of industries, done between 1968 and 1972. The NIOSH scale represents exposure rates that were safe for 97% of the sample (i.e., 3% suffered some loss even at these levels). The OSHA scale represents safe exposure for 75% of the sample. The OSHA scale was derived as a more realistic target for industry compliance. For the rest of us, these scales show that everyone's physiology is different. We are not all equally susceptible and there is currently no way to determine an individual's susceptibility to noise trauma.

	NIOSH	OSHA
Exposure Time	dBA SPL	dBA SPL
8 hrs	85	90
4 hrs	88	95
2 hrs	91	100
1 hr	94	105
30 min	97	110
15 min	100	115
7.5 min	103	120
3.75 min	106	
> 2 min	109	
> 1 min	112	

Figure 4-6. NIOSH/OSHA safe exposure charts.

It is important to see that sustained exposure to even moderate noise can cause damage, just as short-term exposure to loud noise can be "safe". Those involved in efforts to combat noise induced hearing loss will always vie toward the safer, more stringent NIOSH scale.

4.4 The Mechanics of Hearing Loss

Many mistakenly believe that hearing loss is simply the loss of volume sensitivity across the entire frequency band. Worse than that, they also believe that, as one looses that sensitivity, all that is needed is amplification; the more you loose, the more amplification you need. In fact, the danger of over-exposure is always present and excessive exposure will always mean more loss. Not only that, but noise-induced loss is neither broad-band nor of "fixed attenuation." The distortions of hearing loss are both spectral and dynamic.

Certainly, in the most extreme cases, sudden, severe trauma in excess of 160dB SPL (explosions, etc.) can result in instantaneous damage to the organ of corti and hair cells. The focus of this chapter is, rather, the gradual loss experienced from repeated exposure to moderately high levels for long enough periods of time.

4.4.1 Dynamic Sensitivity Loss & Distortion

Moderate noise exposure results in stress, and eventual damage to the stereocilia of the hair cells. Fig. 4-7 shows healthy and increasingly stress-damaged stereocilia. Since they are responsible for "activating" the hair cells, stereocilia damage renders them non-functional. In cases of moderate trauma, this damage occurs primarily in the outer hair cells' stereocilia, as they are anchored in the tectorial membrane. The stereocilia of the more-important inner hair cells are not connected in the same way, and so are "cushioned" from moderate trauma.

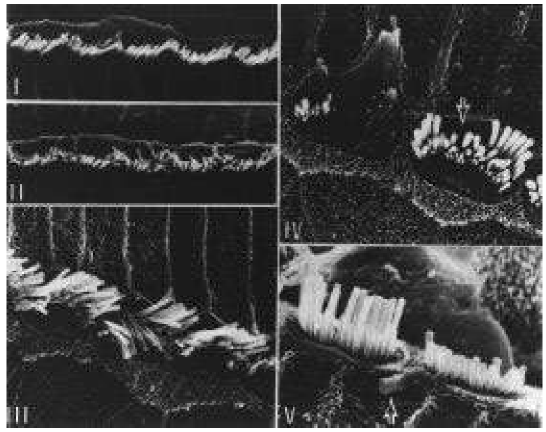


Figure 4-7. Healthy and damaged stereocilia. Courtesy Charles Libermann, PhD.

Thus, in most instances of moderate trauma, actual hearing is not lost, as the inner hair cells survive. Since damage is localized to the outer hair cells, the impairment manifests as the inability to hear *quiet* sound. And since the outer hair cells are not active in loud environments, our sense of what is “loud” stays the same. This is a sensory phenomenon known as “recruitment.” A classic example of recruitment is a person suffering from loss asking that the volume of a TV to be turned up, only to suddenly say that it’s too loud. It is a loss of dynamic sensitivity to low amplitude sounds. From the perspective of signal processing, where healthy outer hair cells behave like upward compressors, when damaged, they behave like downward expanders.

4.4.2 Frequency Sensitivity Loss

As noted, hair cells in the organ of corti are stimulated by the displacement of the basilar membrane. The displacement occurs at different points along its length as a function of the membrane fibers’ sympathetic resonance with the frequencies present in a sound. So, when there is a high concentration of energy in a particular frequency band, there will be high displacement energy and potential stress to the hair cells in that particular region of the organ of corti.

Fig. 4-8 shows two audiograms. On the top is the chart of a college-age individual with normal hearing. On the bottom is the audiogram of another (older, mid 30s) who spent 10+ years as a DJ for hire. Note the loss in the left ear between 2 and 8kHz. There is minimal loss in the right ear because the DJ's right ear was protected by a single-cup headphone for cueing records. The left ear was exposed to the volume of the club system. That 2–8kHz loss a direct result of the natural outer ear resonances discussed in Chapter 3, *Psychoacoustics*, section 3.2.2, Temporal Bones, and illustrated in Fig. 3-4.

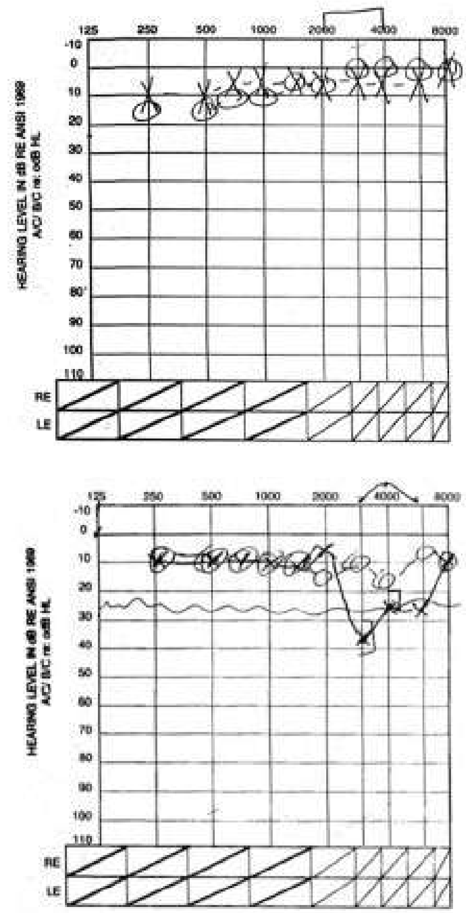


Figure 4-8. audiograms of normal and noise-impaired hearing.

In most environments we are experiencing, on average, broadband acoustic energy. Those outer-ear resonances will always increase exposure in that 2 to 8kHz range. This audiogram loss/curve is commonly referred to as a “noise notch.” There are other hearing disorders that produce loss that is broadband or concentrated in other frequency bands. But noise exposure most often will result in this characteristic curve.

The Result

Moderate, noise-induced hearing loss is characterized as a dynamic distortion in the 2 to 8kHz range. In audio engineering, we know this band as important for developing "presence" in sounds, especially vocals. One can then see how noise-induced loss will manifest itself in an inability to hear consonants and phonemes in speech. And in extreme situations this impairment can make simple conversation, especially in noisy environments, extremely difficult. For the audio engineer or musician, severe impairment will result in the inability to judge equalization or musical timbre.

4.4.3 Temporary Threshold Shift

Most have experienced the feeling of "cotton in the ears" after a night out at a concert or club. Sometimes this will be accompanied by some ringing. For many, this will subside after anywhere from two to four days. This is commonly known as temporary threshold shift or "TTS." Studies show that the temporary loss is stereocilia damage in outer hair cells that heals over time, restoring one's hearing. While there is no definitive "threshold" where the temporary loss becomes permanent, continued excessive exposure will eventually result in permanent loss and permanent tinnitus.

What should be of particular concern to the audio or music professional is the two- to four-day recovery time. For the person who is working/performing 4 to 5 nights a week, there is never sufficient time to fully recover to one's normal hearing thresholds. Considering that the "EQ" of the TTS is the same noise notch, the individual is dealing with a "continuous threshold shift" in that critical 2–8kHz band.

4.4.4 Permanent Threshold Shift

Current research suggests that permanent hearing loss and other disorders are the result of simple wear and tear of the hearing mechanisms over time. Again, everyone's susceptibility is different. The NIOSH studies indicated that most permanent shifts from noise exposure develop over 10 to 15 years of ongoing exposure.

Additionally, research published in 2009 by Sharon Kujawa, PhD and Charles Liberman, PhD,[3] show that noise exposure also traumatizes neural connections at the base of the hair cells. This trauma results in neural damage that doesn't manifest until many years after the trauma.

In the end, permanent hearing loss and other disorders is a combination of the mechanical, chemical, and neural breakdown of the hearing mechanism and the signal paths from the ears to the brain. Much of this is caused by overexposure to excessive levels of sounds from many sources.

4.4.5 Recognizing Noise Hazards

We live in an increasingly noisy world. While we tend to view live music as a prime hearing hazard, many other forms of entertainment and recreation are as loud or louder than a concert. As audio and sound professionals, we must be aware of *all* incidents of noise hazards. A good benchmark for remembering the NIOSH exposure scale is 100dBA as safe for up to 15 minutes, as level changes, for every 3dB up or down, one simply halves or doubles safe exposure time, respectively. Consider the following examples of activities, none are music, yet all have been measured at an average of 100dBA or higher: any power saw, professional hockey, football or soccer, jet skis, snow mobiles, and auto races. At the time of this writing, Kansas City Chiefs fans officially reached 137.5dBA (duration not noted) at a recent game to make the Guinness Book of World Records. A dubious honor, to be sure.

4.4.6 The Hazard of Earbuds

Among the most popular "demons" of hearing health are personal music players and earbuds. It is true that the popularity of listening all day, even at moderate volume, can pose a potential risk. Certainly this applies to those who literally listen all day long. However as the NIOSH/OSHA chart shows, the "fun" of listening loud reduces the safe listening time. There is a hidden hazard where people unintentionally listen too loudly.

The problem is with the earbud, not the player. Earbuds are earphones that sit in the concha of the ear and are nothing more than a speaker placed at the opening of the ear canal. An "insert earphone" is one with a soft tip, and is actually inserted into the ear canal, thus providing some isolation (varying from brand to brand) from the surrounding sound environment.

The potential danger of earbuds is that they do not provide any isolation from the sound environment, thus listeners raise the listening volume to compete with that "noise floor". In 2011 a study of listening volume with earbuds was conducted out of Boston Children's Hospital and the University of Colorado at Boulder. One important finding showed that most people (over 80% of the sample) do not habitually listen at dangerous levels as long as they are listening in a quiet ambient environment. The study went on to show that in an ambient field measuring 80dBA, the average listening level was 85dBA. Another significant finding was volume measurements of different types of earbuds, insert earphones, and headphones as driven by an iPod, shown in Fig. 4-9. An iPod with earbuds becomes a potential hazard above the 75% setting.

Volume control %	Output Level in dBA		
	Earbud	Insert/Isolator	Supra-Aural
60%	77.0	79.5	71.8
70%	83.1	89.7	78.0
80%	89.3	91.8	84.1
90%	95.4	98.0	90.3
100%	101.6	104.1	96.4

RMS average, free-field equivalent output levels.
"Earbud" includes stock iPod earphones.
"Insert/Isolator" includes Etymotic ER6i and Shure E4c earphones.
"Supra-Aural" includes Koss brand headphones.

From Portnuff, Fligor & Arehart (2011).

Figure 4-9. Comparative SPL output of earbuds, earphones, and headphones.[2]

4.4.7 Other Disorders

The most common disorders that can accompany loss are tinnitus and hyperacusis.

Tinnitus ["tin-i-tuss"] is the sensation of sound where there is no actual acoustic stimulus. It most often manifests itself as steady tones; generally single, relatively high frequencies. It also can sound like band-limited noise, "chuffing" sounds, or pulses. Many have low levels of tinnitus that are "audible" only in very quiet environments. Some suffer levels so loud that they never stop hearing the sounds.

Hyperacusis is a term for hypersensitivity to loud sound. Most experience some degree of pain between 115dB SPL and 120dB SPL. The individual suffering from hyperacusis may find levels of 95dB (a loud subway) to be painful.

Currently, there is research to determine cause and treatment of these disorders. Unfortunately there are no cures other than training the brain to ignore tinnitus and the use of earplugs to protect the hyperacusis sufferer from painful loud sound.

4.5 Hearing Conservation Strategies

The “mission” for the audio professional is to be aware of *all* potentially hazardous environments and prepared with effective protection. Hearing loss is an equal-opportunity disability. Loud is loud, whether it comes from a Fender Stratocaster or a DeWalt circular saw. To make conservation “manageable” is to know how hazardous the environment is, and how much protection is needed, given the expected exposure time.

4.5.1 Earplugs

Use of earplugs will provide protection against overexposure in loud sound environments. The benefit of an earplug is simple math. If you are exposed to 110dBA, which would be safe for less than 2 minutes, a plug rated at 20dB would reduce level at the tympanic membrane to 90dBA, providing protection that is safe for 2 hours, based on the more stringent NIOSH scale.

Typical earplugs, the sort found in most drug stores and used in industry, are either soft foam or multi-use “flanged” plugs. While they offer an average of 20dB of attenuation, the rate of attenuation increases by as much as an additional 10dB above 2kHz. The result is the “muffled” sound and widespread opinion that earplugs sound lousy and get in the way of enjoying an event. Thus starts the battle for hearing protection.

More and more music and audio professionals are discovering flat-attenuation plugs, more commonly known as “musician’s earplugs.” A number of years ago, research was conducted to find materials that could fit into a plug aperture and reduce volume evenly across the audio spectrum. Fig. 4-10 shows one brand of generic-flanged musician's plug and its attenuation curve compared to typical foam plugs. There are a number of manufacturers of musician's plugs, all of which offer around 20dB of attenuation and priced between $4.00 and $15.00. The advantage is price, the disadvantage is that there are many situations where 20dB of attenuation is actually more than what would be needed for one or two hours of exposure.

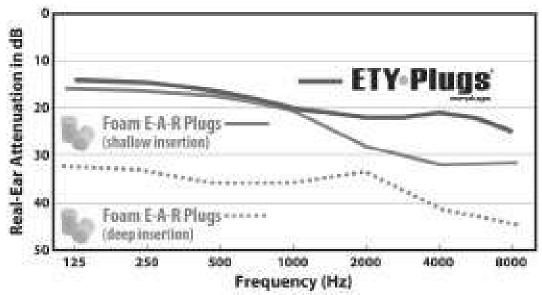


Figure 4-10. Attenuation response of generic musician's foam earplugs. Courtesy of Etymotic Research.

4.5.2 Custom Earplugs

Custom molded earplugs, shown alongside generics in Fig. 4-11 are more expensive, costing about $175.00. But the result is a plug with a perfect fit. They are more comfortable to wear for long periods of time, and can be fitted with interchangeable 9dB, 15dB or 25dB filters. EQ curves for these filters are shown in Fig. 4-12. Note that the 15dB filter has the flattest response of any of the flat-response plugs or filters.[2]

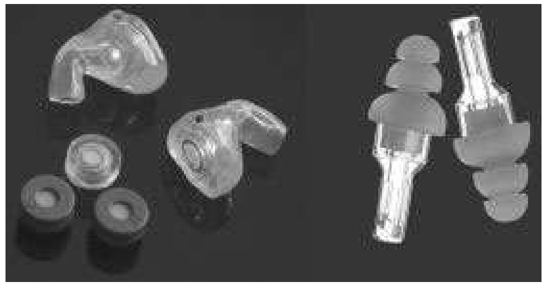


Figure 4-11. Custom and generic flat-attenuation earplugs. Courtesy Sensaphonics and Etymotic Research.

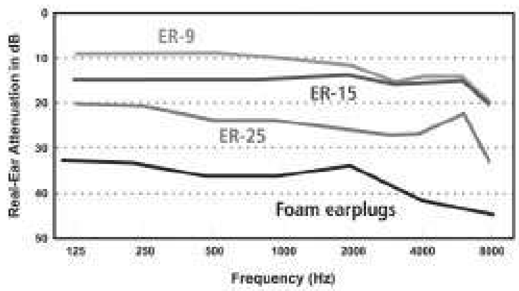


Figure 4-12. Attenuation response of custom plug filters compared to generic foam plugs.

It is important to note that since most custom plugs are molded in soft silicone, they also adjust to the changing shape of the ear canal. Many do not realize that the outer half of the ear canal is surrounded by cartilage, not bone, so the silicone plug can flex with jaw movement. A hard plug cannot provide a continuous, perfect seal.

4.5.3 Proper Fit and Fitting

Regardless of the type of earplug being used, it is critical to attain a snug seal. A poorly inserted plug will vent not only air, but sound as well. In fact, that vent will allow the cavity between the plug and the tympanic membrane to resonate and potentially produce volume levels higher than the ambient noise environment. This is another advantage of the soft-silicone custom plug.

It is also important to find an audiologist who is experienced in the technique of taking what is known as a “deep impression.” The custom plug must go past the “second bend” into the bony portion (inner half) of the ear canal. This way, the plug will not shift as the jaw is moved. Most audiologists are only trained in making impressions for hearing aids, which do not go as deep since they are hard plastic. Once impressions are made, the audiologist will send the impressions to a lab where the plugs are actually made. The customer/patient must be careful to assess the fit immediately when the plugs arrive. A poorly fitting plug has no value.

4.5.4 Adapting to Hearing Protection

It is important to know that earplugs will change one’s perception of an aural environment, and it may take time to get used to the “sound” of plugs. Sometimes it will take a few minutes, sometimes it will take a few uses, and in some instances, individuals may not be able to use them at all. Consider that some of the sound energy received by the cochlea is by way of what is known as “bone conduction.” Musicians who play violin, viola, or any wind instrument, make physical contact with the instrument, and so, some of the sound they normally hear is by way of bone conduction. Since the plug is only attenuating canal born sound, these players will experience a significant EQ shift, as now a higher percentage of their sound is through bone. This is not to suggest that string and wind players cannot use plugs in performance, only to illustrate that one must “recalibrate” for the sound of playing with plugs.

While we cannot expect to successfully protect ourselves from all noise hazards, using protection whenever one can, will at least ensure that the risk of trauma will be minimized.

4.5.5 In-Ear Monitors

The "in-ear monitor" (IEM) has been around for over 20 years (generic and custom versions shown in Fig. 4-13). Use has increased markedly in the past five years, with (at last count) 27 companies offering custom units built with anywhere from one to eight drivers. As with any speaker being used for reference or performance, one must audition a monitor to know if the sound is appropriate for performance. That aside, the IEM provides four advantages to the live/stage performer:

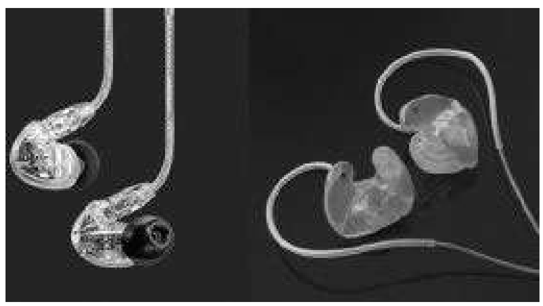


Figure 4-13. Generic and custom in-ear monitors. Courtesy Shure, Inc. and Sensaphonics.

1. Isolation from the "stage sound" to receive a potentially ideal monitor mix for better performance.
2. No danger of monitor feedback for singers.
3. With a wireless receiver (typical), more freedom to move about the stage and even into the audience.
4. With proper measurement of volume level at the tympanic membrane, ability to maintain a safe (or at least safer) monitor level to reduce the risk of noise-induced hearing loss.

There is an almost indistinguishably fine line between an insert-ear phone and a generic in-ear monitor. Those made by recognized audio companies generally offer the best frequency response, good dynamic range, and a good variety of ear tips to ensure the best possible fit. Along with the aforementioned need for a good seal for good isolation, a good seal will also ensure the best bass response. Any venting will not only let ambient sound energy in, but bass energy out.

Much of the effectiveness of the better IEM systems is due to the balanced armature driver (seen in the IEM close-up shot in Fig. 4-14), which does not need to be vented, and so, can be fully encased in a custom earmold. Then too, more than one driver can be built into the piece. Most common are two and three-driver units. They are designed much like loudspeakers with multiple drivers handling different frequency ranges. Again, tastes will determine what configuration is best.

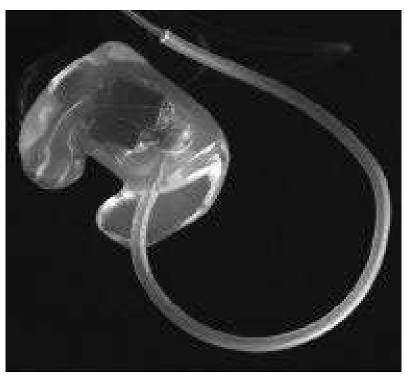


Figure 4-14. Close-up of custom in-ear monitor showing balanced-armature drivers. Courtesy Sensaphonics.

Like custom earplugs, custom IEMs made from soft silicone also benefit from a more comfortable and effective seal for better isolation and bass response. And again, deep molds that extend into the bony portion of the ear canal will give more consistent performance as the tip is well anchored in the canal.

4.5.6 IEM Rumors, Research, and Effectiveness

As IEM usage has grown, some express concern that they are somehow more hazardous, due to the proximity of the driver to the tympanic membrane. Many contended that use of IEMs promote higher levels than floor monitors and are, therefore, more hazardous. Others believe the opposite, that simply using IEMs is a way of protecting one's hearing.

The IEM is not a hearing conservation aid without mindful control of volume. Even the single driver insert earphones noted in Fig. 4-9 are capable of 104dB, driven by an iPod. Higher drive voltages from wireless belt packs and multiple-driver monitors result in levels that can easily reach 120dBA. IEMs can be an effective tool for better hearing health, but it requires a commitment from both the musician and monitor engineer.

In a 2008 study,[3] Matt Federman and Jeremy Ricketts of Vanderbilt University studied musicians' listening volume preferences with both floor and IEM systems. Working with a group of musicians (23-48 years of age and a minimum of 10 years of professional experience), they built a live stage environment where the musicians were able to adjust their vocal monitor levels. In each of a number of trials, the musicians were asked to adjust floor or IEMs to what they are used to, their "preferred listening level" (PLL), and then find the "minimum acceptable listening level" (MALL) or lowest level they could manage before the monitor was of no use. Once the PLL or MALL was established, a probe tube microphone system was used to measure sound pressure levels at the tympanic membrane.

The study made three (among other) very important findings:

1. From trial to trial, there was less than 1dB difference in the PLL for any musician. Musicians are truly “calibrated” to a certain volume to be able to play effectively. This suggests that musicians’ sensitivity to volume and dynamic is as highly tuned as their sensitivity to frequency and timbre.
2. Comparing PLLs between floor monitors and IEMs, there was an average 0.6dB differences in level. While this would be considered statistically significant, in the realm of exposure hazard this is would not change recommended exposure time and shows that loudness perception is the same, regardless of the proximity of the driver to the ear.
3. Average MALLs were 1.8dB lower than PLLs using wedges, but 6dB lower using IEMs. If a musician is *willing* to reduce monitor levels as a means of preserving hearing, use of IEMs will enable him or her to comfortably reduce levels by 6dB, significantly reducing the risk of damage and/or increasing safe exposure time.

Unfortunately, probe-tube measurement systems are expensive and impractical for general use. A few audiologists have such systems and can provide an opportunity for an IEM user to “learn” what his or her PLL and MALL actually is. Another option is a newly-developed technology that can display the level of a particular IEM. Fig. 4-15 shows such a unit, which measures drive voltage and, based on the sensitivity rating of the IEM (loaded into a preference file), displays the actual level at the tympanic membrane as well as safe exposure time based on both the NIOSH and OSHA scales. The last option is to simply make a concerted effort to moderate one's monitor volume as much as possible.

Figure 4-15. In-ear monitor SPL meter. Courtesy Sensaphonics.

4.6 Developing a Practical Strategy

Audio engineers and musicians must acknowledge that healthy hearing is central to practicing their art and trade.

Noise hazards are not limited to live or loud playback of music. Many forms of entertainment and recreation can reach dangerous volume levels, as either crowd, mechanical or electronic noise/sound.

Earplugs can reduce the risk of trauma in any loud and/or long time exposure environments.

In ear monitors can provide the live stage performer a well-controlled, feedback-free, and potentially safevolume monitor mix.

There is a growing network of audiologists who specialize in treating musicians and audio professionals. They will be able to provide a comprehensive preventive hearing health care plan and be a resource for custom earplugs and in-ear monitors.

It is critical to receive a “base-line” audiometric exam, then follow-up with subsequent exams to determine if one is at risk for hearing disorders, be it from noise exposure or other factors.

References

1. S. G. Kujawa, M. C. Liberman (2009). Adding Insult to Injury: Cochlear Nerve Degeneration After “Temporary” Noise-Induced Hearing Loss, *Journal of Neuroscience*, 29(45):14077–14085.

2. Portnuff, Fligor, and Arehart, 2011, Teenage use of portable listening devices: a hazard to hearing, *Journal of the American Academy of Audiology*, 2011 Nov-Dec; 22(10): 663–77. doi: 10.3766/jaaa.22.10.5

3. Jeremy Federman; Todd Ricketts, Preferred and Minimum Acceptable Listening Levels for Musicians While Using Floor and In-Ear Monitors, *Journal of Speech, Language, and Hearing Research*, February 2008, Vol.51, 147–159. doi:10.1044/1092-4388(2008/011).

Chapter 5

Fundamentals of Audio and Acoustics

by Pat Brown

5.1 Introduction

Many people get involved in the audio trade prior to experiencing technical education. Those serious about practicing audio dig in to the books later to learn the physical principles underlying their craft. This chapter is devoted to establishing a baseline of technical information that will prove invaluable to anyone working in the audio field.

Numerous tools exist for those who work with sound systems. The most important are the mathematical tools. Their application is independent of the type of system or its use, plus, they are timeless and not subject to obsolescence like audio products. Of course, one must always balance the mathematical approach with real-world experience to gain an understanding of the shortcomings and limitations of the formulas. Once the basics have been mastered, sound system work becomes largely intuitive.

Audio practitioners must have a general understanding of many subjects. The information in this chapter has been carefully selected to give the reader the big picture of what is important in sound system work. Many of the topics are covered in greater detail in other chapters of this book. In this initial treatment of each subject, the mathematics has been kept to a minimum, opting instead for word explanations of the theories and concepts. This provides a solid foundation for further study of any of the subjects. Considering the almost endless number of topics that could be included here, I selected the following based on my own experience as a sound practitioner and instructor. They are:

1. The Decibel and Levels.
2. Frequency and Wavelength.
3. The Principle of Superposition.
4. Ohm’s Law and the Power Equation.
5. Impedance, Resistance, and Reactance.
6. Introduction to Human Hearing.
7. Monitoring Audio Program Material.
8. Sound Radiation Principles.
9. Wave Interaction.

A basic understanding in these areas will provide the foundation for further study in areas that are of particular interest to the reader. Most of the ideas and principles in this chapter have existed for many years, and their origins have become obscure. While I haven't quoted any of the references in the bibliography verbatim, they get full credit for the bulk of the information presented here.

5.2 The Decibel

Perhaps the most useful tool ever created for audio practitioners is the decibel (dB). It allows changes in system parameters such as power, voltage, and distance to be related to level changes heard by a listener. In short, the decibel is a way to express "how much" in a way that is relevant to the human perception of loudness. We will not track its long evolution or specific origins here. Like most audio tools, it has been modified many times to stay current with the technological practices of the day. Excellent resources are available for that information, notably *Sound System Engineering* by Davis, Patronis, and Brown. What follows is a short study on how to use the decibel for general audio work.

Most of us tend to consider physical variables in linear terms. For instance, twice as much of a quantity produces twice the end result. Twice as much sand produces twice as much concrete. Twice as much flour produces twice as much bread. This linear relationship does not hold true for the human sense of hearing. Using that logic, twice the amplifier power should sound twice as loud. As you will see, this is not true.

Perceived changes in the loudness and frequency of sound are based on the *percentage* change from some initial condition. This means that audio people must be concerned with ratios. A given ratio always produces the same result. Subjective testing has shown that the power applied to a loudspeaker must be increased by about 26% to produce an audible change. Thus a ratio of 1.26:1 produces the minimum audible change, regardless of the initial power quantity. If the initial amount of power is 1 watt (W), then an increase to 1.26W will produce a "just audible" increase. If the initial quantity is 100W, then 126W will be required to produce a just audible increase. A number scale can be linear with values like 1, 2, 3, 4, 5, etc. A number scale can be proportional with values like 1, 10, 100, 1000, etc. A scale that is calibrated proportionally is called a *logarithmic* scale. In fact, *logarithm* means "proportional numbers." For simplicity, base 10 logarithms are used for audio work. Using amplifier power as an example, changes in level are determined by finding the ratio of change in the parameter of interest (e.g. wattage) and taking the base 10 logarithm. The resultant number is the level change between the two wattages expressed in Bels. The base 10 logarithm is determined using a look-up table or scientific calculator. The log conversion accomplishes two things:

1. It puts the ratio on a proportional number scale that better correlates with human hearing.
2. It allows very large numbers to be expressed in a more compact form, Fig. 5-1.

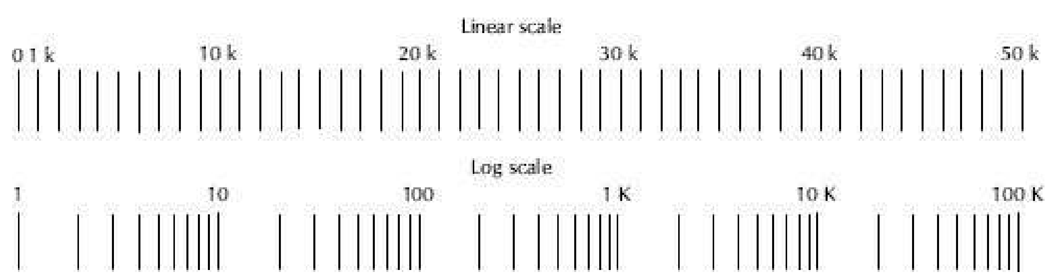


Figure 5-1. A logarithmic scale has its increments marked by a fixed ratio, in this case 10 to 1, forming a more compact representation than a linear scale. Courtesy SynAudCon.

The final step in the decibel conversion is to scale the Bel quantity by a factor of ten. This step converts Bels to decibels and completes the conversion process, Fig. 5-2. The decibel scale is more resolute than the Bel scale.

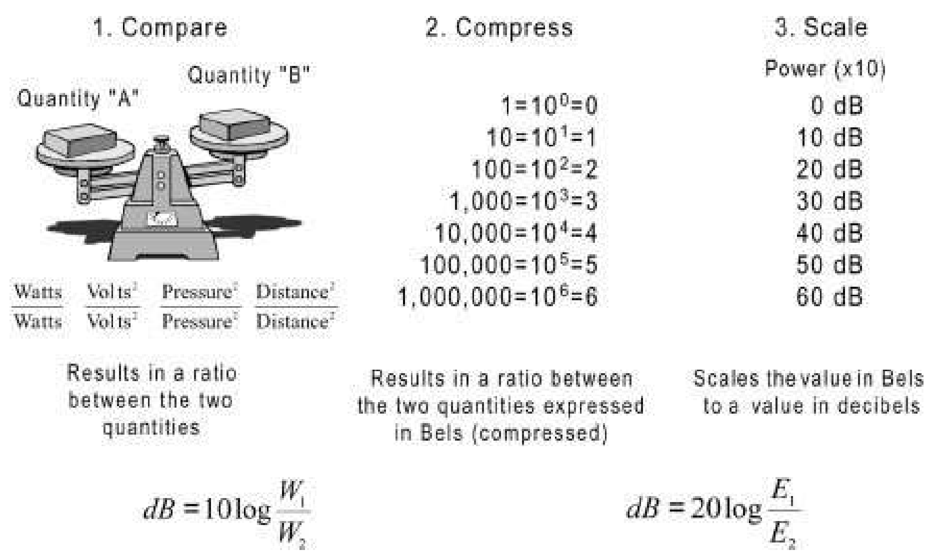


Figure 5-2. The steps to performing a decibel conversion are outlined. Courtesy SynAudCon.

The decibel is always a power-related ratio. Electrical and acoustical power changes can be converted exactly in the manner described. Quantities that are not powers must be made proportional to power—a relationship established by the power equation.

$$W = \frac{E^2}{R} \tag{5-1}$$

where,
W is power in W,
E is voltage in V,
R is resistance in Ω.

This requires voltage, distance, and pressure to be squared prior to taking the ratio. Some practitioners prefer to omit the squaring of the initial quantities and simply change the log multiplier from ten to twenty. This produces the same end result.

Table 5-1 provides a list of voltage, pressure, distance, and power ratios required to produce the indicated dB change. It is a worthwhile endeavor to memorize the changes indicated in bold type and be able to recognize them by listening.

Table 5-1 Some Important Decibel Changes and the Ratios of Power, Voltage, Pressure, and Distance that Produce Them. Courtesy SynAudCon.

Subjective Change	Voltage, Distance, Pressure Ratio	% of Original	Power Ratio	dB Change
	20log		10log	
Barely perceptible	**1.12:1**	**89**	**1.26:1**	**1dB**
	1.26:1	79	1.58:1	2dB
Noticeable to most	**1.41:1**	**71**	**2:1**	**3dB**
	1.58:1	63	2.51:1	4dB
	1.78:1	56	3.16:1	5dB
Goal for system changes	**2:1**	**50**	**4:1**	**6dB**
	2.24:1	45	5:1	7dB
	2.51:1	40	6.3:1	8dB
	2.8:1	36	8:1	9dB
Twice as loud or soft	**3.16:1**	**32**	**10:1**	**10dB**
	10:1	10	100:1	20dB
	31.6:1	3	1000:1	30dB
Limits of audibility	**100:1**	**1**	**10,000:1**	**40dB**
	316:1	0.3	100,000:1	50dB
	1000:1	0.1	1,000,000:1	60dB

A decibel conversion requires two quantities that are in the same unit, i.e., watts, volts, meters, feet. The unit cancels during the initial division process, leaving the ratio between the two quantities. For this reason, the decibel is without dimension and is therefore technically not a unit in the classical sense. If two arbitrary quantities of the same unit are compared, the result is a *relative* level difference. If a standard reference quantity is used in the denominator of the ratio, the result is an absolute level and the unit is dB relative to the original unit. Relative levels are useful for live sound work. Absolute levels are useful for equipment specifications and calibration. Table 5-2 lists some references used for determining absolute levels.

The decibel was originally used in impedance-matched interfaces and always with a power reference. Power requires knowledge of the resistance that a voltage is developed across. If the resistance value is fixed, changes in applied voltage can be expressed in dB, since the power developed will predictably track the applied voltage. In modern sound systems, few device interfaces are impedance matched. Interfaces are actually mismatched to optimize the voltage transfer between components. While the same impedance does not exist at each device interface, the same impedance *condition* does. If a minimum 1:10 ratio exists between the output impedance and input impedance, then the voltage transfer is essentially independent of the actual output or input impedance values. Such an interface is termed *constant voltage*, and the signal source is said to be operating *open circuit* or *un-terminated*. In constant voltage interfaces, open circuit conditions are assumed when using the decibel. This means that the level change at the output of the system is caused by changing the voltage somewhere in the processing chain and is dependent on the voltage change only, not the resistance that it is developed across or the power transfer. Since open-circuit conditions exist almost universally in modern analog systems, the practice of using the decibel with a voltage reference is wide-spread, well-accepted, and essential.

Table 5-2 Some common decibel references used by the audio industry.

Electrical Power	
dBW	1 Watt
dBm	0.001 Watt
Acoustical Power	
dB-PWL or L_w	10^{-12}Watt
Electrical Voltage	

dBV	1 Volt
dBu	0.775 Volts
Acoustical Pressure	
dB SPL or L_p	0.00002 Pascals

One of the major utilities of the decibel is that it provides a common denominator when considering level changes that occur due to voltage changes at various points in the signal chain. By using the decibel, changes in sound level at a listener position can be determined from changes in the output voltage of any device ahead of the loudspeaker. For instance, a doubling of the microphone output voltage produces a 6dB increase in output level from the microphone, mixer, signal processor, power amplifier, and ultimately the sound level at the listener. This relationship assumes linear operating conditions in each device. The 6dB increase in level from the microphone could be caused by the talker speaking 6 dB louder or by simply reducing the miking distance by one-half (a 2:1 distance ratio). The level controls on audio devices are normally calibrated in relative dB. Increasing a fader or rotary knob by 6dB causes the output voltage of the device (and system) to increase by a factor of 2 and the output power from the device (and system) to increase by a factor of four.

Absolute levels are useful for rating audio equipment. A power amplifier that can produce 100 watts of continuous power is rated at

$$\begin{aligned} L_{out} &= 10\log W \\ &= 10\log 100 \\ &= 20\,\text{dBW} \end{aligned} \tag{5-2}$$

This means that the amplifier can be 20dB louder than a 1 watt amplifier. A mixer that can output 10 volts prior to clipping can be rated at

$$\begin{aligned} L_{out} &= 20\log E \\ &= 20\log 10 \\ &= 20\,\text{dBV} \end{aligned} \tag{5-3}$$

If the same mixer outputs 1 volt rms at meter zero, then the mixer has 20dB of peak room above meter zero.

If a loudspeaker can produce a sound level at 1 meter of 90dB ref. 20μPa (micro-Pascals), then at 10 meters its level will be

$$\begin{aligned} L_p &= 90 + 20\log\frac{1}{10} \\ &= 90 + (-20) \\ &= 70\,\text{dB} \end{aligned} \tag{5-4}$$

In short, the decibel says, "The level difference at a point of observation caused by changing a quantity in the signal chain will depend upon the initial value of the quantity and the *percentage* that it is changed."

The applications of the decibel are endless, and the utility of the decibel is self-evident. It forms a bridge between the amount of change of a physical parameter and the loudness change that is perceived by the human listener. The decibel is the language of audio, Fig. 5-3. Once a practitioner is fluent in the use of the decibel, they will find that there may be no need to describe signals in their original units. Voltmeters designed for use in audio systems can read directly in dB, with user-selectable reference values. Here are some examples.

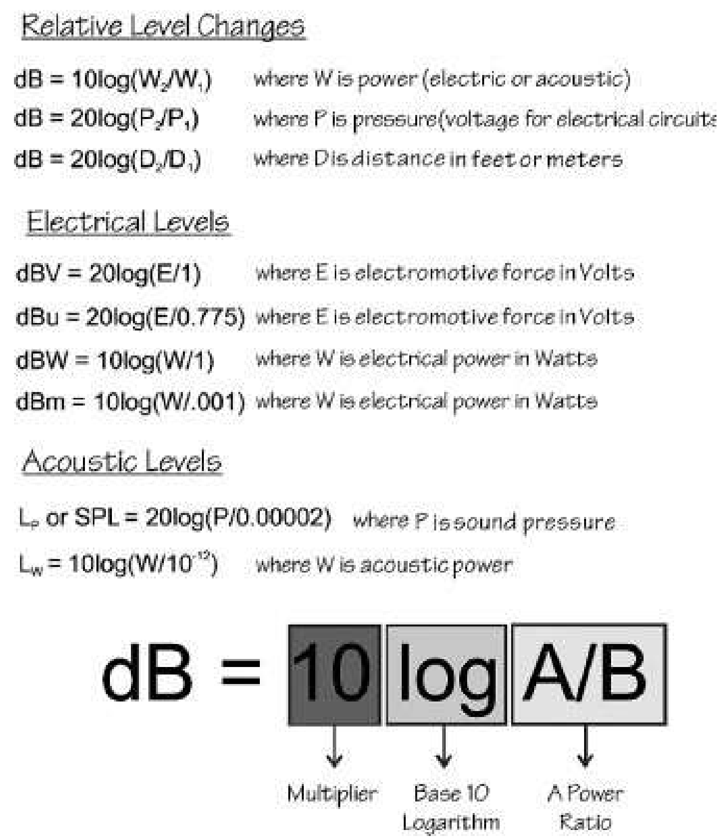


Figure 5-3. Summary of decibel formulas for general audio work. Courtesy SynAudCon.

- The mic is producing -40dBV.
- The output of the mixer is +4dBu.
- The power amp is putting out +20dBW.
- The sound level at the listener is 90dB SPL.

5.3 Loudness and Level

The perceived loudness of a sound event is related to its acoustical level, which is in turn related to the electrical level driving the loudspeaker. Levels are electrical or acoustical pressures or powers expressed in decibels. In its linear range of operation, the human hearing system will perceive an increase in level as an increase in loudness. Since the eardrum is a pressure sensitive mechanism, there exists a threshold below which the signal is indistinguishable from the background noise. This threshold is about 20μPa of pressure deviation from ambient at midrange frequencies. Using this number as a reference and converting to decibels yields

$$
\begin{aligned}
L_p &= 20\log\frac{0.00002}{0.00002} \\
&= 0\,\text{dB}(\text{or } 0\,\text{dB SPL})
\end{aligned}
\tag{5-5}
$$

This is widely accepted as the threshold of hearing for humans at mid-frequencies. Acoustic pressure levels are always stated in dB ref. 0.00002Pa. Acoustic power levels are always stated in dB ref. 1pW (picowatt or 10^{-12}W). Since it is usually the pressure level that is of interest, we must square the Pascals term in the decibel conversion to make it proportional to power. Sound pressure levels are measured using sound level meters with appropriate ballistics and weighting to emulate human hearing. Fig. 5-4 shows some typical sound pressure levels that are of interest to audio practitioners.

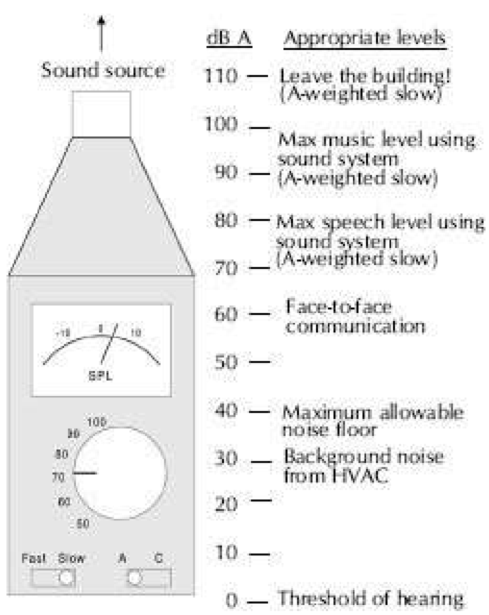


Figure 5-4. Sound levels of interest to system designers and operators. Courtesy SynAudCon.

See Section 5-9, Human Hearing for an important discussion of the weighting scales used by sound level meters.

5.4 Frequency

Audio practitioners are in the wave business. An acoustic wave is produced when a medium is disturbed.

The medium can be air, water, steel, the earth, etc. The disturbance produces a fluctuation in the ambient condition of the medium that propagates as a wave that radiates outward from the source of the disturbance. If one second is used as a reference time span, the number of fluctuations above and below the ambient condition per second is the frequency of the event, and is expressed in cycles per second, or Hertz. Humans can hear frequencies as low as 20Hz and as high as 20,000Hz (20kHz). Audio waves are electromagnetic. Electromagnetic waves do not require a medium, and can propagate through a vacuum. In an audio circuit the fluctuating quantity of interest is usually the electrical voltage or current. In an acoustical circuit it is the air pressure deviation from ambient atmospheric pressure. When the air pressure fluctuations have a frequency between 20Hz and 20kHz they are audible to humans and are called sound waves. A microphone is a transducer that converts sound waves into audio waves. A loudspeaker is a transducer that converts audio waves into sound waves.

As stated in the decibel section, humans are sensitive to proportional changes in power, voltage, pressure, and distance. This is also true for frequency. If we start at the lowest audible frequency of 20Hz and increase it by a 2:1 ratio, the result is 40Hz, an interval of one octave. Doubling 40Hz yields 80Hz. This is also a one-octave span, yet it contains twice the frequencies as the previous octave. Each successive frequency doubling yields another octave increase and each higher octave will have twice the spectral content as the one below it. In spite of this, each octave is perceived as having the same percentage of spectral content (1/10th of the audio spectrum.) This makes the logarithmic scale suitable for displaying frequency. Figs. 5-5 and 5-6 show a logarithmic frequency scale and some useful divisions. The perceived midpoint of the spectrum for a human listener is about 1kHz. Some key frequency ratios exist:

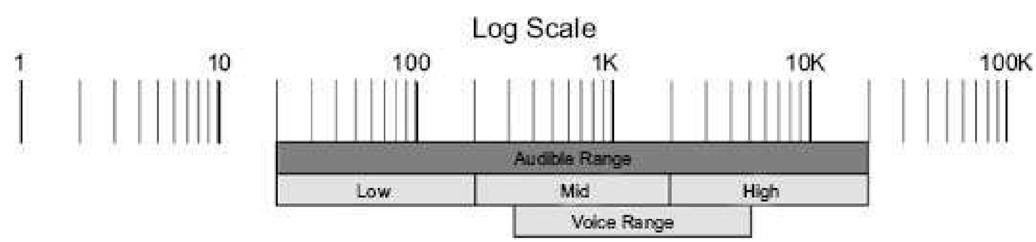


Figure 5-5. The audible spectrum divided into decades (a 10 to 1 frequency ratio). Courtesy SynAudCon.

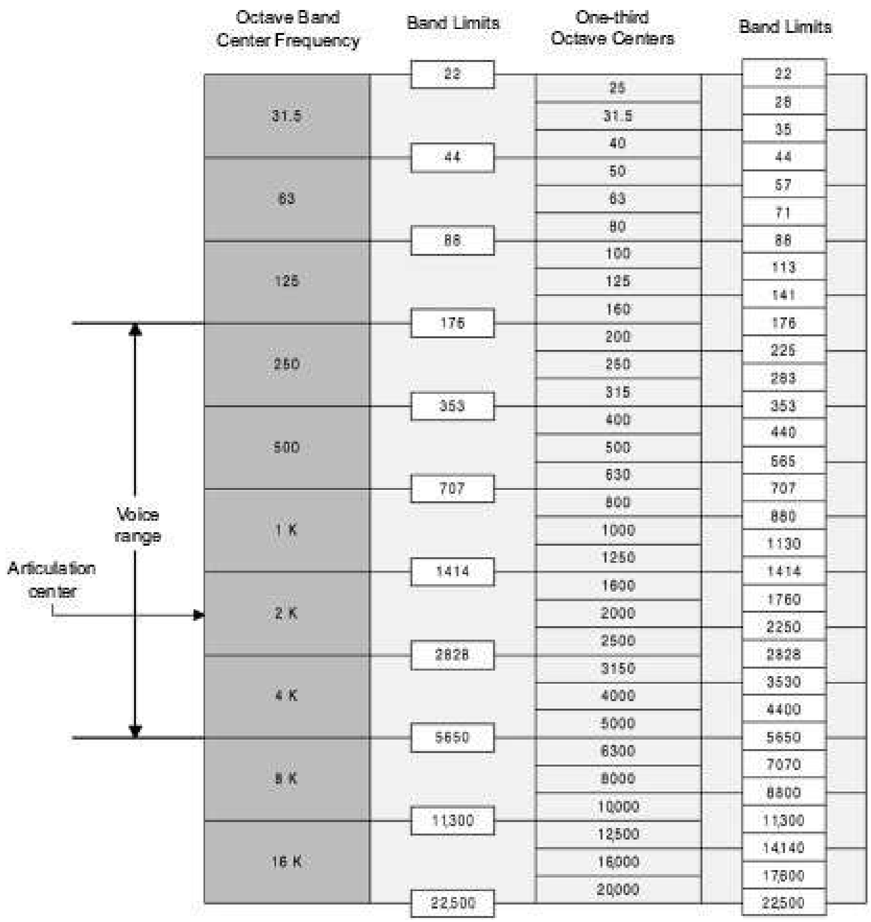


Figure 5-6. The audible spectrum divided into octaves (a 2 to 1 ratio) and one-third octaves. Courtesy SynAudCon.

- 10:1 ratio—decade.
- 2:1 ratio—octave.

The spectral or frequency response of a system describes the frequencies that can pass through that system. It must always be stated with an appropriate tolerance, such as ±3dB. This range of frequencies is the bandwidth of the system. All system components have a finite bandwidth. Sound systems are usually bandwidth limited for reasons of stability and loudspeaker protection. A spectrum analyzer can be used to observe the spectral response of a system or system component.

5.5 Wavelength

If the frequency f of a vibration is known, the time period T for one cycle of vibration can be found by the simple relationship

$$T = \frac{1}{f} \quad (5\text{-}6)$$

The time period T is the inverse of the frequency of vibration. The period of a waveform is the time length of one complete cycle, Fig. 5-7. Since most waves propagate or travel, if the period of the wave is known, its physical size can be determined with the following equation if the speed of propagation is known:

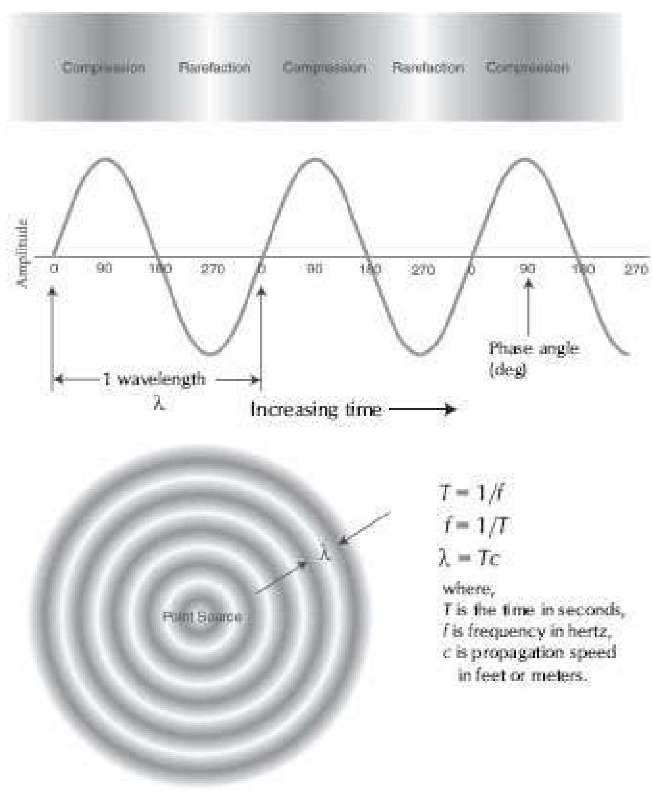


Figure 5-7. The wavelength of an event determines how it interacts with the medium that it is passing through. Courtesy SvnAudCon.

$$\lambda = Tc \tag{5-7}$$

$$\lambda = \frac{c}{f} \tag{5-8}$$

where,
c is the speed of propagation in ft/s or m/s.

Waves propagate at a speed that is dependent on the nature of the wave and the medium that it is passing through. The speed of the wave determines the physical size of the wave, called its *wavelength*. The speed of light in a vacuum is approximately 300,000,000 meters per second (m/s). The speed of an electromagnetic wave in copper wire is somewhat less, usually 90% to 95% of the speed of light. The fast propagation speed of electromagnetic waves makes their wavelengths extremely long at audio frequencies, Table 5-3.

Table 5-3 Acoustic Wavelengths are Relatively Short and Interact Dramatically with their Environment.

	Sound in Air		Copper Wire	
Frequency Hz	U.S. English Ft	SI M	U.S. English Mi	SI KM
31.5	36.0	11.0	5609	9047
63	18.0	5.5	2952	4523
125	9.0	2.7	1476	2261
250	4.5	1.4	738	1130
500	2.3	0.7	369	565
1K	1.13	0.344	184	282
2K	0.56	0.172	92	141
4K	0.28	0.086	46	70
8K	0.14	0.043	23	35
16K	0.07	0.021	11	17.6

Audio wavelengths are extremely long, and phase interaction on audio cables is not usually of concern. Courtesy SynAudCon.

At the higher radio frequencies (VHF and UHF), the wavelengths become very short—1 meter or less. Antennas to that receive such waves must be of comparable physical size, usually one-quarter to one-half wavelength. When waves become too short for practical antennae, concave dishes can be used to collect the waves. It should be pointed out that the highest frequency that humans can hear (about 20kHz) is a very low frequency when considering the entire electromagnetic spectrum.

An acoustic wave is one that is propagating by means of vibrating a medium such as steel, water, or air. The propagation speeds through these media are relatively slow, resulting in waves that are short in length compared to an electromagnetic wave of the same frequency. The wavelengths of audio frequencies in air range from about 17m (20Hz) to 17mm (20kHz). The wavelength of 1kHz in air is about 0.334m (about 1.13ft).

When physically short acoustic waves are radiated into large rooms, there can be adverse effects from reflections. Acoustic reflections occur when a wave encounters a change in acoustic impedance, usually from a rigid surface, the edge of a surface or some other obstruction. The reflection angle equals the incidence angle in the case of a large, rigid boundary. Architectural acoustics is the study of the behavior of sound waves in enclosed spaces. Acousticians specialize in creating spaces with reflected sound fields that enhance rather than detract from the listening experience.

When sound encounters a room surface, a complex interaction takes place. If the surface is much larger than the wavelength, a reflection occurs and an acoustic shadow is formed behind the boundary.

If the obstruction is smaller than the wavelength of the wave striking it, the wave diffracts around the obstruction and continues to propagate. Both effects are complex and frequency (wavelength) dependent, making them difficult to calculate, Fig. 5-8.

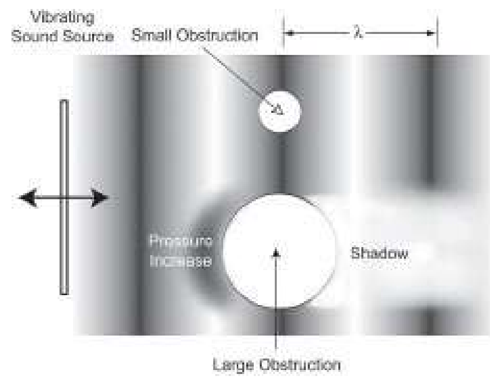


Figure 5-8. Sound diffracts around objects that are small relative to the length of the sound wave. Courtesy SynAudCon.

The reflected wave will be strong if the surface is large and has low absorption. As absorption is increased, the level of the reflection is reduced as some of the energy is converted into heat. If the surface is random, the wave can be scattered, depending on the size relationship between the wave and the surface relief. Commercially available diffusors can be used to achieve uniform scattering in critical listening spaces, Fig. 5-9.

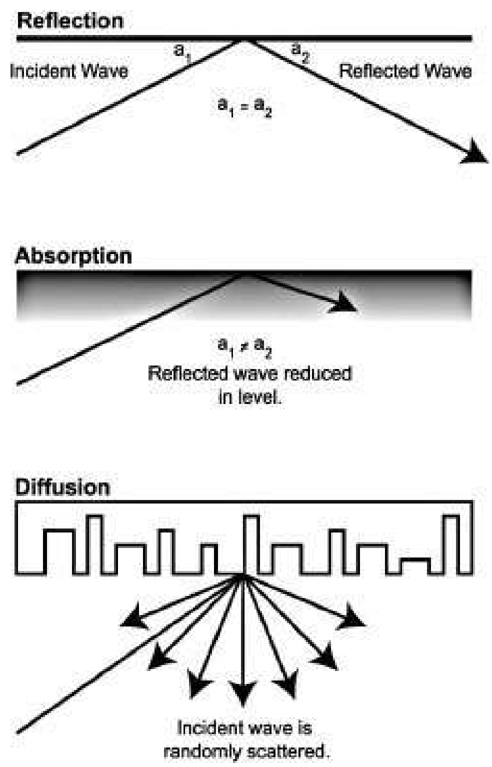


Figure 5-9. Sound waves will interact with a large boundary in a complex way. Courtesy SynAudCon.

5.6 Surface Shapes

The geometry of a boundary can have a profound affect on the behavior of the sound that strikes it. From a sound reinforcement perspective, it is usually better to scatter sound than to focus it. A concave room boundary should be avoided for this reason, Fig. 5-10. Many auditoriums have concave rear walls and balcony faces that require extensive acoustical treatment for reflection control. A convex surface is more desirable, since it scatters sound waves whose wavelengths are small relative to the radius of curvature. Room corners can provide useful directivity control at low frequencies, but at high frequencies can produce problematic reflections.

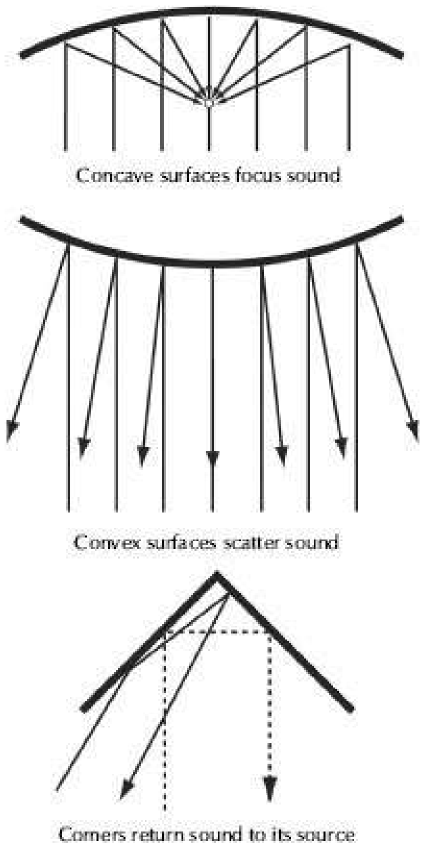


Figure 5-10. Some surfaces produce focused reflections. Courtesy SynAudCon.

Electrical reflections can occur when an electromagnetic wave encounters a change in impedance as it propagates. For such waves traveling down a wire, the reflection is back towards the source of the audio waves. Such reflections are not usually a problem for waves unless there is a phase offset between the outgoing and reflected waves. Note that an audio cable would need to be very long for its length to cause a significant time offset between the incident and reflected wave (many thousands of meters). Since audio waves are very long in length, and audio cables are relatively short, reflections from a load back to the source are not typically of concern, as the incident and reflected waves are coherent. At radio frequencies, reflected waves pose a huge problem, and cables are normally terminated (operated into a matched impedance) to absorb the incident wave at the receiving device and reduce the level of the reflection. The same is true for digital signals due to their very high frequency content.

5.7 Superposition

Sine waves and cosine waves are periodic and singular in frequency. These simple waveforms are the building blocks of the complex waveforms of speech and music. The amplitude of a sine wave can be displayed as a function of time or as a function of phase rotation, Fig. 5-11. The sine wave will serve as an example for the following discussion about superposition. Once the size (wavelength) of a wave is known, it is useful to subdivide it into smaller increments for the purpose of tracking its progression through a cycle or comparing its progression with that of another wave. Since the sine wave describes a cyclic (circular) event, one full cycle is represented by 360°, at which point the wave repeats.

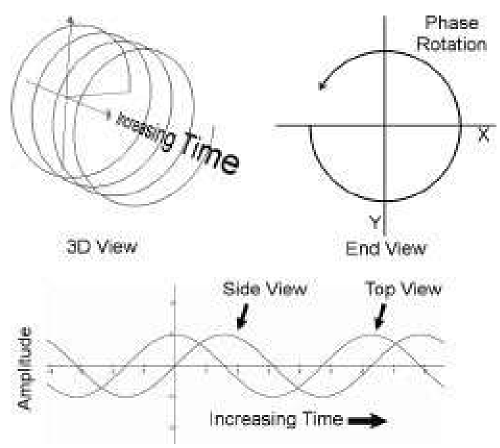


Figure 5-11. Simple harmonic motion can be represented with a sine or cosine wave. Both are viewpoints of the same event from different angles. Courtesy SynAudCon.

When multiple sound pressure waves pass by a point of observation, their responses sum to form a composite wave. The composite wave is the complex combination of two or more individual waves. The amplitude of the summation is determined by the relative phase of the individual waves. Let's consider how two waves might combine at a point of observation. This point might be a listener seat or microphone position. Two extremes exist. If there is no phase offset between two waves of the same amplitude and frequency, the result is a coherent summation that is twice the amplitude of either individual wave (+6dB). The other extreme is a 180° phase offset between the waves. This results in the complete cancellation of the pressure response at the point of observation. An infinite number of intermediate conditions can occur between these two extremes. The phase interaction of waves is not a severe problem for analog audio signals in the electromagnetic domain for sound systems, where the wavelengths at audio frequencies are typically much longer than the interconnect cables. Waves reflected from receiver to source are in phase at any point along the cable and no cancellation occurs. This is not the case for video, radio frequency, and digital signals. The shorter wavelengths of these signals can be dramatically affected by wave superposition on interconnect cables. As such, great attention must be given to the length and terminating impedance of the interconnect cables to assure efficient signal transfer between source and receiver. The practice of impedance matching between source, cable, and load is usually employed.

In sound reinforcement systems, phase interactions are typically more problematic for acoustical waves than electromagnetic waves. Phase summations and cancellations are the source of many acoustical problems experienced in auditoriums. Acoustic wavelengths are often short relative to the room size (at least at high frequency), so the waves tend to bounce around the room before decaying to inaudibility. At a listener position, the reflected waves “superpose” to form a complex waveform that is heard by the listener. The sound radiated from multiple loudspeakers will interact in the same manner, producing severe modifications in the radiated sound pattern and frequency response. Antenna designers have historically paid more attention to these interactions than loudspeaker designers, since there are laws that govern the control of radio frequency emissions. Unlike antennas, loudspeakers are usually broadband devices that cover one decade or more of the audible spectrum. For this reason, phase interactions between multiple loudspeakers never result in the complete cancellation of sound pressure, but rather cancellation at some frequencies and coherent summation at others. The subjective result is *tonal coloration* and *image shift* of the sound source heard by the listener. The significance of this phenomenon is application-dependent. People having dinner in a restaurant would not be concerned with the effects of such interactions since they came for the food and not the music. Concert-goers or church attendees would be more concerned, because their seat might be in a dead spot, and the interactions disrupt their listening experience, possibly to the point of reducing the information conveyed via the sound system. A venue owner may make a significant investment in good quality loudspeakers, only to have their response impaired by such interactions with an adjacent loudspeaker or room surface, Fig. 5-12.

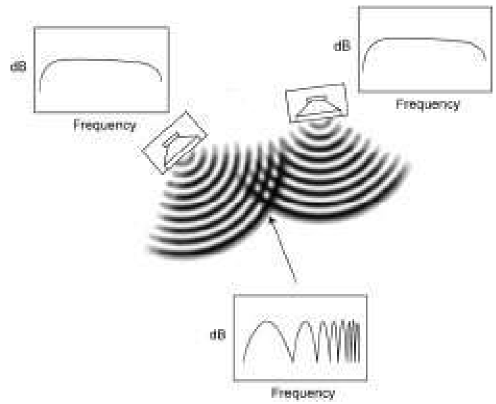


Figure 5-12. Phase interference occurs when waves from multiple sources arrive at different times. Courtesy SynAudCon.

Sound wave interactions are most disruptive in critical listening environments, such as recording studio control rooms or high quality home entertainment systems. Users of these types of systems often make a large investment to maintain sonic accuracy by purchasing phase coherent loudspeakers and appropriate acoustical treatments for the listening space. The tonal coloration caused by wave interference may be unacceptable for a recording studio control room but may be artistically pleasing in a home audio system.

Loudspeaker designers can use wave interaction to their advantage by choosing loudspeaker spacings that form useful radiation patterns. Almost all pattern control in the low frequency decade is achieved in this manner. Uninformed system designers create undesirable radiation patterns by accident in the way that they place and stack loudspeakers. The results are poor coverage and reduced acoustic gain.

The proper way to view the loudspeaker and room are as filters that the sound energy must pass through en route to the listener. Some aspects of these filters can be compensated with electronic filters—a process known as equalization. Other aspects cannot, and electronic *equalization* merely aggravates or masks the problem.

5.8 Ohm's Law

In acoustics, the sound that we hear is nature restoring an equilibrium condition after an atmospheric disturbance. The disturbance produces waves that cause the atmospheric pressure to oscillate above and below ambient pressure as they propagate past a point of observation. The air always settles to its ambient state upon cessation of the disturbance.

In an electrical circuit, a potential difference in electrical pressure between two points causes current to flow. Electrical current results from electrons flowing to a point of lower potential. The electrical potential difference is called an *electromotive force* (EMF) and the unit is the volt (V). The rate of electron flow is called current and the unit is the *ampere* (A). The ratio between voltage and current is called the *resistance* and the unit is the *ohm* (Ω). The product of voltage and current is the *apparent power*, W, that is produced by the source and consumed by the load. Power is the rate of doing work and power ratings must always include a reference to time. A power source can produce a rated voltage at a rated flow of current into a specified load for a specified period of time. The ratio of voltage to current can be manipulated to optimize a source for a specific task. For instance, current flow can be sacrificed to maximize voltage transfer. When a device is called upon to deliver appreciable current, it is said to be operating under load. The load on an automobile increases when it must maintain speed on an uphill grade, and greater power transfer between the engine and drive train is required. Care must be taken when loading audio components to prevent distortion or even damage. Ohm's Law describes the ratios that exist between voltage, current, and resistance in an electrical circuit.

$$R = \frac{E}{I} \tag{5-9}$$

$$E = IR \tag{5-10}$$

$$I = \frac{E}{R} \tag{5-11}$$

where,
E is in V,
I is in A,
R is in Ω.

Direct current (dc) flows in one direction only. In ac (alternating current) the direction of current flow is alternating at the frequency of the waveform. Voltage and current are not always in sync so the phase relationship between them must be considered. Power flow is reduced when they are not in relative phase (synchronization). Voltage and current are in phase in resistive circuits. Phase *shifts* between voltage and current are produced by reactive elements in a circuit. Reactance reduces the power transferred to the load by storing energy and reflecting it back to the source.

Loudspeakers and transformers are examples of sound system components that can have significant reactive characteristics. The combined opposition to current flow caused by resistance and reactance is termed the *impedance* (Z) of the circuit. The unit for impedance is also the ohm (Ω). An impedance can be purely resistive, purely reactive, or most often some combination of the two. This is referred to as a *complex* impedance. Impedance is a function of frequency, and impedance measurements must state the frequency at which the measurement was made. Sound system technicians should be able to measure impedance to verify proper component loading, such as at the amplifier/loudspeaker interface.

$$Z = \sqrt{R^2 + (X_T)^2} \tag{5-12}$$

where,
Z is the impedance in Ω,
R is the resistance in Ω,
X_T is the total reactance in Ω.

Reactance comes is two forms. Capacitive reactance causes the voltage to lag the current in phase. Inductive reactance causes the current to lag the voltage in phase. The total reactance is the sum of the inductive and capacitive reactance. Since they are different in sign one can cancel the other, and the resultant phase angle between voltage and current will be determined by the dominant reactance.

In mechanics, a spring is a good analogy for capacitive reactance. It stores energy when it is compressed and returns it to the source. In an electrical circuit, a capacitor opposes changes in the applied voltage. Capacitors are often used as filters for passing or rejecting certain frequencies or smoothing ripples in power supply voltages. Parasitic capacitances can occur when conductors are placed in close proximity.

$$X_C = \frac{1}{2\pi f C} \tag{5-13}$$

where,
X_C is the capacitive reactance in Ω,
f is frequency in Hz,
C is capacitance in F.

In mechanics, a moving mass is analogous to an inductive reactance in an electrical circuit. The mass tends to keep moving when the driving force is removed. It has therefore stored some of the applied energy. In electrical circuits, an inductor opposes a change in the current flowing through it. As with capacitors, this property can be used to create useful filters in audio systems. Parasitic inductances can occur due to the ways that wires are constructed and routed.

$$X_L = 2\pi f L \tag{5-14}$$

where,
X_L is the inductive reactance in Ω.

Inductive and capacitive reactance produce the opposite effect, so one can be used to compensate for the other. The total reactance X_T is the sum of the inductive and capacitive reactance.

$$X_T = X_L - X_C \tag{5-15}$$

Note that the equations for capacitive and inductive reactance both include a frequency term. Impedance is therefore frequency dependent, meaning that it changes with frequency. Loudspeaker manufacturers often publish impedance plots of their loudspeakers. The impedance of interest from this plot is usually the *nominal* or *rated* impedance. Several standards exist for determining the rated impedance from the impedance plot, Fig. 5-13.

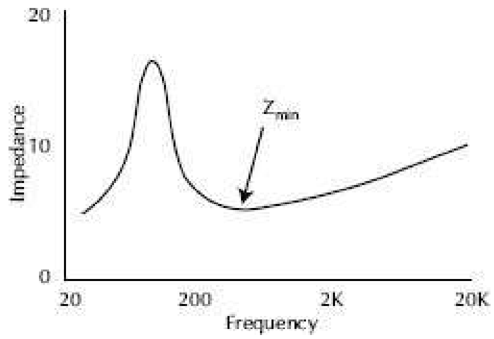


Figure 5-13. An impedance magnitude plot displays impedance as a function of the applied frequency. Courtesy SynAudCon.

An impedance phase plot often accompanies an impedance magnitude plot to show whether the loudspeaker load is resistive, capacitive, or inductive at a given frequency. A resistive load will convert the applied power into heat. A reactive load will store and reflect the applied power. Complex loads, such as loudspeakers, do both. When considering the power delivered to the loudspeaker, the impedance *Z* is used in the power equation. When considering the power dissipated by the load, the resistive portion of the impedance must be used in the power equation. The power factor describes the reduction in power transfer caused by the phase angle between voltage and current in a reactive load. Some definitions are useful.

$$Apparent\ Power\ (Total\ Power) = \frac{E^2}{Z} \qquad (5\text{-}16)$$

$$Active\ Power\ (Absorbed\ Power) = \frac{E^2}{R} \qquad (5\text{-}17)$$

$$Reactive\ Power\ (Reflected\ Power) = \frac{E^2}{Z\cos\theta} \qquad (5\text{-}18)$$

where,
θ is the phase angle between the voltage and current.

Ohm's Law and the power equation in its various forms are foundation stones of the audio field. One can use these important tools for a lifetime and not exhaust their application to the electrical and acoustical aspects of the sound reinforcement system.

5.9 Human Hearing

It is beneficial for sound practitioners to have a basic understanding of the way that people hear and perceive sound. The human auditory system is an amazing device, and it is quite complex. Its job is to transduce fluctuations in the ambient atmospheric pressure into electrical signals that will be processed by the brain and perceived as sound by the listener. We will look at a few characteristics of the human auditory system that are of significance to audio practitioners.

The dynamic range of a system describes the difference between the highest level that can pass through the system and its noise floor. The threshold of human hearing is about 0.00002Pascals (Pa) at mid frequencies. The human auditory system can withstand peaks of up to 200Pa at these same frequencies. This makes the dynamic range of the human auditory system approximately

$$\begin{aligned} DR &= 20\log\frac{200}{0.00002} \\ &= 140\,\text{dB} \end{aligned} \qquad (5\text{-}19)$$

The hearing system can not take much exposure at this level before damage occurs. Speech systems are often designed for 80dB ref. 20μPa and music systems about 90dB ref. 20μPa for the mid-range part of the spectrum. Normal conversation is about 60dB ref. 20 μPa, which is a good target for the playback level of a speech system if it produces adequate signal-to-noise ratio.

Audio practitioners give much attention to achieving a flat spectral response. The human auditory system is not flat and its response varies with level. At low levels, its sensitivity to low frequencies is much less than its sensitivity to mid-frequencies. As level increases, the difference between low- and mid-frequency sensitivity is less, producing a more uniform spectral response. The classic equal loudness contours, Fig. 5-14, describe this phenomenon and have given us the weighting curves, Fig. 5-15, used to measure sound levels and represent them numerically in a way that correlates with human perception.

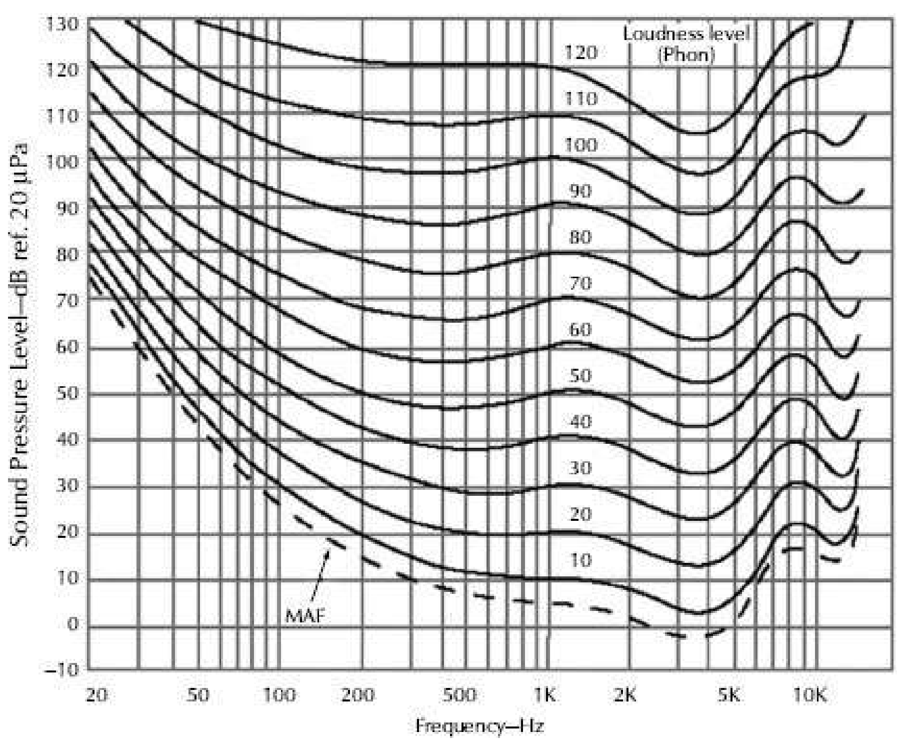


Figure 5-14. The equal-loudness contours. Courtesy SynAudCon.

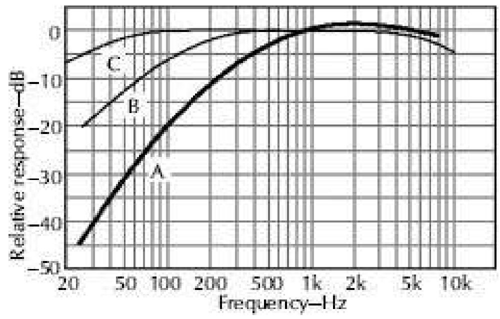


Figure 5-15. Weighting scales for measuring sound levels. Courtesy SynAudCon.

Modern sound systems are capable of producing very high sound pressure levels over large distances. Care must be taken to avoid damaging the hearing of the audience.

The time response of the hearing system is slow compared to the number of audible events that can occur in a given time span. As such, our hearing system *integrates* closely spaced sound arrivals (<50 ms) with regard to level. This is what makes sound indoors appear louder than sound outdoors. While reflected sound increases the perceived level of a sound source, it also adds colorations. This is the heart of how we perceive acoustic instruments and auditoriums. A good recording studio or concert hall produces a musically pleasing reflected sound field to a listener position. In general, secondary energy arrivals pose problems if they arrive earlier than 10ms (severe tonal coloration) after the first arrival or later than 50ms (potential echo), Fig. 5-16.

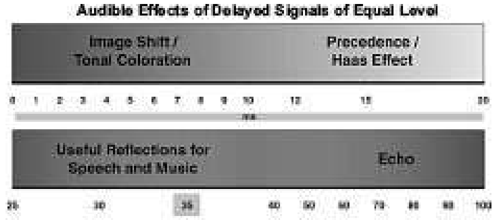


Figure 5-16. The time offset between sound arrivals will

determine if the secondary arrival is useful or harmful in conveying information to the listener. The direction of arrival is also important and is considered by acousticians when designing auditoriums. Courtesy SynAudCon.

The integration properties of the hearing system make it less sensitive to impulsive sound events with regard to level. Peaks in audio program material are often 20dB or more higher in level than the perceived loudness of the signal. Program material that measures 90dB A-weighted (slow response) may contain short term events at 110dB A-weighted or more, so care must be taken when exposing musicians and audiences to high powered sound systems.

The eardrum is a pressure sensitive diaphragm that responds to fluctuations in the ambient atmospheric pressure. Like a loudspeaker or microphone, it can be overdriven and damaged. The Occupational Safety and Health Administration (OSHA) is responsible for assuring that the workplace is in compliance regarding sound exposure. Sound systems are a major source of high level sounds and should work within OSHA guidelines. Tinnitus, or ringing in the ears, is one symptom of excessive sound exposure.

5.10 Monitoring Audio Program Material

The complex nature of the audio waveform necessitates specialized instrumentation for visual monitoring. Typical voltmeters are not suitable for anything but the simplest waveforms, such as sine waves. There are two aspects of the audio signal that are of interest to the system operator. The peaks of the program material must not exceed the peak output capability of any component in the system.

Ironically the peaks have little to do with the perceived loudness of the signal or the electrical or acoustic power generated by it. Both of these parameters are more closely tied to the rms (root-mean-square) value of the signal. Measurement of the true rms value of a waveform requires specialized equipment that integrates energy over a time span, much like the hearing system does. This integrated data will better correlate with the perceived loudness of the sound event. So audio practitioners need to monitor at least two aspects of the audio signal—its relative loudness (related to the rms level) and peak levels. Due to the complexity of true rms monitoring, most meters display an average value that is an approximation of the rms value of the program material.

Many audio processors have instrumentation to monitor either peak or average levels, but few can track both simultaneously. Many mixers have a VI (volume indicator) meter that reads in VU (volume units). Such meters are designed with ballistic properties that emulate the human hearing system and are useful for tracking the perceived loudness of the signal. Meters of this type all but ignore the peaks in the program material, making them unable to display the available headroom in the system or clipping in a component. Signal processors usually have a peak LED that responds fast enough to indicate peaks that are at or near the component's clipping point. Many recording systems have "true peak" meters that monitor the actual peaks but reveal little about the relative loudness of the waveform. They are useful for assuring that the recorded signal does not overdrive the recorder. A third meter type is the peak-program meter (PPM). This is a "quasi" peak meter that ignores very short term events whose clipping may not be audible. So, an audio meter may indicate loudness, peaks, or both.

Fig. 5-17 shows an instrument that monitors both the peaks and loudness of the audio program material. Both values are displayed in dB, and the difference between them is the approximate *crest factor* of the program material (the peak-to-rms ratio). Meters of this type yield a more complete picture of the audio event, allowing both loudness and available peak room to be observed simultaneously.

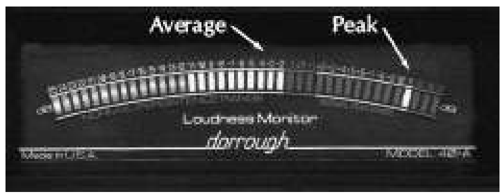


Figure 5-17. A meter that can display both average and peak levels simultaneously. Courtesy Dorrough Electronics.

5.11 Sound Propagation

Sound waves are emitted from acoustic sources—devices that move to modulate the ambient atmospheric pressure. Loudspeakers become intentional acoustic sources when they are driven with waveforms that cause them to vibrate. A point source is a device that radiates sound from one point in space. A true point source is an abstract idea and is not physically realizable, as it would be of infinitesimal size. This does not prevent the use of the concept to describe the characteristics of devices that are physically realizable.

Let us consider the properties of some idealized acoustic sources —not ideal in that they would be desirable for sound reinforcement use, but ideal in respect to their behavior as predictable radiators of acoustic energy.

5.11.1 The Point Source

A point source with 100% efficiency would produce 1 watt of acoustical power from one watt of applied electrical power. No heat would result, since all of the electrical power is converted. The energy radiated from the source would travel equally in all directions from the source. Directional energy radiation is accomplished by interfering with the emerging wave. Since interference would require a finite size, a true infinitesimal point source would be omnidirectional. We will introduce the effects of interference later.

Using 1pW (picowatt) as a power reference, the sound power level produced by 1 acoustic watt will be (5-20)

$$\begin{aligned} L_W &= 10\log\frac{1\,\mathrm{W}}{10^{-12}\,\mathrm{W}} \\ &= 120\,\mathrm{dB} \end{aligned} \tag{5-20}$$

Note that the sound power is not dependent on the distance from the source. A sound power level of L_W = 120dB would represent the highest continuous sound power level that could result from 1W of continuous electrical power. All real-world devices will fall short of this ideal, requiring that they be rated for efficiency and power dissipation.

Let us now select an observation point at a distance 0.282m from the sound source. As the sound energy propagates, it forms a spherical wave front. At 0.282m this wave front will have a surface area of one square meter. As such, the one watt of radiated sound power is passing through a surface area of 1 m^2.

$$L_I = 10\log\frac{1\,\mathrm{W/m^2}}{10^{-12}\,\mathrm{W/m^2}} \qquad (5\text{-}21)$$
$$= 120\,\mathrm{dB}$$

This is the sound intensity level L_I of the source and represents the amount of power flowing through the surface of a sphere of 1 square meter. Again, this is the highest intensity level that could be achieved by an omnidirectional device of 100% efficiency. L_I can be manipulated by confining the radiated energy to a smaller area. The level benefit gained at a point of observation by doing such is called the *directivity index* (DI) and is expressed in decibels. All loudspeakers suitable for sound reinforcement should exploit the benefits of directivity control.

For the ideal device described, the sound pressure level L_P (or commonly SPL) at the surface of the sphere will be numerically the same as the L_W and L_I (L_P = 120dB) since the sound pressure produced by 1W will be 20Pa. This L_P is only for one point on the sphere, but since the source is omnidirectional, all points on the sphere will be the same. To summarize, at a distance of 0.282m from a point source, the sound power level, sound intensity level, and sound pressure level will be numerically the same. This important relationship is useful for converting between these quantities, Fig. 5-18.

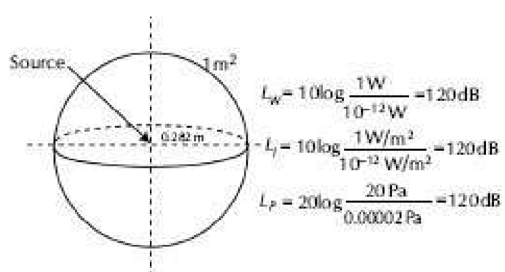


Figure 5-18. This condition forms the basis of the standard

terminology and relationships used to describe sound radiation from loudspeakers. Courtesy SynAudCon.

Let us now consider a point of observation that is twice as far from the source. As the wave continues to spread, its total area at a radius of 0.564m will be four times the area at 0.282m. When the sound travels twice as far, it spreads to cover four times the area. In decibels, the sound level change from point one to point two is

$$\begin{aligned}\Delta L_p &= 20\log\frac{0.564}{0.282}\\ &= 6\,\mathrm{dB}\end{aligned}$$

This behavior is known as the inverse-square law (ISL), Fig. 5-19. The ISL describes the level attenuation versus distance for a point source radiator due to the spherical spreading of the emerging waves. Frequency dependent losses will be incurred from atmospheric absorption, but those will not be considered here. Most loudspeakers will roughly follow the inverse square law level change with distance at points remote from the source, Fig. 5-20.

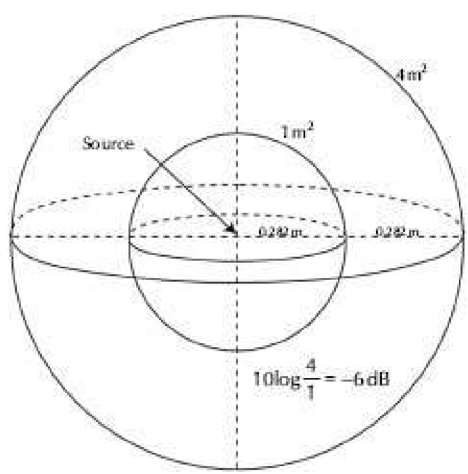


Figure 5-19. When the distance to the source is doubled, the radiated sound energy will be spread over twice the area. Both and

L_P will drop by 6dB. Courtesy SynAudCon.

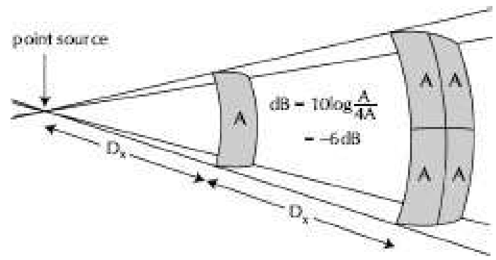


Figure 5-20. The ISL is also true for directional devices in their far field (remote locations from the device). Courtesy SynAudCon.

5.11.2 The Line Source

Successful sound radiators have been constructed that radiate sound from a line rather than a point. The infinite line source emits a wave that is approximately cylindrical in shape. Since the diverging wave is not expanding in two dimensions, the level change with increasing distance is half that of the point source radiator. The sound level from an ideal line source will decrease at 3dB per distance doubling rather than 6dB, Fig. 5-21. It should be pointed out that these relationships are both frequency and line length dependent, and what is being described here is the ideal case. Few commercially available line arrays exhibit this cylindrical behavior over their full bandwidth. Even so, it is useful to allow a mental image of the characteristics of such a device to be formed.

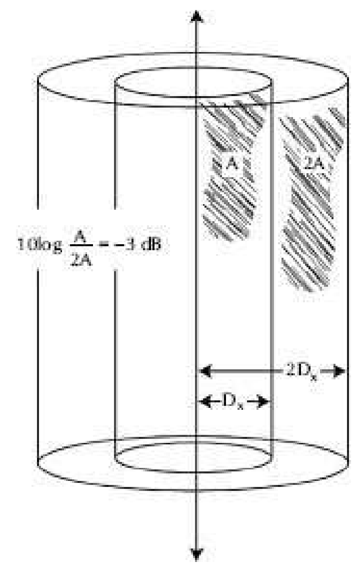


Figure 5-21. Line sources radiate a cylindrical wave (ideal case). The level drop versus distance is less than for a point source. Courtesy SynAudCon.

If the line source is finite in length (as all real-world sources will be), then there will be a phase differential between the sound radiated from various points on the source to a specific point in space. All of the points will be the most phase coherent on a plane perpendicular from the array and equidistant from the end points of the array, Fig. 5-22. As the point of observation moves away from the midpoint, phase interaction will produce lobes in the radiated energy pattern. The lobes can be suppressed by clever design, allowing the wave front to be confined to a very narrow vertical angle, yet with wide horizontal coverage. Such a radiation pattern is ideal for some applications, such as a broad, flat audience plane that must be covered from ear height. Digital signal processing has produced well-behaved line arrays that can project sound to great distances. Some incorporate an adjustable delay for each element to allow steering of the radiation lobe. Useful designs for auditoriums are at least 2 meters in vertical length.

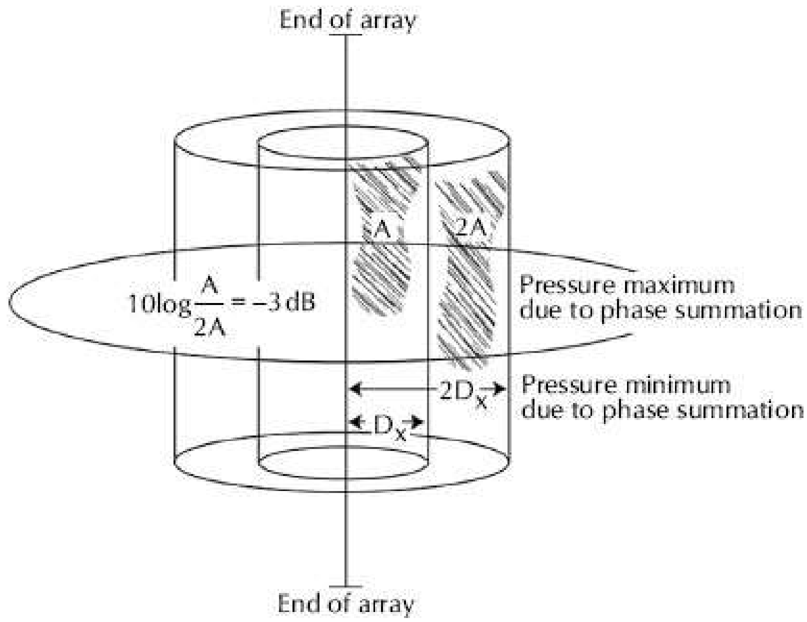


Figure 5-22. The finite line array has gained wide acceptance among system designers, allowing wide audience coverage with minimal energy radiation to room surfaces. Courtesy SynAudCon.

While it is possible to construct a continuous line source using ribbon drivers, etc., most commercially available designs are made up of closely spaced discrete loudspeakers or loudspeaker systems and are more properly referred to as line arrays.

5.12 Conclusion

The material in this chapter was carefully selected to expose the reader to a broad spectrum of principles regarding sound reinforcement systems. As a colleague once put it, "Sound theory is like an onion. Every time you peel off a layer another lies beneath it!" Each of these topics can be taken to higher levels, and many have been by other authors within this textbook. The reader is encouraged to use this information as a springboard into a life-long study of audio and acoustics. We are called upon to spend much of our time learning about new technologies. It must be remembered that new methods come from the mature body of principles and practices that have been handed down by those who came before us. Looking backward can have some huge rewards.

If I can see farther than those who came before me, it is because I am standing on their shoulders.

–Sir Isaac Newton

Bibliography

D. Davis and C. Davis, *Sound System Engineering*, Boston: Focal Press, 1997.

Glen Ballou, *Handbook for Sound Engineers*, 2nd Edition, Boston: Focal Press, 1991.

F. Alton Everest, *Master Handbook of Acoustics*, 3rd Edition, TAB Books, 1998.

Leo Baranek, *Acoustics*, New York: McGraw-Hill, 1954.

Harry Olson, *Acoustical Engineering*, Professional Audio Journals, Inc., Philadelphia, PA.

Part 2

Acoustics

Chapter 6

Small Room Acoustics

by Doug Jones

6.1 Introduction

The acoustics of small rooms is dominated by modes, shape, and reflection management. Acousticians who build large rooms are frequently frustrated with small room design because few of the tools of the trade that work in large rooms can be applied to small rooms. Getting small rooms to sound right involves both art and science. The science part is mostly straightforward. The creative part is quite subjective and a great sounding small room can be just as elusive as a great sounding concert hall.

6.2 The Behavior of Small Rooms as a Function of Wavelength

It is important to understand that small rooms must be treated differently with respect to frequency. Consider Fig. 6-1. These boundaries should not be understood as absolute or abrupt. They are meant to serve as guidelines and the transitions from one region to another are actually very gradual. Region 1 is the region from 0Hz up to the first mode associated with the longest dimension. In this region there is no support from the room at all, and there is not much one can do to treat the room. Region 2 is bounded by the first mode on the low end of the spectrum and f at the high end, where $f \cong 3C/RSD$ (rooms smallest dimension). In this region where the room modes dominate the acoustical performance, wave behavior is the best model and some forms of bass absorption can work well. Region 3 spans from f to roughly four times f. This region is dominated by diffraction and diffusion. This final region is where the wavelengths are generally small relative to the dimensions of the room. In this region, one can use a ray acoustics approach to simplify things as we are dealing with specular reflections.

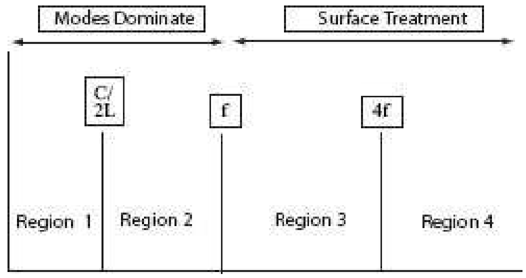


Figure 6-1. Regions for room treatment.

6.3 Room Modes

A room mode is a phenomenon that occurs whenever sound travels between two reflecting surfaces where the distance between the surfaces is such that the impinging wave reflects back on itself creating a standing wave. In 1896 Lord Rayleigh showed that the air enclosed in a rectangular room has an infinite number of normal or natural modes of vibration. The frequencies at which these modes occur are given by the following equation:[1]

$$f = \frac{c}{2}\sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2} \qquad (6\text{-}1)$$

where,

c is the speed of sound, 1130ft/s (or 344m/s),

L is the length of the room in ft (or m),

W is the width of the room in ft (or m),

H is the height of the room in ft (or m),

p, q, and *r* are the integers 0, 1, 2, 3, 4, and so on.

If we consider only the length of the room, we set *q* and *r* to zero, their terms drop out, and we are left with

$$
\begin{aligned}
f &= \frac{c}{2} - \sqrt{\left(\frac{1}{L}\right)^2} \\
&= \left(\frac{c}{2} - \frac{1}{L}\right) \\
&= \frac{1130}{2L} \qquad\qquad (6\text{-}2) \\
&= \frac{565}{L} \text{ in ft} \\
&= \frac{172}{L} \text{ in m}
\end{aligned}
$$

The distribution of modes determines the low frequency performance of a small room. Consider a sound source *S* emitting a sinusoidal signal between two isolated reflecting surfaces as in Fig. 6-2. Starting at a very low frequency, the frequency of the oscillator driving the source is slowly increased. When a frequency of f_0 = 1130/2*L* (*L* is in feet) is reached, a so-called standing wave condition is set up. Consider what is happening at the boundary. Particle velocity must be zero at the wall surface but wherever particle velocity is zero, pressure is maximum. The wave is reflected back out of polarity with itself, that is to say that the reflection is delayed by ½ of the period. This results in a cancellation that will occur exactly midpoint between the reflecting surfaces. If the walls are not perfect reflectors, losses at the walls will affect the heights of the maxima and the depths of the minima. In Fig. 6-2 reflected waves traveling to the left and reflected waves traveling to the right interfere, constructively in some places, destructively in others. This effect can be readily verified with a sound level meter which will show maximum sound pressures near the walls and a distinct null midway between the walls.

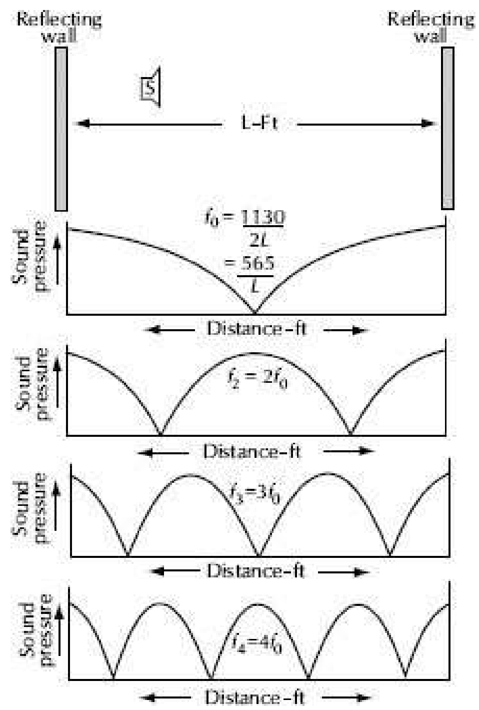


Figure 6-2. The simplest form of room resonance can be illustrated by two isolated, parallel, reflecting wall surfaces.

As the frequency of the source is increased, the initial standing-wave condition ceases, but at a frequency of $2f_0$ another standing wave appears with two nulls and a pressure maximum midway between the walls. Other standing waves can be set up by exciting the space between the walls at whole number multiples of f_0. These are called axial modes as they occur along the axis of the two parallel walls.

The two walls of Fig. 6-2 can be considered the east and west walls of a room. The effect of adding two more pairs of parallel walls to enclose the room is that of adding two more axial standing-wave systems, one along the east-west axis and the other along the north-south axis. In addition to the two axial systems that are set up, there will be a standing wave associated with two times the path length that involves all four surfaces. These modes are called tangential modes, Fig. 6-3. Most rooms of course will have at least six boundaries and there are modes that involve all six surfaces as well, Fig. 6-4. These modes are called oblique modes.

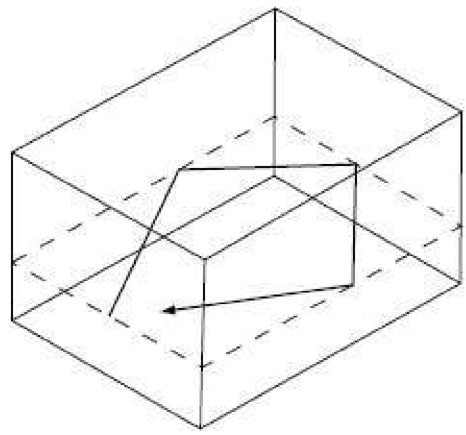


Figure 6-3. Tangential room modes.

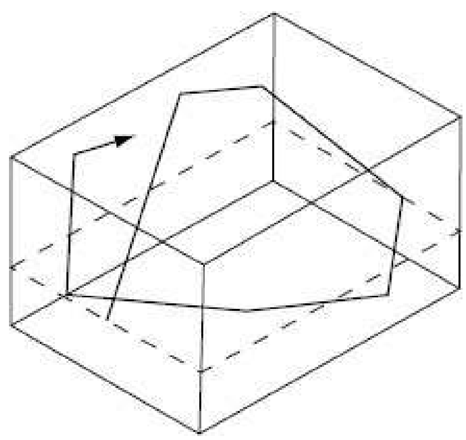


Figure 6-4. Oblique room modes.

Eq. 6-1 is a 3D statement based on the orientation of our room on the *x*, *y*, and *z* axes, as shown in Fig. 6-5. The floor of the room is taken as the *x* plane, and the height is along the *z* axis. To apply Eq. 6-1 in an orderly fashion, it is necessary to adhere to standard terminology. As stated, *p, q*, and *r* may take on values of zero or any whole number. The values of *p, q*, and *r* in the standard order are thus used to describe any mode. Remember that:

- p is associated with length L.
- q is associated with width W.
- r is associated with height H.

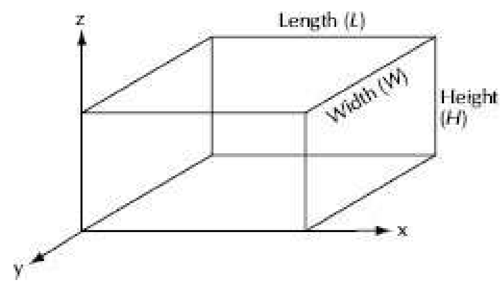


Figure 6-5. The floor of the rectangular room under study is taken to be in the *xy* plane and the height along the *z* axis.

We can describe the four modes of Fig. 6-2 as 1,0,0; 2,0,0; 3,0,0; and 4,0,0. Any mode can be described by three digits. For example, 0,1,0 is the first-order width mode, and 0,0,2 is the second-order vertical mode of the room. Two zeros in a mode designation mean that it is an *axial mode*. One zero means that the mode involves two pairs of surfaces and is called a *tangential mode*. If there are no zeros in the mode designation, all three pairs of room surfaces are involved, and it is called an *oblique mode*.

6.4 Room Mode Distribution

In order to better understand how to evaluate the distribution of room modes, we calculate the modal frequencies for three rooms. Let us first consider a room with dimensions that *are not* recommended for a sound room. Consider a room with the dimensions of 12ft long, 12ft wide by 12ft high (3.66m × 3.66m × 3.66m), a perfect cube. For the purposes of this exercise, let us assume that all the reflecting surfaces are solid and massive. Using

Eq. 6-1 to calculate only the axial modes (for now) we see a fundamental mode at 565/12 or 47.08Hz, Fig. 6-6. We saw in Section 6.2 above that modes dominate the acoustics of small rooms up to a frequency roughly equal to 3*C/RSD*. This means that we can stop calculating the modes at around 280 to 300Hz. But in order to illustrate the modal distribution, lets continue the series, looking at the 1,0,0; 2,0,0; 3,0,0…10,0,0 modes we see the results in Table 6-1.

Before we continue the calculation, let us examine what this table is indicating. The frequencies listed are those and only those that are supported by these two walls; that is to say there will be some resonance at these frequencies but at no others. When the source is cut off, the energy stored in a mode decays logarithmically. The actual rate of decay is determined by the type of mode and the absorptive characteristics of whatever surfaces are involved with that mode. An observer in this situation, making a sound with frequency content that includes 141Hz, may hear a slight increase in amplitude depending on the location in the room. The observer will also hear a slightly longer decay at 141Hz. At 155Hz, for example, there will be no support or resonance anywhere between these two surfaces. The decay will be virtually instantaneous as there is no resonant system to store the energy. Of course, in a cube the modes supported by the other dimensions (0,1,0; 0,2,0; 0,3,0 … 0,10,0 and 0,0,1; 0,0,2; 0,0,3… 0,0,10) will all be identical, Table 6-2.

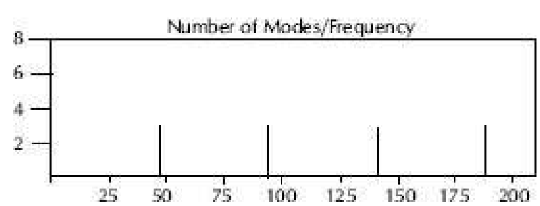


Figure 6-6. Number of axial modes and frequencies for a cube

room.

Table 6-1. Modal Frequencies for a 12 Ft Cube Room (Axial Only)

Length	Modes	Length	Modes
47.08	1,0,0	282.50	6,0,0
94.17	2,0,0	329.58	7,0,0
141.25	3,0,0	376.67	8,0,0
188.33	4,0,0	423.75	9,0,0
235.42	5,0,0	470.83	10,0,0

Table 6-2. Axial Modes in a Cube Supported in Each Dimension

Length	Modes	Width	Modes	Height	Modes
47.08	1,0,0	47.08	0,1,0	47.08	0,0,1
94.17	2,0,0	94.17	0,2,0	94.17	0,0,2
141.25	3,0,0	141.25	0,3,0	141.25	0,0,3
188.33	4,0,0	188.33	0,4,0	188.33	0,0,4
235.42	5,0,0	235.42	0,5,0	235.42	0,0,5
282.50	6,0,0	282.50	0,6,0	282.50	0,0,6
329.58	7,0,0	329.58	0,7,0	329.58	0,0,7
376.67	8,0,0	376.67	0,8,0	376.67	0,0,8
423.75	9,0,0	423.75	0,9,0	423.75	0,0,9
470.83	10,0,0	470.83	0,10,0	470.83	0,0,10

In the cube, all three sets of surfaces are supporting the same frequencies and no others. Talking in such a room is like singing in the shower. The shower stall supports some frequencies, but not others. You tend to sing at those frequencies because the longer decay at those frequencies adds a sense of fullness to the sound. Table 6-2 can be made more useful by listing all the modes in order to better examine the relationship between them. Table 6-3 is such a listing. In this table, we have included the spacing in Hz between a mode and the one previous to it.

Table 6-3. All of the Axial Modes of the Cube in Table 6-1 and Table 6-2

Frequency	Modes	Spacing	Frequency	Modes	Spacing
47.08	1,0,0		282.50	6,0,0	47.08
47.08	0,1,0	0.00	282.50	0,6,0	0.00
47.08	0,0,1	0.00	282.50	0,0,6	0.00
94.17	2,0,0	47.08	329.58	7,0,0	47.08
94.17	0,2,0	0.00	329.58	0,7,0	0.00
94.17	0,0,2	0.00	329.58	0,0,7	0.00
141.25	3,0,0	47.08	376.67	8,0,0	47.08
141.25	0,3,0	0.00	376.67	0,8,0	0.00
141.25	0,0,3	0.00	376.67	0,0,8	0.00
188.33	4,0,0	47.08	423.75	9,0,0	47.08
188.33	0,4,0	0.00	423.75	0,9,0	0.00
188.33	0,0,4	0.00	423.75	0,0,9	0.00
235.42	5,0,0	47.08	470.83	10,0,0	47.08
235.42	0,5,0	0.00	470.83	0,10,0	0.00
235.42	0,0,5	0.00	470.83	0,0,10	0.00

We can clearly see the triple modes that occur at every axial modal frequency, and there is 47Hz (equal to f_0) between each cluster. The space between each cluster is important because if a cluster of modes or even a single mode is separated by more than about 20Hz from its nearest neighbor, it will be quite audible as there is no masking from nearby modes. Consider another room that does not have a good set of dimensions for a sound room, but represents a typical room size because of standard building materials. Our next test room is 16ft long by 12ft wide with a ceiling height of 8ft (4.88m × 3.66m × 2.44m). Table 6-4 shows the length, width, and height modes, and Table 6-5 shows the same data sorted into order according to frequency.

If we examine the data in Fig. 6-7 we see that there are some frequencies which are supported by only one dimension. 35Hz, for example, is only supported by the 16ft (4.88m) dimension. Other frequencies, like 70Hz, occur twice and are supported by length and height. Still others like 141Hz occur three times and are supported by all three dimensions. In the nomograph of Fig. 6-7, the height of the line indicates the magnitude of the mode. This room is clearly better than a cube, but it is far from ideal as there are many frequencies which will stand out. 70 Hz, 141Hz, 211Hz, 282Hz, and 253Hz are all going to be problem frequencies in this room.

Table 6-4. Modes of a 16Ft × 12Ft × 8Ft Room Sorted by Frequency

Frequency	Modes	Spacing	Frequency	Modes	Spacing
35.31	1,0,0		282.50	8,0,0	35.31
47.08	0,1,0	11.77	282.50	0,6,0	0.00
70.63	2,0,0	23.54	282.50	0,0,4	0.00
70.63	0,0,1	0.00	317.81	9,0,0	35.31
94.17	0,2,0	23.54	329.58	0,7,0	11.77
105.94	3,0,0	11.77	353.13	10,0,0	23.54
141.25	4,0,0	35.31	353.13	0,0,5	0.00
141.25	0,3,0	0.00	376.67	0,8,0	23.54
141.25	0,0,2	0.00	423.75	0,9,0	47.08
176.56	5,0,0	35.31	423.75	0,0,6	0.00
188.33	0,4,0	11.77	470.83	0,10,0	47.08
211.88	6,0,0	23.54	494.38	0,0,7	23.54
211.88	0,0,3	0.00	565.00	0,0,8	70.63
235.42	0,5,0	23.54	635.63	0,0,9	70.63
247.19	7,0,0	11.77	706.25	0,0,10	70.63

Table 6-5. Data for a Room 23Ft × 17Ft × 9Ft

Frequency	Modes	Spacing	Frequency	Modes	Spacing
24.57	1,0,0		196.52	8,0,0	8.19
33.24	0,1,0	8.67	199.41	0,6,0	2.89
49.13	2,0,0	15.90	221.09	9,0,0	21.68
62.78	0,0,1	13.65	232.65	0,7,0	11.56
66.47	0,2,0	3.69	245.65	10,0,0	13.01
73.70	3,0,0	7.23	251.11	0,0,4	5.46
98.26	4,0,0	24.57	265.88	0,8,0	14.77
99.71	0,3,0	1.45	299.12	0,9,0	33.24
122.83	5,0,0	23.12	313.89	0,0,5	14.77
125.56	0,0,2	2.73	332.35	0,10,0	18.46
132.94	0,4,0	7.39	376.67	0,0,6	44.31
147.39	6,0,0	14.45	439.44	0,0,7	62.78
166.18	0,5,0	18.79	502.22	0,0,8	62.78
171.96	7,0,0	5.78	565.00	0,0,9	62.78
188.33	0,0,3	16.38	627.78	0,0,10	62.78

Now consider a room that has dimensions that might be well suited for an audio room; 23ft long by 17ft wide by 9ft high ceiling (7m × 5.18m × 2.74m). The sorted data and nomograph are shown in Fig. 6-8 and Table 6-6.

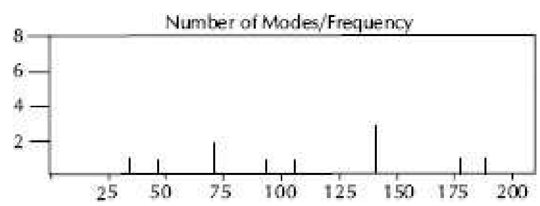


Figure 6-7. Number of modes and frequencies for a room 16ft × 12ft × 8ft.

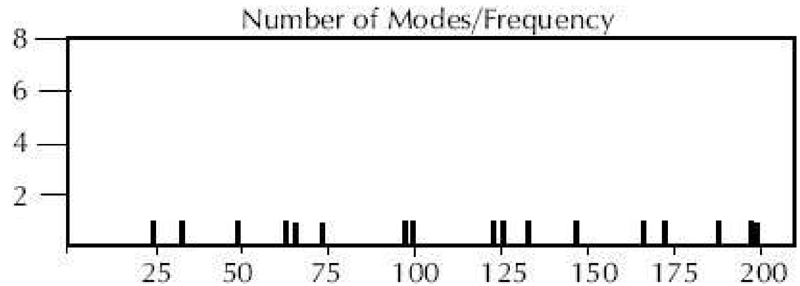


Figure 6-8. Number of modes and frequencies for a room 23ft × 17ft × 9ft.

The data in Fig. 6-8 look quite different from the data in Figs. 6-6 and 6-7. There are no instances where all three dimensions support the same frequency. There is also a reasonably good distribution of modes across the spectrum. There are a few places where the difference between the modes is quite small, like the space between the 4,0,0 and the 0,3,0. These three rooms, the cube, the room with dimensions determined by the builder, and the last room demonstrate the first important principle when dealing with room modes. The ratio of the dimensions determine the distribution of modes. The ratio is determined by letting the smallest dimension be 1 and dividing the other dimensions by the smallest. Obviously, the cube with its ratio of 1:1:1 results in an acoustic disaster. Even though a 12 ft (3.66m) cube will sound differently from a 30ft (9m) cube, both rooms will exhibit the same modal distribution and it is the distribution that overwhelmingly determines the low frequency performance of a small room. The second room determined by the dimension of common building materials has a ratio of 1:1.5:2.0. The ratio is determined by letting the smallest dimension be 1 and dividing the other dimensions by the smallest. So an 8ft by 16ft by 24ft room would have the ratio of 1:2:3. The third room, which seems reasonably good at this point, has a ratio of 1:1.89:2.56. From this we can see that in order to have a reasonable modal distribution one should avoid whole number ratios and avoid dimensions that have common factors.

6.4.1 Comparison of Modal Potency

To this point we have only considered the axial modes. The three types of modes, axial, tangential, and oblique differ in energy level. Axial modes have the greatest energy because there are the shortest distances and fewest surfaces involved. In a rectangular room tangential modes undergo reflections from four surfaces, and oblique modes six surfaces. The more reflections the greater the reflection losses. Likewise the greater the distance traveled the lower the intensity. Morse and Bolt[5] state from theoretical considerations that, for a given pressure amplitude, an axial wave has four times the energy of an oblique wave. On an energy basis this means that if we take the axial waves as 0dB, the tangential waves are −3 dB and the oblique waves are −6dB. This difference in modal potency will be even more apparent in rooms with significant acoustical treatment. In practice it is absolutely necessary to calculate and consider the axial modes. It is a good idea to take a look at the tangential modes because they can sometimes be a significant factor. The oblique modes are rarely potent enough in small rooms to make a significant contribution to the performance of the room.

6.4.2 Modal Bandwidth

As in other resonance phenomena, there is a finite bandwidth associated with each modal resonance. The bandwidth will, in part, determine how audible the modes are. If we take the bandwidth as that measured at the half-power points (−3dB or $\sqrt{2}$), the bandwidth is

$$\Delta f = f_2 - f_1 = \frac{k_n}{\pi} \tag{6-3}$$

where,

Δf is the bandwidth in Hz,

f_2 is the upper frequency at the −3dB point,

f_1 is the lower frequency at the −3dB point,

k_n is the damping factor determined principally by the amount of absorption in the room and by the volume of the room. The more absorbing material in the room, the greater k_n.

If the damping factor k_n is related to reasonable assumptions about the decay time of a small room, the expression for Δf becomes

$$\Delta f = \frac{6.91}{\pi T} = \frac{2.2}{T} \tag{6-4}$$

where,

T is the decay time of a mode in s.[1]

From Eq. 6-4 a few generalizations may be made. For decay times in the range of 0.3 to 0.5s, typical of what is found in small audio rooms, the bandwidth is in the range 4.4 to 7.3Hz. It is a reasonable assumption that most audio rooms will have modal bandwidths of the order of 5Hz. Referring to Table 6-6 it can be seen that in a few instances there are modes that are within 5Hz of each other. These modes will fuse into one and occasionally some beating will be audible as the modes decay. Modal frequencies which are separated on both sides by 20Hz or more will not fuse at all, and will be noticeable as well, although not as noticeable as a double or triple mode. Consider a room with the dimensions of 18ft × 13ft × 9ft (5.48m × 3.96m × 2.74m). The axial frequencies are listed in Table 6-7. There are some frequencies which double, such as 62Hz and 125Hz. These are obvious problems. However 282Hz is also a problem frequency because it is separated by more than 20Hz on either side.

6.5 Criteria for Evaluating Room Modes

So far we have shown that there are a few general guidelines for designing small rooms with good distribution of room modes. We know that if two or more modes occupy the same frequency or are bunched up and isolated from neighbors, we are immediately warned of potential coloration problems. Over the years, a number of authors have suggested techniques for the assessment of room mores and methods for predicting the low frequency response of rooms based on the distributions of room modes. Most notably Bolt,[2] Gilford,[3] Louden,[4] Bonello,[5] and D'Antonio[6] have all suggested criteria. We will examine Bonello's approach as an exampe.

Bonello's number one criterion is to plot the number of modes (all the modes, axial, tangential, and oblique) in one-third octave bands against frequency and to examine the resulting plot to see if the curve increases monotonically (i.e., if each one-third octave has more modes than the preceding one or, at least, an equal number). His number two criterion is to examine the modal frequencies to make sure there are no coincident modes, or, at least, if there are coincident modes, there should be five or more modes in that one-third octave band. By applying Bonello's method to the 23ft × 17ft × 9ft (7m × 5.18m × 2.74m) room, we obtained the graph of Fig. 6-9. The conditions of both criteria are met. The monotonic increase of successive one-third octave bands confirms that the distribution of modes is favorable.

Table 6-6. Axial Modes for a Room 18Ft × 13Ft × 9Ft

Frequency	Spacing	Frequency	Spacing
31.39		219.72	2.41
43.46	12.07	251.11	31.39
62.78	19.32	251.11	0.00
62.78	0.00	260.77	9.66
86.92	24.15	282.50	21.73
94.17	7.24	304.23	21.73
125.56	31.39	313.89	9.66
125.56	0.00	313.89	0.00
130.38	4.83	345.28	31.39
156.94	26.56	347.69	2.41
173.85	16.90	376.67	28.97
188.33	14.49	376.67	0.00
188.33	0.00	391.15	14.49
217.31	28.97		

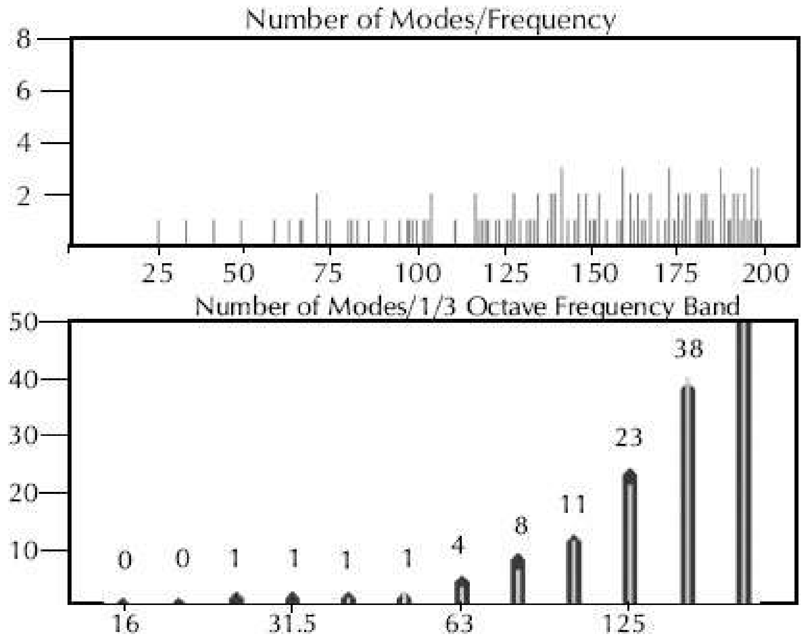


Figure 6-9. The application of Bonello's criterion 1 to the 23ft × 17ft × 9ft room.

It is possible that the critical bands of the ear should be used instead of one-third octave bands. Actually, one-sixth octave bands follow critical bandwidths above 500Hz better than do one-third octave bands. Bonello considered critical bands in the early stages of his work but found that one-third octave bands better show subtle effects of small changes in room dimensions.[5] Another question is whether axial, tangential, and oblique modes should be given equal status as Bonello does when their energies are, in fact, quite different. In spite of these questions, the Bonello criteria are used by many designers and a number of computer programs are using the Bonello criteria in determining the best room mode distributions.

D'Antonio et al, have suggested a technique which calculates the modal response of a room, simulating placing a measurement microphone in one corner of a room then energizing the room with a flat power response loudspeaker in the opposite corner.[6] The authors claim that this approach yields significantly better results than any other criteria.

Another tool which historically has been used to help choose room dimensions is the famous Bolt footprint shown in Fig. 6-10. Please note the chart to the right of the footprint which limits the validity of the footprint. The ratios of Fig. 6-10 are all referenced to ceiling height.[2]

6.6 Modes in Nonrectangular Rooms

Nonrectangular rooms are often built to avoid flutter echo and other unwanted artifacts. This approach is usually more expensive, therefore it is desirable to see what happens to modal patterns when room surfaces are skewed. At the higher audio frequencies, the modal density is so great that sound pressure variations throughout a rectangular room are small and there is little to be gained except, of course, the elimination of flutter echoes. At lower audio frequencies, this is not the case. The modal characteristics of rectangular rooms can be readily calculated from Eq. 6-1. To determine modal patterns of nonrectangular rooms, however, requires one of the more complex methods, such as the use of finite elements. This is beyond the scope of this Chapter. We, therefore, refer to the work of van Nieuwland and Weber of the Philips Research Laboratories of Eindhoven, the Netherlands, on reverberation chambers.[7]

In Fig. 6-11 the results of finite element calculations are shown for 2D rectangular and nonrectangular rooms of the same area (377 ft^2 or $35m^2$). The lines are contours of equal sound pressure. The heavy lines represent the nodal lines of zero pressure of the standing wave. In Fig. 6-11 the 1,0,0 mode of the rectangular room, resonating at 34.3Hz, is compared to a 31.6Hz resonance of the nonrectangular room. The contours of equal pressure are decidedly nonsymmetrical in the latter. In Fig. 6-10 the 3,1,0 mode of the rectangular room (81.1Hz) is compared to an 85.5Hz resonance in the nonrectangular room. Increasing frequency in Fig. 6-10, the 4,0,0 mode at 98Hz in the rectangular room is compared to a 95.3Hz mode in the nonrectangular room. Fig. 6-10 shows the 0,3,0 mode at 102.9Hz of a rectangular room contrasted to a 103.9Hz resonance in the nonrectangular room. These pressure distribution diagrams of Fig. 6-11 give an excellent appreciation of the distortion of the sound field by extreme skewing of room surfaces.

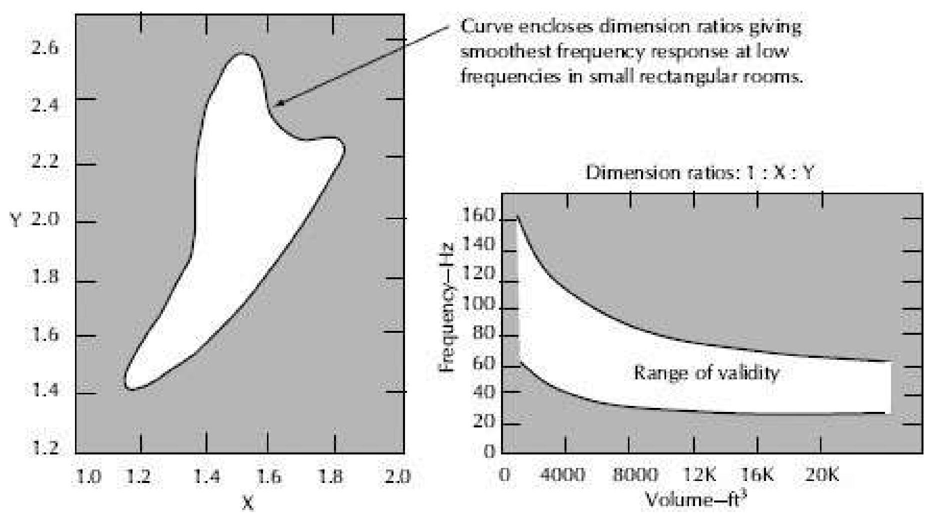


Figure 6-10. Room proportion criterion.

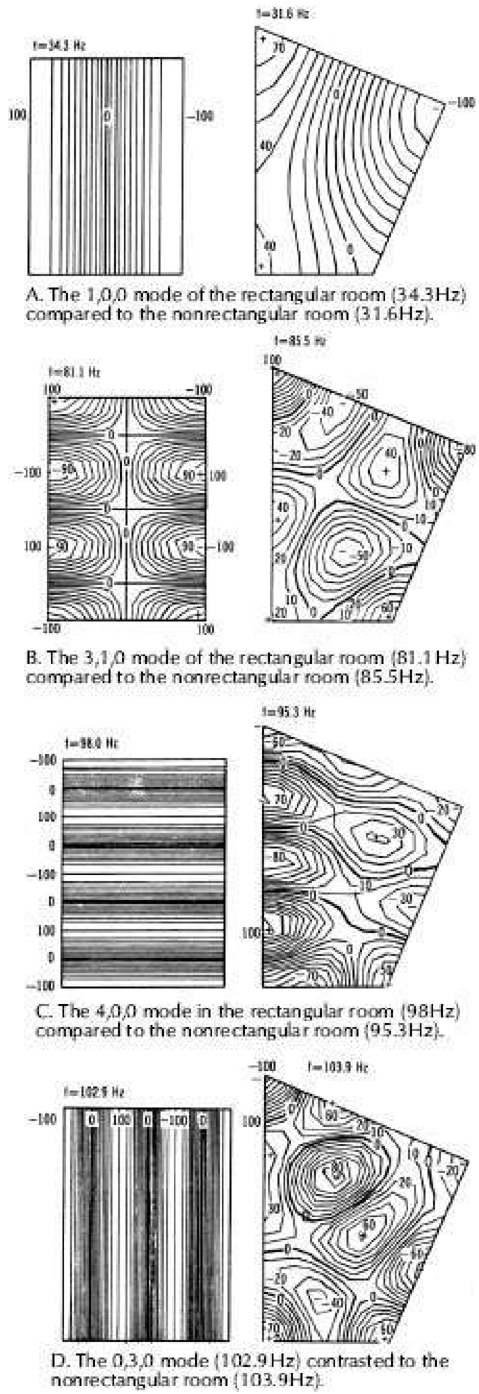


Figure 6-11. A comparison of calculated 2D sound fields in rectangular and nonrectangular rooms having the same areas.[8]

When the shape of the room is irregular as in Fig. 6-11, the modal pressure pattern is also irregular. The number of modes per frequency band in the irregular room is about the same as the regular room because it is determined principally by the volume of the room rather than its shape. Instead of axial, tangential, and oblique modes characteristic of the rectangular room, the resonances of the nonrectangular room all have the character of 3D (oblique like) modes. This has been demonstrated by measuring decay rates and finding less fluctuation from mode to mode. Note that the modes did not go away and that there was not a significant change in the frequency of the modes nor in the distribution of the modes relative to frequency. What changed was the distribution of the modes in the physical space. The benefits of asymmetrical, non-rectangular designs must be measured against the drawbacks as we shall see later on in this chapter.

6.7 Summation of Modal Effect

Room modes determine the performance of small rooms below $f_{l,}$ a frequency roughly equal to $F = 3C/RSD$ (rooms smallest dimension). The following criteria should be applied when evaluating room ratios or dimensions in terms of modal distribution. When considering axial modes, there should be no modes within 5Hz of each other, and no mode should be greater than 20Hz from another. Since the modal bandwidth in small rooms is approximately 5Hz, any modes that are within 5Hz of each other will effectively merge into one. Modes that are isolated by more than 20Hz will not have masking from any other modes nearby and will likely stand out. Obviously there should not be any double or triple modes. Some criteria should also be applied to all the modes, the axial, tangential, and oblique. There are many excellent tools for calculating modal distribution.

6.8 The Question of Reverberation in Small Rooms

W.C. Sabine, who first formulated the equation to calculate reverberation time, described reverberation in his seminal paper on the topic. He said "reverberation results in a mass of sound filling the whole room and incapable of analysis into its distinct reflections."[10] What Sabine was saying, although he did not use these terms, was that for true reverberation to exist, there needs to be a homogenous and isotropic sound field. Usually such conditions are approached in physically large rooms that do not contain much absorption. Unfortunately the term reverberation is popularly understood to be equivalent to decay. Does reverberation time refer to the decay of a well-established, totally homogenous, diffuse sound field that exhibits no net energy flow due to the richness of the reflections present or does reverberation time refer to the decay of any sound in a room no matter what the nature of the sound is, even if it is not diffuse? To some extent, this is a question of semantics. It is interesting to note that maybe Sabine himself perhaps anticipated the confusion that would eventually arise because in the same paper he wrote:

> *The word "resonance" has been used loosely as synonymous with "reverberation" and even with "echo" and is so given in some of the more voluminous but less exact popular dictionaries. In scientific literature the term has received a very definite and precise application to the phenomenon where ever it may occur. A word having this significance is necessary; and it is very desirable that the term should not, even popularly, by meaning many things, cease to mean*

anything exactly.[8]

It is the opinion of this author that this is precisely where we find ourselves today. Without a rigorous definition and application of the concept of reverberation, we are left with something which ceases to mean anything exactly.

When Sabine first measured the decay of the reverberation in Fogg Lecture Hall at Harvard University, he did it with an organ pipe and a stopwatch. He had no way of examining the fine detail of the reflections or any of the components of the sound field, nor was he initially looking at decay as a function of frequency. (Later on he looked at decay as a function of frequency, but never connected this to room size or shape.) He could only measure the decay rate of the 513Hz pipe he was using. The volume of the lecture hall was approximately 96,700ft^3.[9] The room was large enough that 512Hz was not going to energize any of the normal room modes. Since there was virtually no absorption in the room whatsoever, it is likely that Sabine was measuring a truly diffuse sound field. It is interesting to note that in Sabine's early papers he rarely mentions the dimensions other than the volume of the rooms he was working in. He was convinced that it was the volume of the room that was important. The mean free path was also central to his thesis. The *MFP* is defined as the average distance a given sound wave travels in a room between reflections.[10] The equation for finding the mean free path is

$$MFP = \frac{4V}{S} \qquad (6\text{-}5)$$

where,

V is the volume of the room,

S is the total surface area.

Consider a small room with dimensions of 12ft × 16ft × 8ft high (3.66m × 4.88m × 2.44m). This room will have a volume of 1536ft^3 (43.5m^3) and a total surface area of 832ft^2 (77.3m^2). Putting these numbers into Eq. 6-5 yields a result of a MFP of about 7.38ft. At the average speed of sound (1130ft/s or 344m/s) this distance will be covered in 0.00553s or 5.53ms. It is generally accepted that in small rooms, after approximately four to six bounces, a sound wave will have lost most of its energy to the reflecting surfaces and will become so diffuse as to be indistinguishable from the noise floor. This of course depends on the amount of absorption in the room. In very absorptive rooms there may not be even two bounces. In very live hard rooms a wave may bounce more than six times. In this room a single wave will take only 32.6ms to bounce five times and be gone. Compare this with a large room. Consider a room that is 200 ft long by 150 ft wide with a 40 ft ceiling (61m × 45.7m 12.2m). This room will have a *MFP* of 54.5ft (16.61m). It will take 241.3ms for a single wave to bounce five times and dissipate.

Sabine was not interested in the shape of the room or even in the distribution of the absorptive material. He focused on the statistical nature of the diffuse sound field and on the rate of decay. Other researchers looked at similar issues eventually dividing the time domain performance into smaller and smaller regions and examining their contributions to the subjective performance of rooms.

In the acoustically small room there will not be enough bounce to create the homogeneous mixing that is one of the hall marks of true reverberation much less achieve isotropy.

Fig. 6-12 is an Envelope Time Curve of a large, reverberant room, measured using time delay spectrometry. The left side of the graph represents *t*0 or the beginning of the measurement. Note that this is the point in time what the signal leaves the loudspeaker. If the microphone represents the observer in this system, the observer would not hear anything until $t(0) + t(x)$ where $t(x)$ is the time of takes for the sound to leave the loudspeaker and arrive at the observation point. In this measurement, that time is 50 ms because the loudspeaker was about 56ft (17m) away from the microphone. This first arrival is known as the direct sound because it is the sound energy that first arrives at the listener or microphone, before it reflects off of any surface. A careful examination of this graph shows a small gap between the direct sound and the rest of the energy arriving at the microphone. This is known as the initial time gap (ITG) and it is a good indicator of the size of the room. In this room it took about 50ms for the sound to travel from the loudspeaker to the microphone then another 40ms (90ms total) for sound to leave the speaker and bounce off of some surface to arrive at the microphone. Therefore, in this room the ITG is about 40ms wide.

Fig. 6-13 is an enlargement of the first 500ms of Fig. 6-12. The ITG can be clearly seen and is about 40ms long. The sound then takes about 130ms or so to build up to a maximum at around 270 ms. Fig. 6-12 shows that the sound then decays at a fairly even rate over the next 4s till the level falls into the noise floor.

If we perform a Schroeder integration[11] of the energy then measure the slope and extrapolate down to 60dB below the peak, we see the reverb time of this room to be on the order of 6.8s at 500Hz, Fig. 6-14.

It is useful to take a look at the ETC of a small room for comparison, Fig. 6-15.

Careful examination of Fig. 6-15 reveals a room dominated by strong discrete reflections that start coming back to the observation point within a few ms of the direct sound. By 30ms after the direct sound, the energy has decayed into the noise floor.

Since there is no significant diffuse or reverberant field in acoustically small rooms, equations having reverb time as a variable are not appropriate.

The discussion of how to quantify the decay of sound in small rooms continues. Most recording studios, control rooms, and listening rooms are too small to have a completely diffuse sound field, especially at the lower frequencies. In small, acoustically dead rooms the only frequencies that have any significant decay are those that are at or very near to the natural resonances or modes of the room, and will only be an issue if the mode is supported by 2 or more dimensions. This decay time is, strictly speaking, a relative of reverberation and perhaps can be treated as such. However the decay of the mode is not homogeneous, as the measurement will likely change as you move round the room. Furthermore, the intensity should be clearly observable and will not be seen to be isotropic. The Sabine equation and its offspring are not going to help in predicting how much absorption is going to be needed to modify such a room. One of the problems facing small rooms especially is the lack of a well-defined vocabulary of subjective descriptors which are mapped into objective measures.

The treatment of small room therefore cannot be accomplished with a statistical approach. Absorption cannot be specified simply on the basis of a desired total number of Sabines. More discussion and research are needed to be able to fully quantify the behavior of absorption in small rooms.

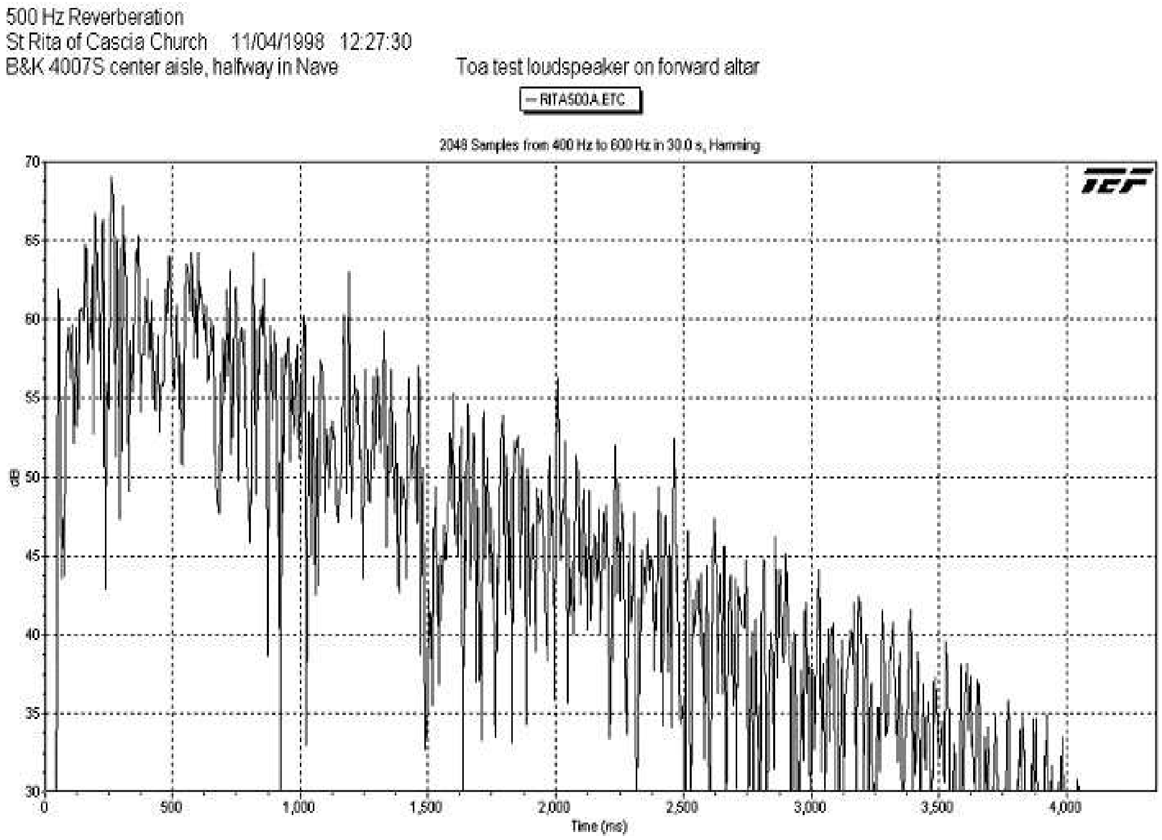


Figure 6-12. ETC of a large reverberant church. Measurement courtesy of Jim Brown.

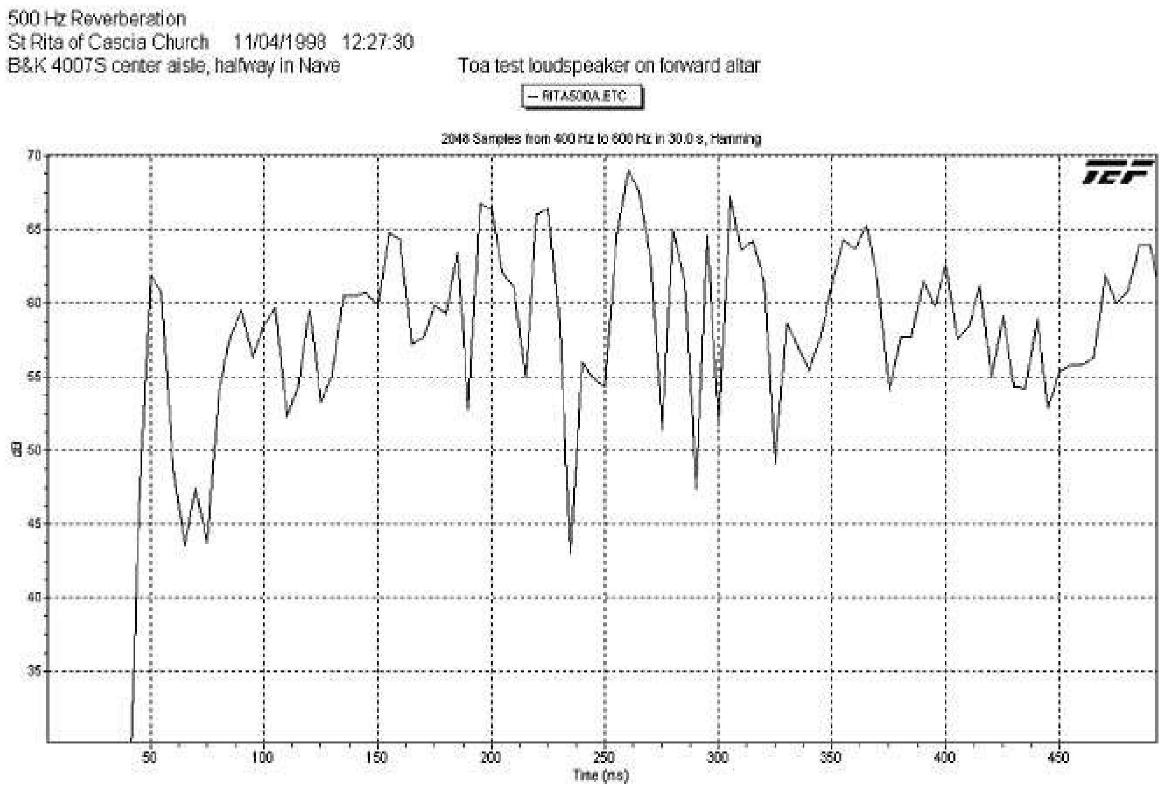


Figure 6-13. Enlargement of the first 500ms of Fig. 6-12.

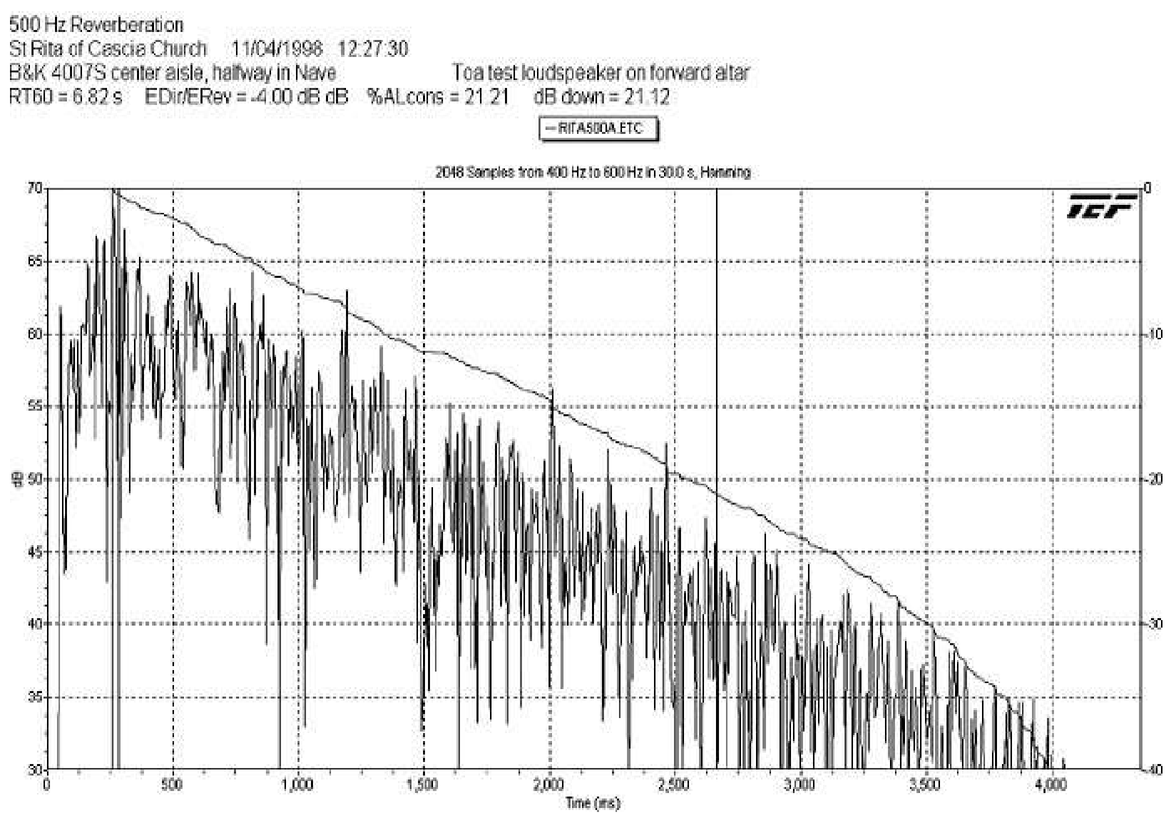


Figure 6-14. Schroeder integration of Fig. 6-12.

As is mentioned in Chapter 5, *Fundamentals of Audio and Acoustics*, the standard way to measure acoustical treatment is according to ASTM C 423-00. This is an indirect method that looks at the impact that the material has on a diffuse sound field. In rooms that do not exhibit any diffuse sound field in the frequency range of interest, we need another way to measure acoustic treatment. An alternative method is outlined in Chapter 8, *Acoustical Treatment for Indoor Areas*.

6.9 Room Shape

We have referred to the statistical approach (e.g., reverberation) and the wave approach (modes) to acoustical problems, and now we come to the geometrical approach. The one overriding assumption in the application of geometrical acoustics is that the dimensions of the room are large compared to the wavelength of sound being considered. In other words, we must be in region 3 of Fig. 6-1 where specular reflection prevails and the short wavelengths of sound allow treating sound like rays.

A sound ray is nothing more than a small section of a spherical wave originating at a particular point. It has a definite direction and acts as a ray of light following the general principles of geometric optics. Geometrical acoustics is based on the reflection of sound rays. This is where the shape of the room is the controlling acoustical aspect. Like the quest for room ratios the search for the perfect room shape is also elusive. Some have suggested that nonparallel surfaces are a must, however there are no perfect shapes. There are some shapes that work well for some applications.

6.9.1 *Reflection Manipulation*

In open space (air-filled space, of course) sound travels unimpeded in all directions from a point source. In real rooms we don't have point sources, we have loudspeakers or other sound sources such as musical instruments that do not behave like the theoretical point source. Real sources have characteristic radiation or directional patterns. Of course in real rooms the sound does not travel unimpeded for very long, depending on the MFP. After the sound leaves its source it will bounce off of some surface and will interact with the unreflected sound. This interaction can have a profound impact on the perception of the original sound. There is an elegant way to model the reflections in a room. The reflection can be considered to come from an image of the source on the opposite side of the reflecting surface and equidistant from it. This is the simple case: one source, one surface, and one image. If this reflecting surface is now taken to be one wall of a room, the picture is immediately complicated. The source now has an image in each of five other surfaces, a total of six images sending energy back to the receiver. Not only that, images of the images exist and have their effect, and so on. A physicist setting out to derive the mathematical expression for sound intensity from the source in the room at a given receiving point in the room must consider the contributions from images, images of the images, images of the images of the images, and so on. This is known as the image model of determining the path of reflections. The technique is fully described in Chapter 13, *Acoustical Modeling and Auralization*.

6.9.2 Comb Filters

When the direct sound and a reflection combine at some observation point, a spectral perturbation often called a comb filter is produced. The frequency of the first notch and the spacing of the rest of the notches is base on the delay between the two arrivals. The first notch *F* in hertz is calculated by

$$F = \frac{1}{2t} \qquad (6\text{-}6)$$

where,
t is the delay in seconds.

Each successive notch will be at

$$\frac{1}{t}$$

Fig. 6-15 shows the response of a system with a delay of 1.66ms between the two signals. Reflections can dramatically change the way program material sounds depending on the time of arrival, the intensity, and the angle of incidence relative to the listener. For a more in-depth treatment of how comb filters are created.[12]

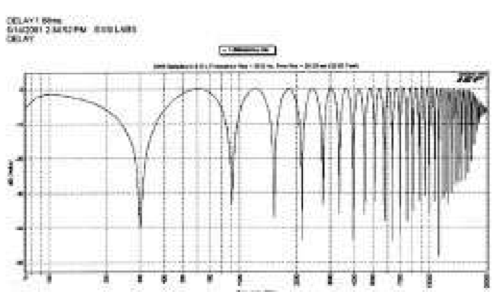


Figure 6-15. Response of a system with a delay of 1.66ms between the two sources.

In 1971 M. Barron wrote a paper exploring the effects of reflections on the perception of a sound.[13] He was trying to quantify the effects of lateral reflections in concert halls. Although his work was conducted in large reverberant spaces, a number of small room designers look to his work with great interest as he is considering the first 100ms of a sound field in a room. In small rooms that is often all you will get. Fig. 6-16 is a graphic summary of the effects of a single lateral reflection. It can be seen that the very early reflections, on the order of 0 to 5ms, can cause image shifts even when very low in amplitude relative to the direct sound. This can be important as one considers, for example, the accepted practice of placing loudspeakers on the meter bridge of a recording console.

Fig. 6-17A shows an ETC of a popular near field loudspeaker placed on the meter bridge of a recording console, measured at the mix position. The first spike is of course the direct arrival of the signal from the loudspeaker. The second spike is the reflection off of the face of the console, approximately 1.2ms later. Fig. 6-17B shows the resulting frequency response when these two signals arrive at the microphone. Finally Fig. 6-17C is the frequency response of three loudspeaker with the reflection removed. This author finds it curious indeed that this practice of placing a loudspeaker on the console ostensibly to remove the effects of the room and get a more accurate presentation actually results in seriously coloring the response of the speaker, and will have a significant impact on the ability to accurately perceive a stereo image.

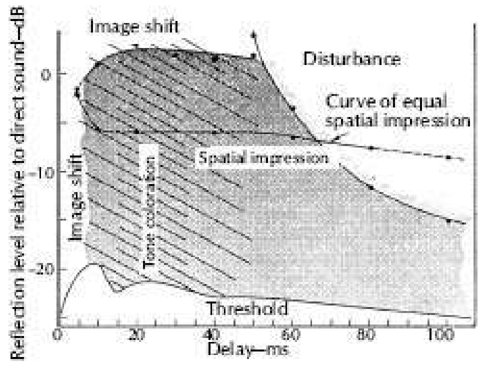


Figure 6-16. Graphic summary of the effects of a single lateral reflection.

Although Barron did not look at the effect of reflections arriving from below as in the case of a console reflection, the effect is clearly audible. In 1981 C.A.P. Rodgers[14] noted a similarity between the spectral notches created as a result of loudspeaker misalignment and those created by the pinna which have been shown to play an important role in localization. She postulated that the presence of spectral notches would impair or at least confuse the auditory system's ability to decode position cues. This could explain the phenomenon noted by Barron. The very early reflections are those that cause notches similar in spectral positioning to those caused by the pinna. These are the reflections that cause image shifts.

6.10 Small Room Design Factors

The basic tools for looking at small room performance have been addressed. We now turn our attention to small room design factors. We have divided small rooms into three broad categories; precision listening rooms, rooms for microphones, and rooms for entertainment. We are not trying to imply that there is only one way to build a control room or an entertainment room. There are different design criteria for different outcomes. The categories presented here are not intended to be exhaustive; rather they are intended to be general and representative. It should also be noted that we are not including the noise control issues that are often an important part of room design. The reader is referred to Chapter 3 for noise control information.

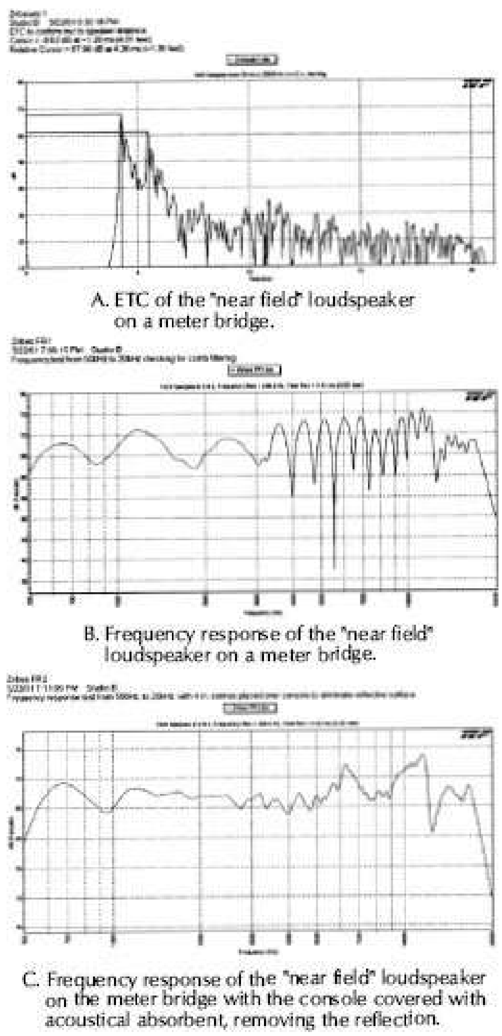


Figure 6-17. The effect of reflections bouncing off a studio console. Measurement courtesy of Mathew Zirbes.

6.10.1 The Control Room Controversy

Since most control rooms are acoustically small, it is appropriate to discuss control room design in general in this context. Some insist that control rooms should be as accurate as possible. Others insist that since music is rarely listened to in highly precise analytic rooms, recorded music would be better served if control rooms were more like entertainment rooms; not so sterile, but rather designed so that everything sounds subjectively great. Indeed many recordings are made in rooms that are not close to precision listening rooms. This debate will probably never be resolved as long as there are deductive and inductive reasoners, left-brain and right-brain people, artists and engineers. In the next few sections we are not attempting to solve this debate, rather we are trying to set out some simple guidelines. The most important task for the room designer is to listen to the client and not make assumptions about what it is he or she is looking for.

6.10.2 Precision Listening Rooms

These are rooms where the primary goal is for the listener to have as much confidence as possible that what is heard is precisely what is being or has been recorded. Frequently, users of these rooms are performing tasks that require listening analytically to the program and making decisions or judgments about what is heard. Examples of rooms in this category are recording control rooms, mastering suites, and audio production rooms. The state of the art at this writing does not permit us to design transducers or electronics that are perfect so as to afford the user 100% confidence that what is heard is fully equivalent to what has been or is being recorded. We can, however, design rooms that fully meet this criterion. An anechoic chamber would indeed be 100% neutral to the loudspeaker, allowing the user to hear precisely and only what is coming out of the speaker. The problem is that anechoic chambers are quite possibly the most acoustically hostile places we can imagine. It is difficult to spend a few minutes in an anechoic chamber let alone try to be creative and make artistic decisions about music in one. The challenge is to build a room that will not significantly interact with the loudspeaker by means of room modes or reflections that arrive at the listening position and still be a place that is subjectively acceptable to the user. There have been a number of good approaches to this problem over the years starting with LEDE™,[15] Reflection Free Zone™ (RFZ),[16] and the Selectively Anechoic Space™.[17] Later came Tom Hidley's neutral room or non-environment design and more recently, David Moulton has proposed his wide-dispersion design. These approaches all endorse attenuating or completely eliminating all the early reflections, creating a space that is essentially anechoic when energized by the loudspeakers and listened to in the prescribed position, but in all other ways it is an average room. Reflections can be eliminated or reduced at the listening position by changing the angle of the reflector, by using an absorber, or by using a diffuser. It should be noted that Angus questioned the use of diffusion in controlling lateral reflections.[18]

On the surface one might wonder why all sound rooms are not built this way. The reason is that most people do not listen to music analytically. In precision rooms, music that is poorly recorded will sound that way. One can certainly design rooms where the music sounds better than it does in a precision room. There are artifacts that one can build into a room that are subjectively very pleasing, but they are part of the *room* and not the recording. The recording engineer generally wants to know what exactly is in the recording.

The engineer generally listens to the product in a number of different environments before releasing it to insure that it does hold up even under non-ideal conditions.

So-called good sounding artifacts can be observed in the frequency domain as well as the time domain. For example, if a room has an audible room mode at 120Hz music might sound full and rich in the upper low end and be quite pleasing, however the fullness is in the room, not the recording. The recording may in fact be “thin” or lacking in the low end because the room is adding to the mix. In the time domain, a reflection that occurs in the first 10ms or so and comes from the side (a lateral reflection) might result in a perception of a stereo image that is much wider than the physical separation of the speakers might allow. This might be perceived as a very good sound stage, but it is an artifact of the room and not actually encoded in the recording.[19]

Designing such a room is an art and a science. It is beyond the scope of this chapter to detail a complete room design protocol, however, the steps in designing such a room must include:

1. Choosing a set of room ratios that yield a modal distribution that will result in the best possible low frequency performance.
2. Choosing a symmetrical room shape so that each loudspeaker interacts with the room in exactly the same way.
3. Choosing and placing acoustical treatment so that the early reflections (at least the first 18 ms) are attenuated at and are at least 18dB below the direct sound. Care should be taken to insure that the treatment chosen exhibits a flat absorption characteristic at the frequency range of interest and at the angles of incidence. The energy time curve should be measured at different positions to insure that the direct sound is not

compromised over the entire listening area.

4. Placing equipment and furniture in the room in such a way as to not interfere with the direct sound. It should be noted that the recording console is often the most significant acoustical element in the control room.
5. Insuring that there are enough live and diffusive surfaces in the room so that the overall subjective feel of the room is that of a normal room and not an anechoic chamber.

6.10.3 Rooms for Microphones

Designers are frequently asked to design rooms that are intended for recording or use with live microphones. Recording studios, vocal booths, and even conference rooms could be part of this category. The criteria in these rooms are almost all subjective. End users want rooms that sound good and that are comfortable to work in. The acoustician is well advised to work with a good interior decorator as a significant part of what makes someone feel comfortable in a room is the way the room is decorated and lit. Obviously noise control is a large part of the design criteria. There are a few general rules that will help with the acoustics of these small rooms:

1. Like the precision rooms, these rooms will work better if the proportions of the room result in optimal modal distribution.
2. Unlike the precision rooms, studios and vocal booths often work best when they are *not* symmetrical.
3. Avoid parallel surfaces if possible.
4. Use treatment that is as linear as possible, both statistically and by direct measurement of reflected sound.
5. Avoid treating entire surfaces with a single form of treatment. For example, covering an entire wall with an absorber will

usually be less effective than treating some areas and leaving some alone.

6. Listen carefully to the kinds of words the end user employs to describe the space either in terms of what is desired or in terms of something that need modification. Words like *intimate, close, dark, dead, quiet* are usually associated with the use of absorption. Words like *open, live, bright, airy* are often used in conjunction with diffusion.
7. Placing absorption in the same plane as the microphone will increase the apparent MFP and result in a longer ITG (initial time gap). This often makes the room seem larger. For example, in a vocal booth that is normally used by standing talent, place the absorption on the walls such that both the talent and the microphone are in the same plane as the absorptive area. In a conference room placing a band of absorption around the room at seated head height will help improve the ability to communicate in the room.

6.11 Rooms for Entertainment

There was a serious temptation to call this section "Rooms That Sound Good." The temptation was resisted to avoid the criticism that the section titles would thus imply that precision rooms don't sound good. It is a matter of goals. As was pointed out, the purpose of the precision room is analysis. This section will cover rooms that are designed for entertainment.

Of course it is much more difficult to set out design criteria for a good sounding room. As with any subjective goal it comes down to the tastes and preferences of the end user. To a great extent how one approaches an entertainment room depends on the type of system to be used, and the type of entertainment envisioned. An audiophile listening room will be treated differently from a home theater. It should be noted that in the world of home entertainment there exists a very rich audio vocabulary. Some of the words that are used like *spaciousness* and *localization* have meanings that are consistent with the use of these words in the scientific audio community. Subjective words like air, *grain, definition, impact*, and *brittleness* are much more ambiguous and are not yet mapped into the physical domain so that we know how to control them. One of the challenges is when the end user wants two mutually exclusive aspects optimized! The Nippon-Gakki experiments of 1979[20] quite elegantly showed how different subjective effects can be created by simply moving acoustic treatment to different locations in a room, Fig. 6-18. Note that when localization is rated good, spaciousness is rated poor and vice versa.

Some general points:

1. In home entertainment systems the distribution of room modes is somewhat less important. Having modal support in the low end although inaccurate can result in rooms that sound fuller. This might enhance a home theater system.
2. Absorption should be used sparingly. These rooms should be quiet, not dead. If absorption is to be used, it must be linear.
3. Remember that everything in a room contributes to the acoustics of the room. Most home entertainment rooms will have plush furniture that will be a significant source of absorption.

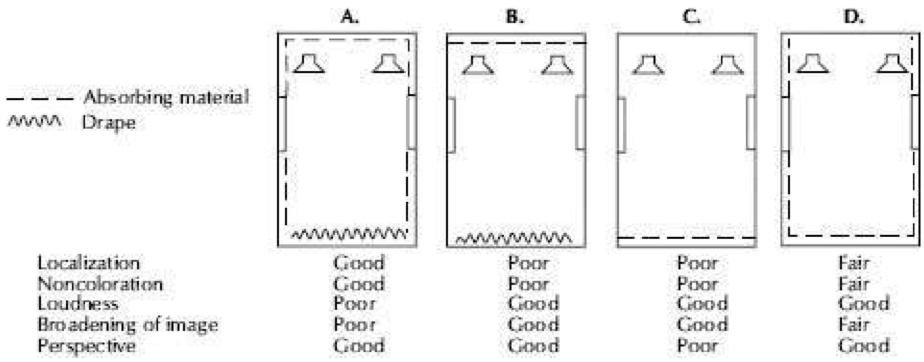


Figure 6-18. Summary of the results of the Nippon-Gakki psychoacoustical experiments.

The furnishings should be in place before the final treatment is considered.

1. Lateral reflections should be emphasized by using critically placed diffusers. Lateral reflections can dramatically increase the sense of spaciousness in a room.
2. Absorptive ceilings tend to create a sense of intimacy and a feeling of being in a small space. If this is not desired, use some absorption to control the very early reflections but leave the rest live.

References

1. L. L. Beranek, *Acoustics*, Second Edition, New York: McGraw-Hill, 1986.
2. R. H. Bolt, "Note on Normal Frequency Statistics for Rectangular Rooms," *JASA*, vol. 18 no. 1, July 1946, pp. 130–133.
3. O. J. Bonello, "A New Criterion for the Distribution of Normal Room Modes," *Journal of the Audio Engineering Society*, vol. 29, no. 9, pp. 597–606, September 1981.

4. M. M. Louden, "Dimension Ratios of Rectangular Rooms with Good Distribution of Eigentones," *Acoustica*, no. 24, pp. 101–104, 1971.

5. C. L. S. Gilford, "The Acoustic Design of Talk Studios and Listening Rooms," *Journal of the Audio Engineering Society*, vol. 27, pp. 17–31, 1979.

6. T. Cox and P. D'Antonio, "Room Sizing and Optimization at Low Frequencies," *Journal of the Acoustical Society of America*, vol. 52, no. 6, pp. 640–653, June 2004.

7. J. M. van Nieuwland and C. Weber, "Eigenmodes in Nonrectangular Reverberation Rooms," *Noise Control Engineering*, vol. 13, no. 3, pp. 112–121, November-December 1979.

8. W. C. Sabine, *Collected Papers on Acoustics*, Peninsula Publishing, Los Altos, CA, 1992, pp. 9.

9. Ibid. pp. 8 and 9.

10. L. L. Beranek, *Acoustics*, Second Edition, New York: McGraw-Hill, 1986.

11. M. R. Schroeder, "New Method of Measuring Reverberation Time," *Journal of the Audio Engineering Society*, vol. 35, pp. 299–305, 1987.

12. F. A. Everest, *Master Handbook of Acoustics*, Third Edition, TAB Books, 1998.

13. M. Barron, "The Subjective Effects of First Reflections in Concert Halls-The Need for Lateral Reflections," *Journal of Sound and Vibration*, vol. 15/4 pp. 475–494 (1971).

14. C. A. P. Rodgers "Pinna Transformations and Sound Reproduction," *Journal of the Audio Engineering Society*, Vol 29 pp. 226–234 (1981).

15. Don Davis and Chips Davis, "The LEDE Concept for the Control of Acoustic and Psychoacoustic Parameters in Recording Studio Control Rooms." *Journal of the AES*, vol. 28, no. 9, pp. 585, 595. September 1980

16. "The RFZ/RPG Approach to Control Room Monitoring." 76 AES Convention, NY 1984, preprint 2157.

17. D. Jones, G. Kendall, and W. Martins, "Designing a Room for Spatial Hearing Research using the TEF," Proceedings of 1984 International Computer Music Conference, IRCAM, Paris.

18. J. A. Angus "The Effects of Specular vs. Diffuse Reflections on the Frequency Response at the Listener," *Journal of the Audio Engineering Society*, vol. 49, no. 3, March 2001.

19. D. Jones, G. Kendall, and W. Martins, "Optimizing Control Rooms for Stereo Imagery," Paper presented at 1985 Acoustical Society Convention, Nashville TN.

20. Kishinaga, Shimizu, Ando, and Yamaguchi, "On the Room Acoustic Design of Listening Rooms," Audio Engineering Society 64th Convention (Nov. 1979) preprint #1524.

Chapter 7

Acoustical Noise Control

by Doug Jones

7.1 Introduction

This chapter will address issues pertaining to noise control in rooms. Sometimes there is confusion with respect to exactly what this means. In this chapter we are defining noise as unwanted sound. Noise control therefore means keeping unwanted sound out of a space. For the purposes of this chapter the term isolation is essentially synonymous with noise control.

Isolation, to a great extent, must be designed into a project. It can be very difficult, if not impossible to retroactively improve the isolation of a room.

It is often possible to change the way a room sounds subjectively by simply modifying the wall treatment. Wall treatment however generally has little impact on isolation. It is beyond the scope of this chapter to analyze what makes a sound unwanted. It is much more difficult to measure and quantify the extent to which any given sound annoys any given individual. We are letting the definition be purely subjective. To the Harley rider driving his Fat Boy™ past your studio, the exhaust is music to his ears but noise to you. Your music is noise to the therapist trying to conduct a session of a very different sort next door. The music in studio A is noise to the band trying to record in studio B down the hall.

Throughout this chapter, the term sound room will be used to designate any room that requires some measure of quiet in order for the room to serve its purpose.

Noise migrates from one area to another in two ways. It travels through the air and it travels through the structure. To reduce or eliminate airborne noise, one must eliminate all air paths between the spaces. To reduce structure-borne noise one must create isolation systems that eliminate mechanical connections between spaces. It is a rather simple matter to make these statements. Implementing the solutions is obviously much more difficult.

The first part of this chapter will present the most common metrics used in noise control. This will be followed by discussion of techniques and materials which may be useful in improving isolation using a case study approach.

7.2 Noise Criteria

If the definition of noise is unwanted sound, how can it be quantified? One way is to measure the ambient or background sound in a space. Assuming that whatever sound is there is undesired, (e.g. air conditioning noise, or being able to hear the train going by) we need to have a way to determine permissible levels. When specifying permissible noise levels, it is customary to use some form of the noise criteria (NC). The beauty of the NC contours is that a spectrum specification is inherent in a single NC number. The NC contours of Fig. 7-1 are helpful in setting a background noise goal for a sound room. Other families of NC contours have been suggested such as the PNC, Fig. 7-2 which adds an additional octave to the low end of the scale, and NR (noise rating), Fig. 7-3, used in Europe. In 1989 Beranek proposed the NCB or Balanced Noise Criteria.[1] The NCB adds the 16Hz octave band and the slopes of the curves are somewhat modified relative to the NC or PNC curves, Fig. 7-4. Beranek also proposed NCB limits for various applications as shown in Table 7-2.

Evaluating the spectrum of noise is far superior to using a single, wideband noise level. However, if desired, each NC contour can be expressed as an overall decibel level by adding the sound power in each octave band as in Table 7-1. These overall levels are convenient for rough appraisal of noise levels from a single sound level meter (SLM) reading. For example, if the SLM reads 29dB on the A-weighting scale for the background noise of a studio, it could be estimated that the NC of that room is close to NC-15 on the assumption that the noise spectrum of that room matched the corresponding NC contour, and that there are no dominant pure tone components.

Table 7-1. Noise Criteria (NC) Overall Levels*

NC Contour	Equivalent Wideband Level (A-weighted)
15	28
20	33
25	38
30	42
35	46
40	50
45	55
50	60
55	65
60	70
65	75

* Source: Rettinger[2]

It is helpful to see recommended NC ranges for recording studios and other rooms compared to criteria applicable to spaces used for other purposes, Table 7-2. The NC goals for concert halls and halls for opera, musicals, and plays are low to assure maximum dynamic range for music and greatest intelligibility for speech.

This same reasoning applies to high-quality listening rooms such as control rooms. For recording studios, stringent NC goals are required to minimize noise pickup by the microphone. Levels below NC-30 are generally considered "quiet," but there are different degrees of quietness. An NC-25 is at the low end of what is expected of an urban residence. This means that if an NC-25 were met in an urban residence, it is likely that the occupants would perceive it as being quiet. That same NC-25 represents the *upper* limit of what is acceptable for a recording studio. Most recording studios, especially those that record material with wide dynamic content, will require NC-20 or even 15. Levels lower than NC-15 require expensive construction and are difficult to achieve in urban settings.

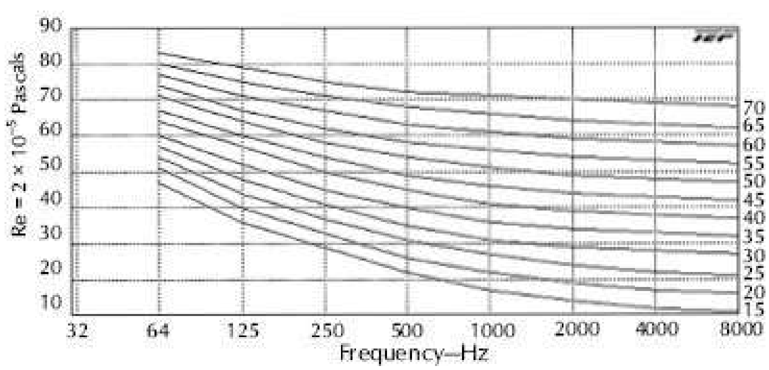


Figure 7-1. Noise criteria (NC) curves.

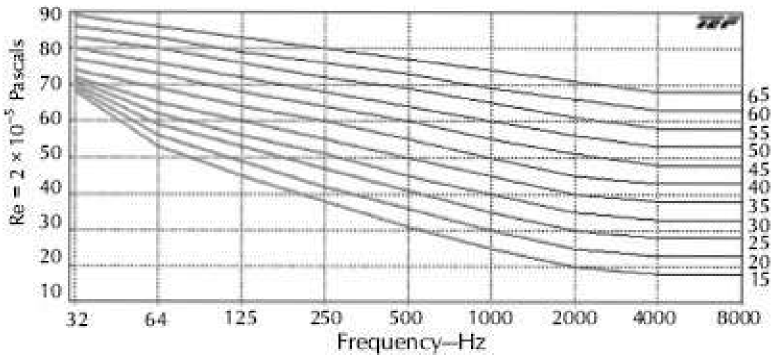


Figure 7-2. PNC. Note extra octave.

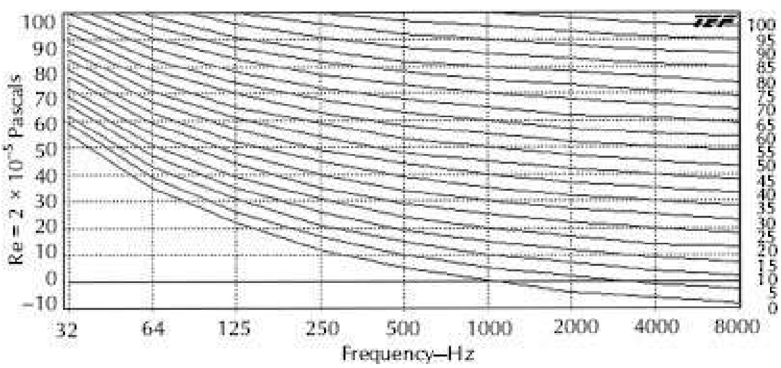


Figure 7-3. NR—The European noise rating curves. Note the extended range.

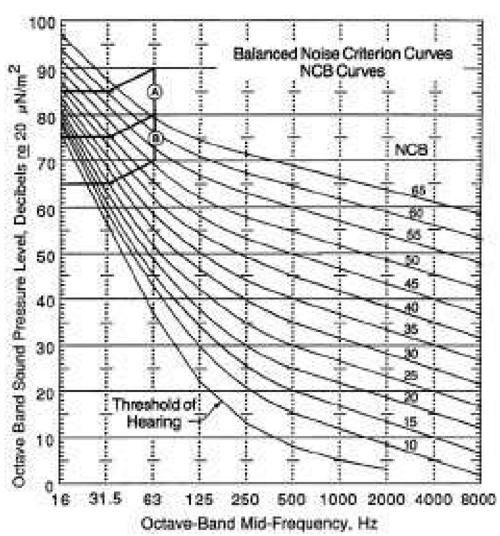


Figure 7-4. Balanced noise-criterion (NCB) curves for occupied room.

Table 7-2. Noise Criteria Ranges, NC and NCB*

Use of Space	Noise Criteria Range	NCB
Private urban residence, corridors	25 – 35	25 – 40
Private rural residence	20 – 30	na
Hotel rooms	30 – 40	25 – 40
Hospital, private rooms	25 – 35	25 – 40
Hospital, lobby, corridors	35 – 45	40 – 50
Office, executive	30 – 40	30 – 40
Offices, open	35 – 45	35 – 45
Restaurant	35 – 45	35 – 45
Church sanctuary	20 – 30	20 – 30
Concert, opera halls	15 – 25	10 – 15
Studios, recording and sound reproduction	15 – 25	10

* Selected from references 1, 3, and 4

7.2.1 Transmission Loss (TL)

Another tool for use in specifying or evaluating isolation is Transmission loss (TL). Transmission loss is the loss that occurs when a sound goes through a partition or barrier. A higher TL number means more loss, i.e., less acoustic energy gets through. If the desired NC or noise limit is known, and the noise load is known, a designer must then design barriers or partitions that have appropriate TL to meet the design goal. As is the case with virtually everything in acoustics the TL will vary dramatically with frequency, or more precisely the wavelengths involved.

7.2.2 Sound Transmission Class (STC)

The noise criterion approach is convenient and valuable because it defines a permissible noise level and spectrum by a single NC number. It is just as convenient and valuable to be able to classify the transmission loss of a barrier versus frequency curve by a single number. The STC or sound transmission class, is a single number method of rating partitions.[5] A typical standard contour is defined by the values in Table 7-3. A plot of the data in Table 7-3 is shown in Fig. 7-5. Only the STC-40 contour is shown in Fig. 7-5, but all other contours have exactly the same shape. It is important to note that the STC is not a field measurement. The field STC, or FSTC, is provided for in ASTM E336-97 annex a1. The FSTC is often 5dB or so worse than the laboratory STC rating. Therefore a door rated at STC-50 can be expected to perform around STC-45 when installed. Nonetheless, the STC provides a standardized way to compare products made by competing manufacturers.

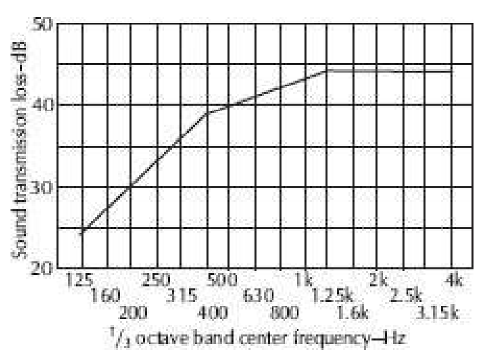


Figure 7-5. The standard shape used in determining the sound transmission class (STC) of a partition (ASTM E413-87).

Assume that a TL versus frequency plot of a given partition is at hand and that we want to rate that partition with an STC number. The first step is to prepare a transparent overlay on a piece of tracing paper of the standard STC contour (the STC-40 contour of Table 7-3 and Fig. 7-5) to the same frequency and TL scales as the

TL graph. This overlay is then shifted vertically until some of the measured TL values are below the contour and the following conditions are fulfilled:[6]

1. The sum of the deficiencies (i.e., the deviations below the contour) shall not be greater than 32dB.
2. The maximum deficiency at any single test point shall not exceed 8dB.

Table 7-3. Standard STC Contour

Frequency in Hz	⅓ Octave Sound Transmission Loss in dB	Frequency (Hertz)	⅓ Octave Sound Transmission Loss in dB
125	24	800	42
160	27	1000	43
200	30	1250	44
250	33	1600	44
315	36	2000	44
400	39	2500	44
500	40	3150	44
630	41	4000	44

When the contour is adjusted to the highest value that meets these two requirements, the sound transmission class of that partition is the TL value corresponding to the intersection of the contour and the 500Hz ordinate. An example of the use of STC is given in Fig. 7-6. To determine the STC rating for the measured TL curve shown in Fig. 7-6, the STC overlay is first aligned to 500Hz and adjusted vertically to read some estimated value, say, STC-44. The difference between the measured TL level and the STC curve is recorded at each of the octave points. These data are added together. The total, 47 dB, is more than the 32 dB al-lowed. The STC overlay is next lowered to an estimated STC-42, and a total of 37dB results. Lowering the overlay to STC-41 yields a total of 29dB, which fixes the STC-41 contour as the rating for the TL curve of Fig. 7-6.

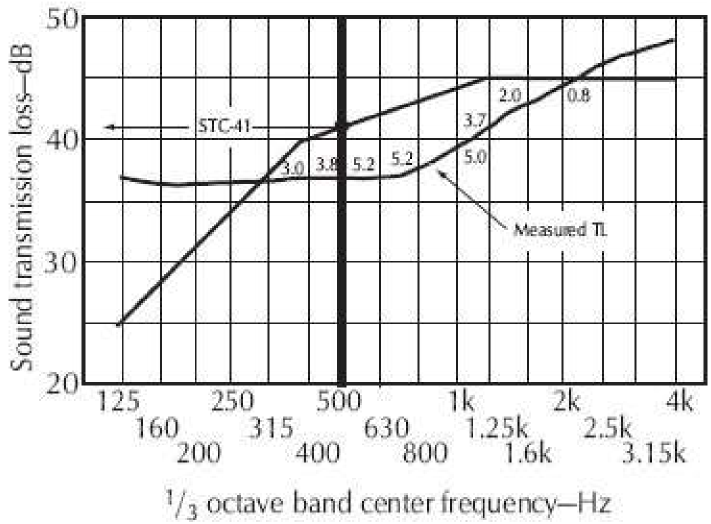


Figure 7-6. The method of determining the single-number STC rating of a barrier from its measured *TL* graph.

The final illustration of STC methods is given in Fig. 7-7. In this case, a pronounced coincidence dip appears at 2500Hz. This illustrates the second STC requirement, “the maximum deficiency at any single test point shall not exceed 8dB.” This 8dB requirement fixes the overlay at STC-39, although it might have been considerably higher if only the 32dB sum requirement applied.

The shape of the standard STC contour may be very different from the measured TL curve. For precise work, using measured, or even expertly estimated, TL curves may be desirable rather than relying on STC single number ratings. Convenience usually dictates use of the STC shorthand system, but it is, at best, a rather crude approximation to the real-world TL curves.

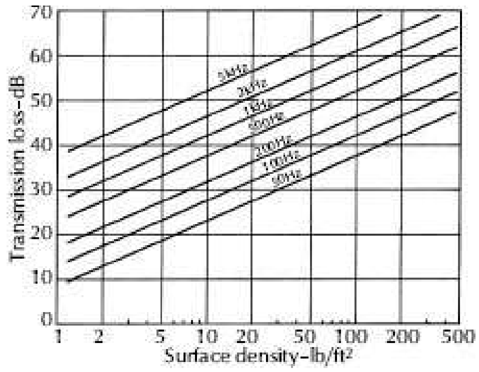


Figure 7-7. The second rule for STC determination that a maximum deficiency of 8dB is allowed.

7.2.3 Impact Insulation Class (IIC)

In multi-unit dwellings, noise from upper story units can be very disruptive to those living below. This noise most commonly is structure borne rather than air borne. There is some degree of commonality between airborne and structure borne noise, however structure borne noise is generally much harder to measure and control. ASTM E492.3 describes a machine known as a tapping machine which creates a standard test "signal" by dropping calibrated weights from a controlled height onto a surface, generally a floor. Measurements are made in the unit below. These measurements are then used to calculate the IIC according to a fairly complex method outlined in ASTM E1007-13b. The measurements are made in 16 bands from 100Hz to 3.15 kHz, and the levels are compared to a standard contour. The acoustics of the sending room and receiving room are also variables in this method. The result is a one number rating that can be used to describe the isolation between vertically stacked dwellings. It does a pretty good job of predicting impact sounds like high heeled shoes, but because of the 100Hz LF limit, it does not predict very well how much of the LF "thud" will come through.

7.2.4 Outdoor–Indoor Transmission Class (OITC)

The OITC is a standard test for rating the sound transmission effectiveness of walls, doors windows and other building elements. The OITC is similar to the STC in that it is a one number system for rating building elements, the higher the number, the higher the isolation. ASTM E1332 specifies that the measurements be made over a range of 80HZ to 4kHz and is calculated by subtracting the log summation of the TL values of a building element from the measured log sum of the impinging noise, or from a standard file.

7.3 A Case Study of Noise Control

In order to illustrate the major steps that must be taken to design a quiet space, we will take an imaginary design for a recording studio as a case study.

7.3.1 Site Selection

Part of meeting a noise goal is the careful selection of a building site, a site that is appropriate for the application and where the NC is achievable and affordable. It is one thing to build a room meeting an NC-15 in a cornfield in central Iowa. It is another thing altogether to build an NC-15 room in downtown Manhattan. When surveying a site, watch for busy roads, especially freeways; elevated, ground level, or underground railroads, busy intersections, airports, and fire stations. When economic or other factors make such a location imperative, allowance must be made for the extra cost of the structure to provide the requisite protection from such noise. When considering space in an existing building, inspect all neighboring spaces and be wary of adjacent spaces that are vacant unless the owner of the sound-sensitive space also controls the vacant space. Sometimes empty spaces can attract very noisy neighbors!

Buildings can be very noisy spaces. Sources of noise include elevator doors and motors, heating, ventilating, and air-conditioning equipment; heel taps on hard floors, plumbing, and business machines.

If selecting a plot of land, a limited amount of protection can be achieved by erecting earthen embankments or masonry walls between the structure and the noise source. These are reasonably effective at high frequencies, but low-frequency components of noise whose wavelengths are large relative to the size of the embankment tend to diffract over the top. A stand of dense shrubbery might yield as much as 10dB of overall attenuation. Physical separation of the proposed structure from the noise source is helpful but limited by the inverse-square law. The 6dB per distance double rule applies only to point sources in free-field conditions but it is useful for rough estimation. Going from 50ft to 100ft (a change of 50ft) from the source yields the same reduction of noise level as going from 100ft to 200ft (a change of 100ft). Clearly, increasing separation counts most when close to the source. At any given site, locating the sound-sensitive rooms on the face of the building away from a troublesome noise source is favorable, especially if no reflective structures are there to compromise the barrier effect.

As mentioned in the introduction there are two ways that noise travels from the source to the observer. It is either transmitted through the air—airborne noise—or carried through the structure or the earth—structure borne noise. A highway carrying heavy truck traffic or an overhead or subway railroad, may literally shake the earth to such an extent that large amplitude, low-frequency vibrations of the ground may be conducted to the foundation of the structure and carried to all points within that structure. Even if such vibrations are subsonic, they have been known to shake microphones with good low-frequency response so as to overload low level electronic circuits. Vibration, both subsonic and sonic, is carried with amazing efficiency throughout a reinforced concrete structure. The speed of sound in air is 344m/s whereas the speed of sound in reinforced concrete, for example, is on the order of 3700m/s.[7] A large-area masonry wall within a structure, when vibrated at high amplitude, can radiate significant levels of sound into the air by diaphragmatic action. It is possible by using a combination of vibration-measuring equipment and calculations (outside the scope of this treatment) to estimate the sound pressure level radiated into a room via such a structure-borne path. In most cases noise is transmitted to the observer by both air and structure.

7.3.2 Site Noise Survey

A site survey gives the designer a good idea of the noise levels present at the proposed building site. It is important to know how much noise exists in the immediate environment so that appropriate measures can be taken to reduce it to acceptable levels.

Ambient noise is a very complex, fluctuating mixture of traffic and other sounds produced by a variety of human and natural sources. The site noise should be documented with the appropriate test equipment. Subjective approaches are unsatisfactory. Even a modest investment in a studio suite or a listening room justifies the effort and expense of a site noise survey which provides the basis for designing walls, floor, and ceiling to achieve the desired noise criterion.

One approach to a noise survey of the immediate vicinity of a proposed sound room is to contract with an acoustical consultant to do the work and submit a report. If technically oriented persons are available, they may be able to turn in a credible job if supplied with the right equipment and given some guidance.

The easy way to survey a proposed site is to use one of the more sophisticated microprocessor-based recording noise analyzers available today. There are a number of fine units that are capable of producing reliable and very useful site surveys. Fig. 7-8 is an example of a 24 hour site survey made with the Gold Line TEF 25 running Noise Level Analysis™ software. One can also use a handheld sound level meter (SLM) if outfitted with the appropriate options. Some real-time frequency analyzers such as the Brüel and Kjaer 2143 are also appropriate. Dosimeters such as the Quest Technologies model 300 can also be used, but make sure that the dosimeter is capable of measuring the levels that you expect to find at the site. Dosimeters often will not measure levels below 40dB, Fig. 7-9.

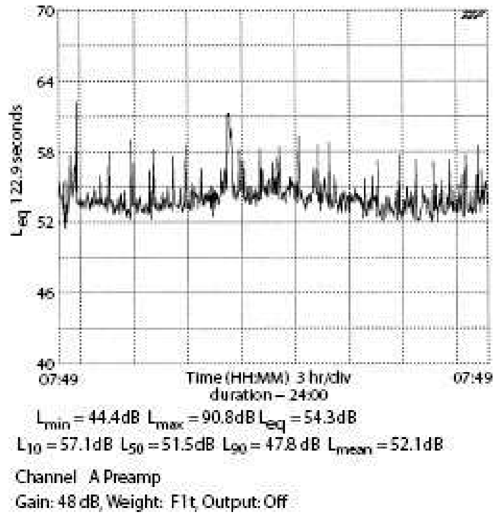


Figure 7-8. A 24 hour NLA site survey made with the Gold Line TEF 25.

No matter which analyzer is used, the system must be calibrated using a microphone calibrator. The weather conditions, especially temperature and relative humidity, should be noted at the time of calibration. The measuring microphone may be mounted in a weatherproof housing at the desired location with the microphone cable running to the equipment indoors. There are a number of terms which will appear on any display of a noise survey. There will be a series of L_n levels indicated, see Table 7-4. These are called *exceedance* or *percentile levels*. L_{10} refers to the noise level exceeded 10% of the time, L_{50} the level exceeded 50% of the time, L_{90} the level exceeded 90% of the time and so forth. In the United States L_{10} is considered to indicate the *average* maximum level and L_{90} the *average* minimum or background level.[8] Since many noise levels vary dramatically over time, it is useful to have a number which represents the equivalent constant sound pressure level. The L_{eq} is the steady continuous level that would yield the same energy over a given period of time as the measured levels. L_{dn} indicates a 24 hour L_{eq} with 10dB added to the levels accumulated between

2200 and 0700h to account for the increased annoyance potential during the nighttime hours. The Community Noise Equivalent Level (CNEL) also is used to document noise levels over a 24 hour period. It differs from the L_{dn} as weighting factors for the evening period between 1900 to 2200 h are included. The L_{eq} for evening hours is increased by 5dB while the L_{eq} for the nighttime hours is increased by 10dB.

Figure 7-9. Quest Technologies model 300 dosimeter. Courtesy Quest Technologies.

Table 7-4. Common Level Designations in Noise Surveys

L_{10}	Noise level exceeded 10% of the time
L_{50}	Noise level exceeded 50% of the time
L_{90}	Noise level exceeded 90% of the time
L_{dn}	24 hour L_{eq}
l_{mean}	Arithmetic mean of measured levels

The l_{mean} is the arithmetic mean of the measured levels. L_{min} and

L_{max} refer to the lowest and highest measured instantaneous levels, respectively.

Ideally, the site survey should take place over a minimum of 24 hours. A 24 hour observation captures diurnal variations; observations on selected days of the week capture especially noisy events varying from day to day or occurring at certain times during the week.

Most of the inexpensive analyzers will capture only the sound pressure levels over time. Others like the Brüel and Kjaer 2143, will also capture the spectrum of the noises as well. The 2143 will also capture the spectrum of the noise as well. If an analyzer such as the 2143 is not available, it is advisable to make a number of measurements of the spectrum as well as the time-stamped level record.

The data collected from the site survey should be combined with projections of the levels anticipated in the sound room and what will be tolerated in adjacent spaces.

7.3.3 Sound Barriers

The purpose of a sound barrier is to attenuate sound. To be effective, the barrier must deal with airborne as well as structure-borne noise. Each barrier acts as a diaphragm, vibrating under the influence of the sound impinging upon it. As the barrier vibrates, some of the energy is absorbed, and some is reradiated. The simplest type of barrier is the limp panel or a barrier without any structural stiffness. Approached theoretically, a limp panel should give a transmission loss increase of 6dB for each doubling of its mass. In the real world, this figure turns out to be nearer 4.4dB for each doubling of mass. The empirical mass law deduced from real-world measurements can be expressed as

$$TL = 14.5 \log M + 23 \qquad (7\text{-}1)$$

where,

TL is the transmission loss in dB,

M is the surface density of the barrier in lb/ft^2.

Fig. 7-10 is plotted from the empirical mass law stated in Eq. 7-1, which is applicable to any surface density and any frequency, as long as the mass law is operating free from other effects.

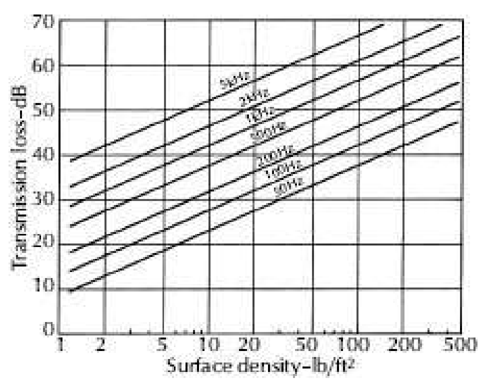


Figure 7-10. The empirical mass law based on real-world measurements of transmission loss. Surface density is the weight of the wall corresponding to a 1ft^2 wall surface.

From Fig. 7-10 several general conclusions can be drawn. One is that at any particular frequency, the heavier the barrier, the higher the transmission loss. A concrete wall 12in (30cm) thick with a surface density of 150lb/ft^2 (732kg/m^2) gives a higher transmission loss than a ¼in (6mm) glass plate with a surface density of 3lb/ft^2 (14.6kg/m^2). Another conclusion is that for a given barrier the higher the frequency, the higher the transmission loss.

The straight lines of Fig. 7-10 give only a partial picture since barrier effects other than limp mass dominate. Fig. 7-11 shows four different regions in the frequency domain of a barrier. At extremely low frequencies, stiffness of the barrier dominates. At somewhat higher frequencies, resonance effects control as the barrier vibrates like a diaphragm. Above a critical frequency, a coincidence effect controls the transmission loss of the barrier. The mass law is an important effect in determining barrier performance, but resonance and coincidence cause significant deviations.

The low-frequency resonance effect is due to the mechanical resonance of the barrier. For heavier barriers, the resonant-frequency is usually below the audible limit. As the panel vibrates at resonance, there is virtually no transmission loss. At frequencies above resonance, the mass law is in effect, and the function stays fairly linear until the coincidence effect. The coincidence effect occurs when the wavelength of the incident sound coincides with the wavelength of the bending waves in the panel. For a certain frequency and a certain angle of incidence, the bending oscillations of the panel will be amplified, and the sound energy will be transmitted through the panel with reduced attenuation. The incident sound covers a wide range of frequencies and arrives at all angles, but the overall result is that the coincidence effect creates an "acoustical hole" over a narrow range of frequencies giving rise to what is called the c*oincidence dip* in the transmission loss curve. This dip occurs above a critical frequency, which is a complex function of the properties of the material. Table 7-5 lists the critical frequency for some common building materials.

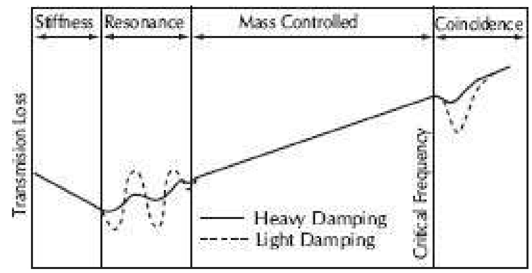


Figure 7-11. The performance of a barrier is divided into four regions controlled by stiffness, resonance, mass, and coincidence.

Table 7-5. Critical Frequencies

Material	Thickness (In)	Critical Frequency (Hz)
Brick wall	10	67
Brick wall	5	130
Concrete wall	8	100
Glass plate	¼	1600
Plywood	¾	700

*Calculated from Rettinger[7]

Assume that a goal of NC-20 has been chosen for a sound room. The noise survey indicates a noise level and spectrum as shown in Fig. 7-12. What wall construction will bring the noise of Fig. 7-12 down to the NC-20 goal we have set for the interior? Fig. 7-13 shows that a wall having a rating of STC-55 is required. The next step is to explore the multitude of possible wall configurations to meet the STC-55 requirement as well as other needs.

If the NC curve in Fig. 7-12 is subtracted from the measured noise curve, this will indicate the raw data that indicates the amount of loss needed to achieve the desired NC. This is plotted in Fig. 7-13. The standard STC template is laid over the graph and the needed STC is read opposite the 500Hz mark.

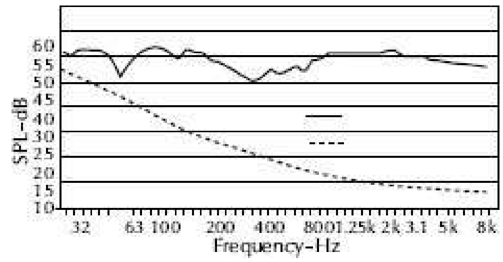


Figure 7-12. Noise spectrum from noise survey.

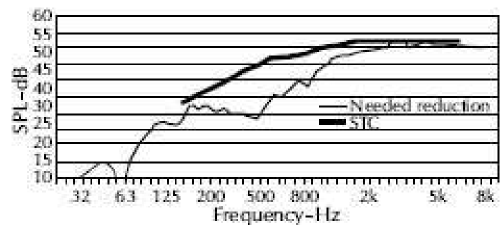


Figure 7-13. The sound barrier attenuation required for the sound room example of Fig. 7-12 is specified here as STC-47.

7.4 Isolation Systems

Isolation systems must be dealt with holistically. One must consider walls, ceilings, floors, windows, doors, etc. as parts of a whole isolation system. Vibration takes every path possible when traveling from one spot to another. For example, if one intends to build a sound room directly below a bedroom of another tenant in a building, one might assume that special attention must be paid to the ceiling. Of course this is correct. However, there are often paths that would permit the vibration to bypass the ceiling. All these flanking paths must be accounted for if isolation between two spaces is desired.

It should be noted that in some parts of the country (most notably California) building codes require seismic engineering. Make sure that the isolation systems that are under consideration do not violate any local seismic codes or require additional seismic restraints. Mason Industries has published a bulletin that is quite instructive in seismic engineering.[9]

7.4.1 Wall Construction

Acoustic partitions are complex entities. As was previously noted, walls exhibit different degrees of isolation in different segments of the spectrum. It is therefore imperative that you know what frequencies you are isolating. (Refer to Fig. 7-11.) The more massive the wall and the more highly damped the material, the fewer the problems introduced by diaphragmatic resonance. In comparing the relative effectiveness of various wall configurations, the mass law offers the most easily accessible rough approximation. However, most practical acoustical partitions actually perform better—that is, they achieve more loss—than what is predicted by the mass law. To assist in the computation of isolation based on mass, the densities of various common building materials are listed in Table 7-6. If an air space is added as in double wall construction, this introduces an element other than mass and generally leads to higher transmission loss.

Table 7-6. Building Material Densities

Material (inches)	Density (lb/ft^3)	Surface Density (lb/ft^2)
Brick	120	
4		40.0
8		80.0
Concrete: light wt.	100	
4		33.0
12		100.0
Concrete: dense	150	
4		50.0

12		150.0
Glass	180	
¼		3.8
½		7.5
¾		11.3
Gypsum wall-board	50	
½		2.1
⅝		2.6
Lead	700	
		3.6
Particle Board	48	
¾		1.7
Plywood	36	
¾		2.3
Sand	97	
1		8.1
4		32.3
Steel	480	
¼		10.0
Wood	24–28	
1		2.4

under one gypsum board face and/or mounting

7.4.2 High-Loss Frame Walls

The literature describing high TL walls is extensive. Presented here is a dependable, highly simplified overview of the data with an emphasis on practical solutions for sound room walls. The summary shown in Table 7-7 describes eight frame constructions including the STC performance of each.[10,11] In each of these constructions Gypsum wallboard is used because it provides an inexpensive and convenient way to get necessary wall mass and as fire retardant properties. Two lightweight concrete block walls, systems 9 and 10, fall in the general STC range of the gypsum wallboard walls 1 to 8, inclusive.

The three papers by Green and Sherry report measurements on many wall configurations utilizing gypsum wallboard.[10] Fig. 7-14 describes three of them yielding STCs from 56 to 62.

An expression of the empirical mass law stated as an STC rating rather than transmission loss[10] is shown in Fig. 7-15. This makes it easy to evaluate the partitions of Table 7-7 and Fig. 7-14 with respect to partition surface weight. The numbered STC shaded ranges of Fig. 7-15 correspond to the same numbered partitions of Table 7-7, and the A, B, and C points refer to the A, B, and C constructions of Fig. 7-14. From Fig. 7-15 it can be seen that the performance of wall types 1 and 9 can be predicted from the mass law. The other wall types perform better than what the mass law curve predicts. This better performance results primarily from decoupling one leaf of a structure from the other.

In recent years there have been new developments in wallboard. QuietRock™ is an internally damped wall-board.[11] Although it is considerably more expensive than standard gypsum board, it far outperforms conventional drywall, and for a given STC because less material is needed when using QuietRock, the cost can be offset.

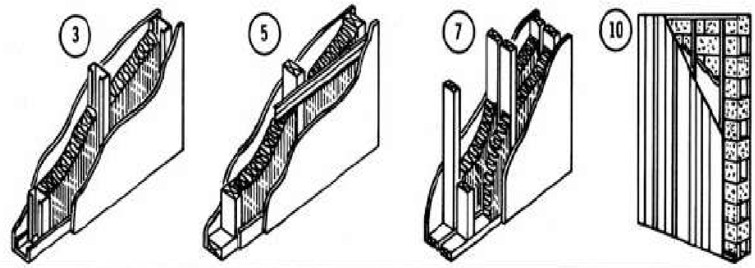
Following are ten points to remember concerning frame walls for highest STC ratings:

1. It is theoretically desirable to avoid having the coincidence dip associated with one leaf of a wall at the same frequency as that of the other leaf. Making the two leaves different with coincidence dips appearing at different frequencies should render their combined effect more favorable. However, Green

and Sherry found this effect negligible when partitions having equivalent surface weights were compared.[10]

2. The two leaves of a wall can be made different by utilizing gypsum board of different thickness, mounting a soft fiber (sound-deadening) board gypsum board on resilient channels on one side.

Table 7-7. Sound Transmission Class of Some Common Building Partitions*



Wall System	Test Sponsor[1]	Laboratory Test Reference	Surface Weight, lb/ft²	STC Rating: No Cavity Absorption	STC Rating: Cavity Absorption
1. Single-row 2 x 4 wood stud (16-inch on center), single-layer $\frac{5}{8}$-in gypsum board panels each side, *direct attached*	OCF NGC	OCF 424 & OCF 423 NGC 2403 & NGC 2166	— 6.1	34 35	36 ($3\frac{1}{2}$-inch glass fiber) 38 ($3\frac{1}{2}$-inch glass fiber)
2. *Same as 1*, except double-layer $\frac{1}{2}$-inch gypsum board each side	OCF	OCF W-23-69 & OCF W-25-69	—	39	45 ($3\frac{1}{2}$-inch glass fiber)
3. Single-row 24-gage $3\frac{5}{8}$-inch steel stud (24-inch on center) single $\frac{5}{8}$-in gypsum board panels each side, *direct attached*	NGC	NGC 2385 & NGC 2386	5.2	42	47 ($2\frac{1}{2}$-inch glass fiber)
4. *Same as 3*, except double-layer $\frac{1}{2}$-inch gypsum board each side	NGC	NGC 2282 & NGC 2288	8.9	48	53 (3-inch glass fiber)
5. Single-row 2 x 4 wood stud, single-layer $\frac{5}{8}$-inch gypsum board panels, *direct attached* one side, attached to metal *resilient channels* other side	FPL OCF GA	TL-73-72 OCF 431 & OCF 427 TL 77-138	6.4 — —	— 40 —	47 ($2\frac{1}{2}$-inch glass fiber) 46 ($3\frac{1}{2}$-inch glass fiber) 50 ($3\frac{1}{2}$-inch glass fiber)
6. *Same as 5*, except double-layer $\frac{5}{8}$-inch gypsum board each side	USG NGC	TL 67-212 & TL 67-239 NGC 2368 & NGC 2365	10.6 11.3	49 50	59 ($3\frac{1}{2}$-inch mineral fiber) 54 ($3\frac{1}{2}$-inch glass fiber)
7. Double-row 2 x 4 wood stud, 1-inch plate separation, single layer $\frac{5}{8}$-inch gypsum board each side	FPL OCF	TL 75-83 OCF W-43-69 & OCF 448	7.6 —	— 45	57 (double $3\frac{1}{2}$-inch glass fiber) 56 (3-inch glass fiber)
8. *Same as 7*, except double-layer gypsum board each side	FPL OLF	TL 75-82 OCF W-42-69 & OCF W-40-69	12.2 —	— 58	63 (double $3\frac{1}{2}$-inch glass fiber) 62 ($1\frac{1}{2}$-inch glass fiber)
9. 8-inch lightweight hollow concrete block both sides sealed with latex paint	ABPA	TL 70-16	34	46	—
10. *Same as 9*, with addition of furred out wall: $1\frac{5}{8}$-inch 24-gage metal studs, runners placed $\frac{1}{4}$-inch from concrete wall, covered with $\frac{1}{4}$-inch prefinished hardboard facing	ABPA	TL 70-14	36	—	57 ($1\frac{1}{2}$-inch mineral fiber)

[1] OCF—Owens-Corning Fiberglas Corporation
NGC—Gold Bond Building Products Division
FPL—USDA Forest Products Laboratory
GA—Gypsum Association
USG—United States Gypsum
ABPA—American Board Products Association
* Source: Reference 9

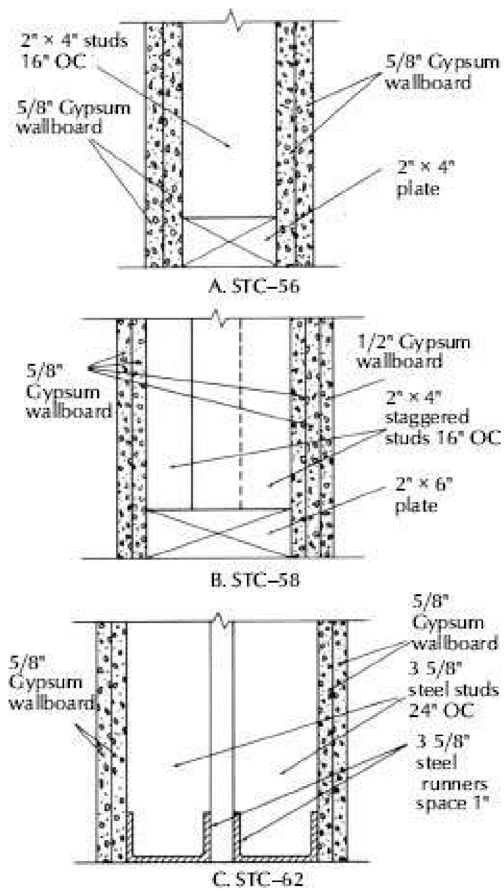


Figure 7-14. Three arrangements of gypsum wallboard two-leaf partitions having progressively higher STC ratings. (After Green and Sherry, Reference 10.)

3. Resilient channels are more effective on wood studs than on steel studs.
4. Steel stud partitions usually have an STC from two to ten points

higher than the equivalent wood stud partition. The flange of the common C-shaped steel stud is relatively flexible and transmits somewhat less sound energy from face to face.

5. If multiple layers of gypsum board are used, mounting the second layer with adhesive rather than screws can affect an STC increase by as much as six points. This is especially helpful with higher density walls.
6. A fiberglass cavity filler (such as R-7) may increase STC by five to eight points. It is more effective in multilayer partitions if the second layer is attached with adhesive.

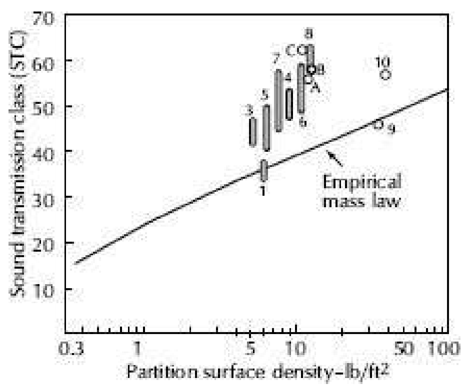


Figure 7-15. A variation of the empirical mass law expressed in terms of sound transmission class rather than TL. The numbers refer to the partitions of Table 7-7; the letters refer to Fig. 7-14.

7. A slight increase in STC results from increasing stud spacing from 16in to 24in on center.
8. Increasing stud size from 2in to 3in does not significantly increase either transmission loss or STC in steel-stud partitions with filler in the cavity.
9. Additional layers of gypsum wallboard increase STC and TL, but the greatest improvement is with lighter walls. Adding layers increases stiffness, which tends to shift the coincidence

dip to a lower frequency.

10. Attaching the first wallboard layer to studs with adhesive actually reduces STC.

7.4.3 Concrete Block Walls

Concrete block walls behave much like solid walls of the same surface weight and bending stiffness. In Table 7-7, wall system 9 is a lightweight, hollow, concrete block wall with both sides sealed with latex paint. In Fig. 7-15 we see that the performance of this specific wall falls close to pure mass operation. The STC-46 is matched or exceeded by many frame walls listed before it in Table 7-7. Wall system 10 in Table 7-7 is the same as 9, except wall system 10 has a new leaf, is furred out, and has mineral fiber added to one side in the cavity. These additions increase the STC from 46 to 57. It should be noted that there are less expensive frame structures that perform just as well. The performance of concrete walls can be improved by increasing the thickness of the wall, by plastering one or both faces, or by filling the voids with sand or well-rodded concrete, all of which increase wall mass. The STC performance of such walls can be estimated from Fig. 7-15 when the pounds-per-square-foot surface density is calculated. To further improve the performance one must add a furred-out facing (such as 10) or adding a second block wall with an air space.

7.4.4 Concrete Walls

The empirical mass law line in Fig. 7-15 goes to 100lb/ft^2 (488kg/m^2), just far enough to describe an 8inch concrete wall of 150b/ft^3 density (surface density 100lb/ft^2 or 732kg/m^2). This wall gives a rating close to STC-54. By extending the line we would find that a 12 inch wall would give STC-57, and a 24 inch concrete wall, about STC-61. The conclusion is inescapable. This brute-force approach to sound TL is not the cheapest solution. High TL concrete walls can be improved by introducing air space—e.g., two 8 inch walls spaced a foot or so apart. Such a wall requires specialized engineering talent to study damping of the individual leaves of the double wall, the coupling of the two leaves by the air cavity, the critical frequencies involved, the resonances of the air cavity, and so on.

7.4.5 Wall Caulking

There is continual movement of all building components due to wind, temperature expansion and contraction, hygroscopic changes, and deflections due to creep and loading. These movements can open up tiny cracks that are anything but tiny in their ability to negate the effects of a high-loss partition. An acoustical sealant is required to caulk all joints of a partition if the highest TL is to be attained. This type of sealant is a specialty product with nonstaining, nonhardening properties that provides a good seal for many years. Fig. 7-16 calls attention to the importance of bedding steel runners and wood plates in caulking to defeat the irregularities always present on concrete surfaces. A bead of sealant should also be run under the inner layer of gypsum board. The need for such sealing is as important at the juncture of wall-to-wall and wall-to-ceiling as it is at the floor line. The idea is to seal the room hermetically. Fig. 7-17 is a nomograph that illustrates what happens if there is leakage in a partition. The X axis represents a partition that is not compromised by any leaks. The family of curves are gaps or holes expressed as percentages of the whole surface area of the partition. This nomograph shows that a partition rated at a TL of 45 with no penetrations would perform as a TL-30 wall if only 0.1% of the wall were open. Consider what this means in real terms. A partition has a surface area of $10m^2$, 0.1% of $10m^2$ amounts to an opening with an area of a square centimeter (cm^2). This could be a gap in the wall/floor junction where the caulking was omitted, or it could be the area left open by the installation of an electrical box in a partition. This small gap will reduce the performance of the wall by a significant amount. All of the engineering and calculations that have been discussed so far can be rendered meaningless if sufficient care is not taken to seal *all* holes in a partition.

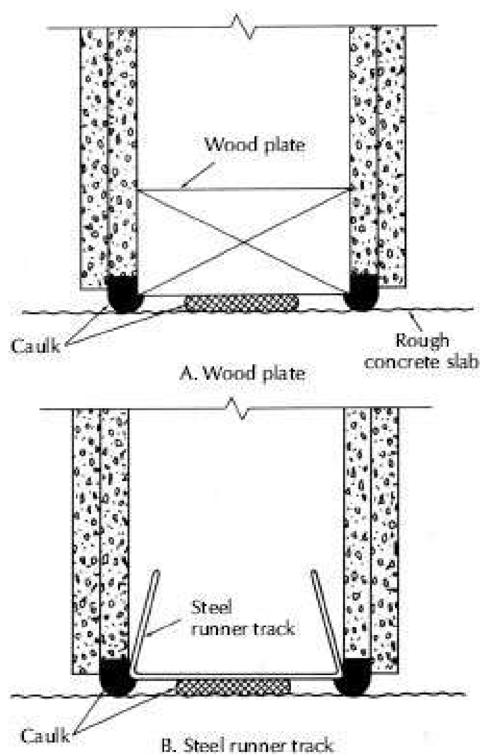


Figure 7-16. Caulking methods used for partitions.

7.4.6 Floor and Ceiling Construction

Building high TL walls around a sound room is futile unless similar attention is given to both the floor/ceiling system above the room and to the floor of the sound room itself. Heel and other impact noise on the floor above the room is readily transmitted through the ceiling structure and radiated into the sound room unless precautions are taken. The floor and ceiling structure of Fig. 7-18A is the type common in most existing frame buildings. Impact noise produced on the floor above is transmitted through the joists to the ceiling diaphragm below and radiated with little loss into the room below. Carpet on the floor above softens heel taps, but is low mass, and therefore has little effect on transmission of structure-borne sounds. Some decoupling of the floor membrane from the ceiling membrane is introduced in Fig. 7-18B in the form of resilient mounting of the ceiling gypsum board. Placing absorbent material in the cavity is also of modest benefit. In Table 7-8 four floor and ceiling structures are described along with STC ratings for each, as determined from field TL measurements.

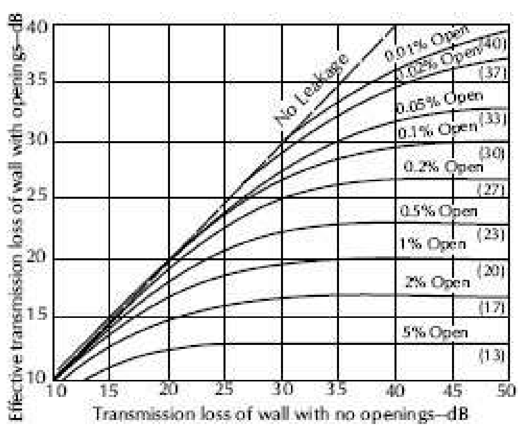


Figure 7-17. Effect of gaps on transmission loss. Courtesy of Russ Berger, Russ Berger Designs.

Another means of decoupling the floor above from the sound room ceiling involves suspending the entire ceiling by a resilient suspension, such as in Fig. 7-19. Mason Industries, Inc. reports one test that demonstrates the efficacy of this approach.[12] They started with a 3in concrete floor that alone gave STC-41. With a 12in air gap, a ⅝ in gypsum board ceiling was supported on W30N spring and neoprene hangers, resulting in STC-50. By adding a second layer of ¾ in gypsum board and a sound-absorbent material in the air space, an estimated STC-55 was realized. The W30N hanger uses both a spring and neoprene. This combination is effective over a wide frequency range. The spring is effective at low frequencies and the neoprene at higher frequencies.

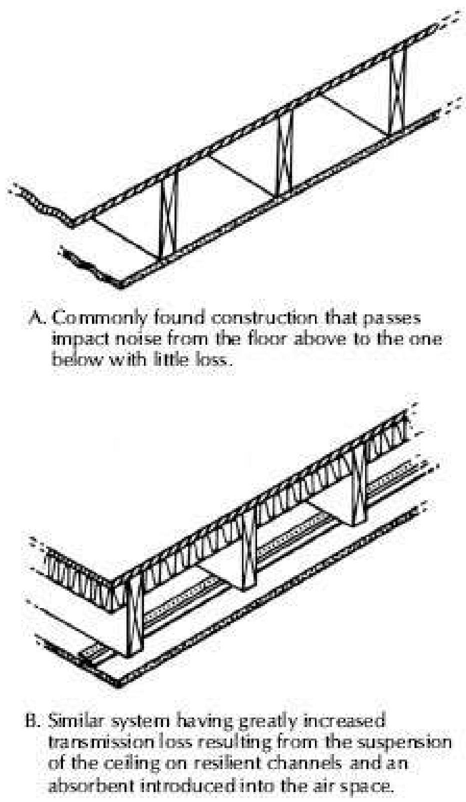


Figure 7-18. Floor and ceiling systems.

Table 7-8. Floor and Ceiling Systems

Ceiling Treatment	Treatment of Floor Above	Sound Transmission Class*

½ inch gypsum wallboard nailed to joists	1½inch lightweight concrete on ⅝inch plywood, 2in × 12in joists 16in on centers	STC-48
3 inch mineral wool. Resilient channels 2ft-0in on center ½ inch sound deadening board, ⅝ inch gypsum board	1⅛ inch plywood on 2in × 10in joists 16in oc	STC-46
3 inch mineral wool. Resilient Channels 2ft-0in, oc ⅝ inch gypsum board	1½ inch lightweight concrete on ½ inch sound deadening board on ⅝ inch plywood, 2in × 10in joists, 16in oc	STC-57
2 inch mineral wool. ½ inch sound deadening board. Resilient channels 2ft-0in oc, ⅝ inch gypsum board	1½ inch lightweight concrete on ⅝ inch plywood, 2in × 10in joists, 16in oc	STC-57

* These are FSTC ratings.[12]

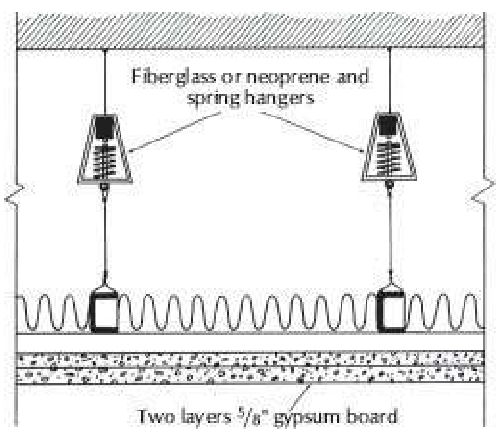


Figure 7-19. A method of suspending a ceiling that gives a great improvement in STC rating of the floor and ceiling combination.

7.4.7 Floor Construction

Many variables must be considered when designing isolated floors. These variables include cost, load limits of the existing structure, the desired isolation, and the spectrum of the noise. Every successful system uses a combination of mass and resilient support designed to work above the resonance point of the system, and thus achieve isolation. There are three general approaches to floating or isolated floors; the continuous underlayment, the resilient mount, and the raised slab, Fig. 7-20.

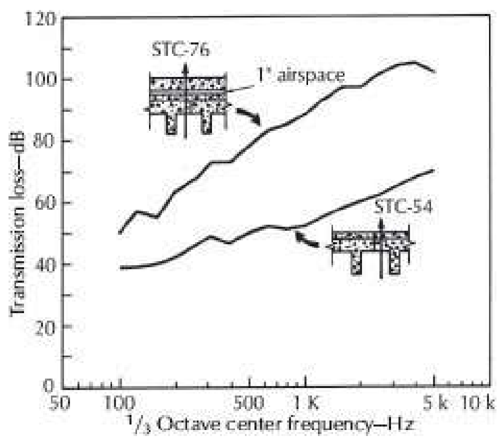


Figure 7-20. The dramatic improvement of sound transmission class (STC) from 54 to 76 by adding a 4 in floating floor with a 1in air gap between it and the structural floor. (Riverbank TL-71-247 test reported by Mason Industries, Inc., in Reference 13.)

Floating Floors. Once again, simply increasing mass is often the least productive way to make significant gains in STC. For example, a 6in solid concrete floor has an STC of 54, and doubling the thickness to 12in raises it only to STC-59. There are many recording studios and other sound-sensitive rooms that require floors greater than STC-54. The answer is in dividing available mass and placing an air space between. The results of an actual test, sponsored by Mason Industries, Inc., are given in Fig. 7-21.[13] The TL of basic T sections (4in floor thickness) with 2 in of poured concrete gives a total thickness of 6in and the STC-54 mentioned previously. Adding a 4in concrete floor on top of the same structural floor with 1in of air gap gives a healthy STC-76, which should be adequate for all but the most critical applications. A 4in slab added to the 6in floor without an air space gives only STC-57. A 19 dB improvement can be attributed directly to the air space.

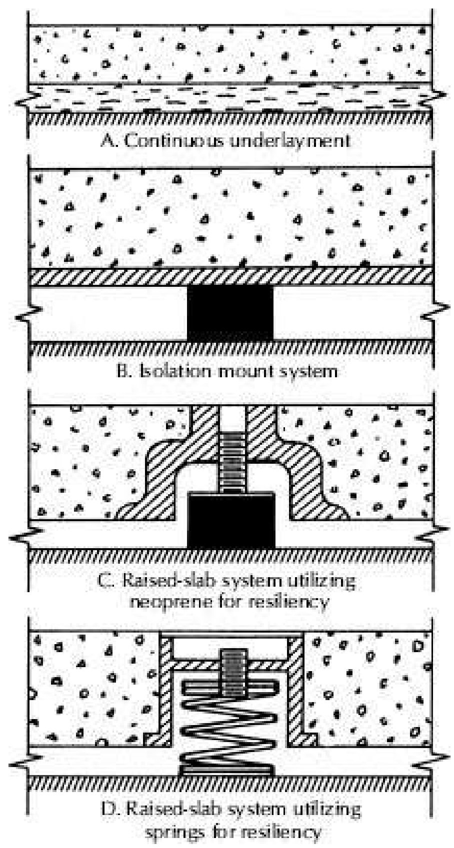


Figure 7-21. Four methods used in floating floors for increasing transmission loss.

Continuous Underlayment. The continuous underlayment is the simplest and easiest form of floating floor to construct. It is most often used for residential and light commercial applications where surface loads are relatively light. The technique consists of laying down some sort of vibration-absorbing mat and then constructing a floor on top of the mat, taking care not to penetrate the mat with any fasteners. The perimeter is surrounded with a perimeter isolation product and sealed with a nonhardening acoustical sealant. Maxxon offers a number of products including Acouti-Mat 3, Acousti-MatII-Green, and Enkasonic®. These are all underlayments that form a resilient layer upon which a wood floor can be constructed, Fig. 7-22, or can be part of a poured concrete system.

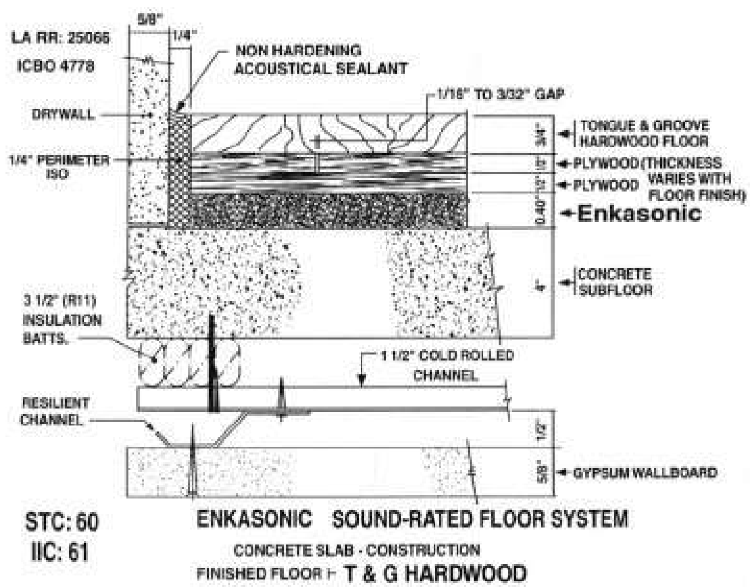


Figure 7-22. Enkasonic floor system.

Isolation Mount Systems. If heavier loads are anticipated and greater isolation is needed, an isolation mount system should be considered. Various manufacturers build systems for isolating either wood floors or concrete slabs. Wood floors can be isolated as shown in Fig. 7-23. This system offered by Kinetics utilizes encapsulated fiberglass pads, imbedded in a roll of low-frequency fiberglass designed to fill the air space.

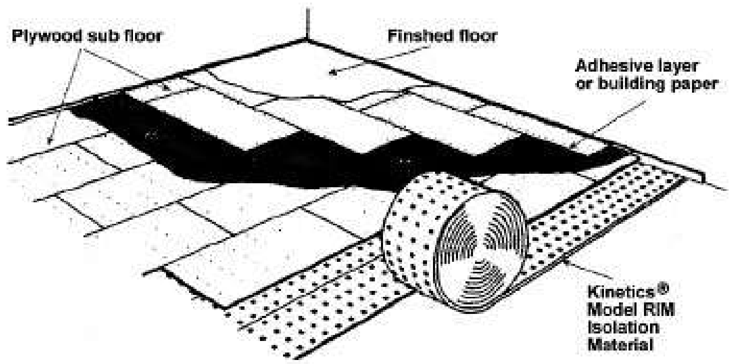


Figure 7-23. Kinetics Floating Wood Floor. Courtesy Kinetics Corp.

Another approach by Mason Industries is to build a grid supported on neoprene mounts or, if greater isolation is needed, on combination spring and neoprene as shown in Fig. 7-24. A wood floor is then built on the substructure.

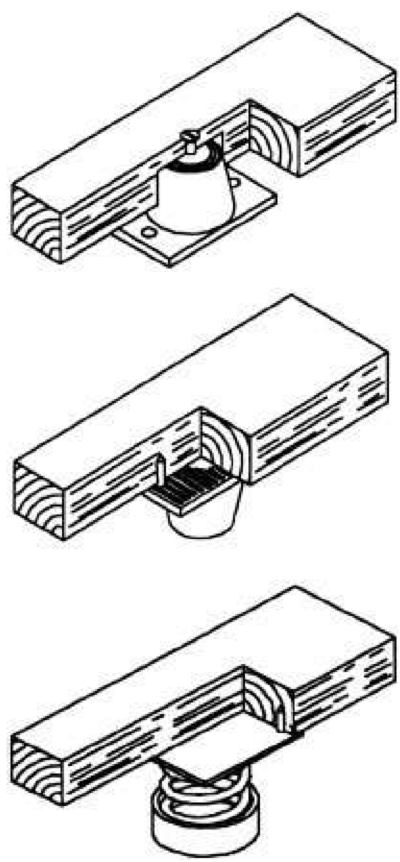


Figure 7-24. Mason Industries Floating Wood Floor systems using either springs or neoprene. Courtesy Mason Industries

In some situations a floating concrete slab is indicated in Fig. 7-25, concrete slab is supported by the model RIM mat. The roll-out mat is ordered with the pad spacing based on the expected load. When the mat is unrolled:

1. The plywood panels are then put in place
2. The plastic sheet laid over the plywood.
3. The concrete poured, Fig. 7-25.

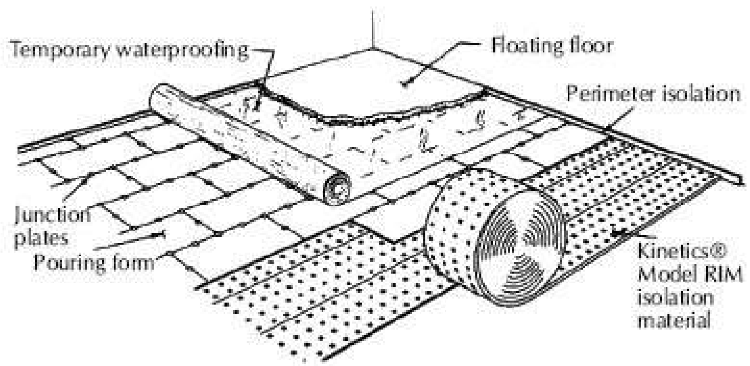


Figure 7-25. The roll-out mat system of constructing floating floors. Courtesy Kinetics Inc.

A perimeter board isolates the floating floor from the walls. The plastic film protects the plywood and helps to avoid *bridges*.

Raised-Slab or Jack-Up System. This system is for heavy duty applications where high STC ratings are needed. In Fig. 7-26 the individual isolators are housed in metal canisters, Fig. 7-27, that are placed typically on 36in to 48in centers each dimension. The metal canisters are arranged to tie into the steel reinforcing grid and are cast directly in the concrete slab. After sufficient curing time (about

28 days), it is lifted by judicious turning of all the screws one-quarter or one-half turn at a time. This is continued until an air space of at least 1in is achieved. Fig. 7-28 shows an alternative raised slab system utilizing springs instead of neoprene or fiberglass mounts. After the slab is raised to the desired height, the screw holes are filled with grout and smoothed. Fig. 7-29 further describes the elements of the raised-slab system. Turning the screws in the load-bearing isolation mounts raises the cured slab, producing an air space of the required height. This system requires heavier reinforcement rods in the concrete than the system of Fig. 7-25.

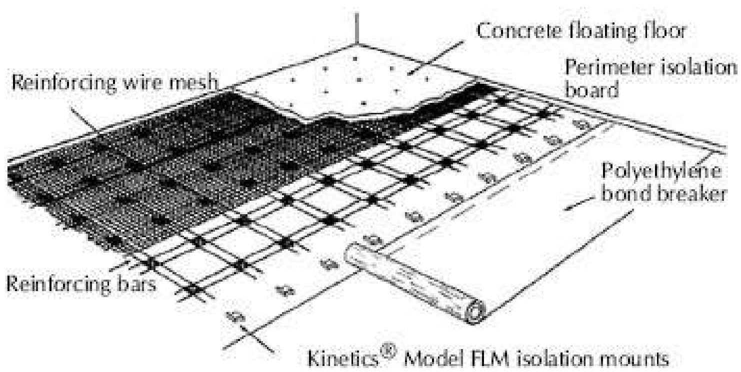


Figure 7-26. Kinetics FLM Jack Up Concrete Floor system. Courtesy Kinetics Corp.

7.4.8 Summary of Floating Floor Systems

Loading must be calculated for each type of floating floor systems discussed. If the resilient system is too stiff, vibration will travel through the isolator rendering it ineffective. Likewise if the springs are too soft, they will collapse under the weight of the structure and also be ineffective.

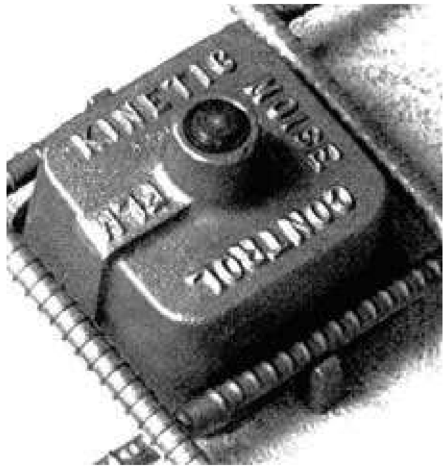


Figure 7-27. Kinetics FLM isolation mount. Courtesy Kinetics Corp.

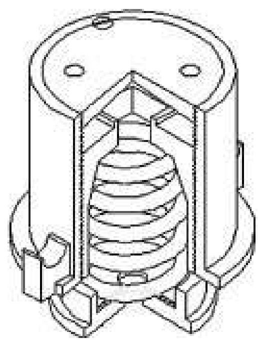


Figure 7-28. Mason Industries FS spring jack up floor system. Courtesy Mason Industries.

Each floating floor system has its advocates. No one type of floor will suit all situations. The designer is urged to consider all the variables before making a decision. For example, there are pros and cons concerning use of neoprene versus the compressed, bonded, and encased units of glass fiber. Most of the arguments have to do with deterioration of isolating ability with age and freedom from oxidation, moisture penetration, and so on.

Fig. 7-30 combines several features that have been discussed in a "room within a room." The walls are supported on the floating floor and stabilized with sway braces properly isolated. The ceiling is supported from the structure with isolation hangers. This type of hanger incorporates both a spring, which is particularly good for isolation from low-frequency vibration, and a Neoprene or a fiberglass element in series, which provides good isolation from higher-frequency components. An important factor is the application of a non-hardening type of acoustical sealant at the points marked “S.” An even better approach would be to support the ceiling from the walls by using joists or trusses spanning the room. Such a room should provide adequate protection from structure-borne vibrations originating within the building as well as from those vibrations transmitted through the ground to the building from nearby truck, surface railroad, or subway sources.

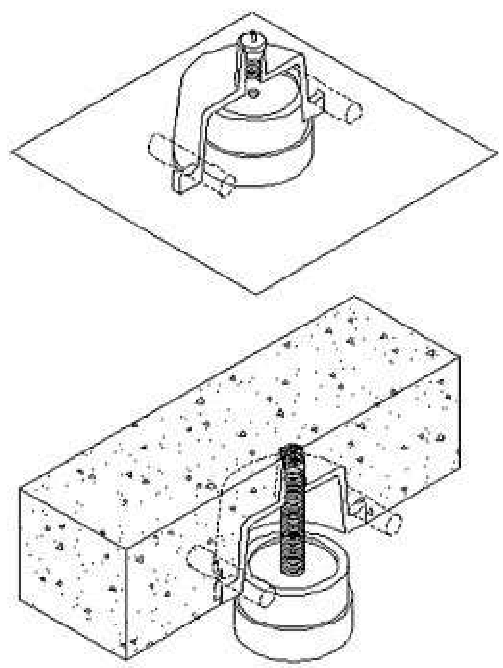


Figure 7-29. Details of a jack mount. Courtesy Mason Industries.

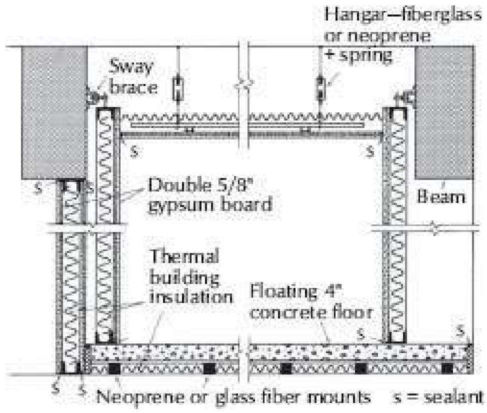


Figure 7-30. A "room-within-a-room" exemplifying the principles discussed in the text.

The design of rooms to achieve maximum isolation from airborne and structure-borne sounds is a highly specialized undertaking, ordinarily entrusted to consultants expert in that branch of acoustics. However, a sound engineer, charged with the responsibility of working with a consultant or doing the design personally, is advised to become familiar with the sometimes conflicting claims of suppliers and the literature on the subject.

Acoustical Doors. Every part of an acoustical door is critical to its performance. Special metal acoustical doors are available with special cores, heavy hinges, including special sealing and latching hardware. Their acoustical performance is excellent and their higher cost must be evaluated against high labor costs in constructing an alternative. There are two design elements required in considering what kind of door to utilize. There is the transmission loss of the door itself and there is the sealing system. The sealing system is the more critical of the two. Whatever system is used, it must hold up over time and withstand the wear and tear of use. Doors and their seals are difficult to build and are often the weak point of a sound room. There is good reason to design sound room access and egress in such a way that excessively high performance is not required of a single door. Use of a sound lock corridor principle places two widely spaced doors in series, relieving the acoustical requirements of each, Fig. 7-31.

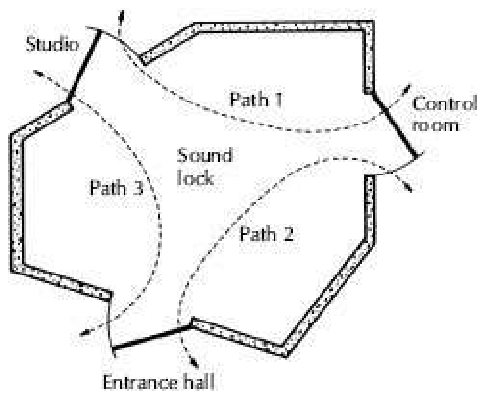


Figure 7-31. Sound lock corridor.

Homemade Acoustical Doors. An inexpensive door, satisfactory for less demanding applications, can be built from void-free plywood or high density particle board. It is also possible to start with a core material of particle board and laminate it with gypsum board if sufficient care is taken to protect the fragile edges of the gypsum board. Doors for acoustical isolation must have a solid and void-free core and be as massive as practical. Most residential grade doors are hollow and approach acoustical transparency. Some commercially available solid core doors are made of laminated wood; others, of particle board with composition board facing. The latter has the greater surface density. The 5.2lb/ft^2 of the particle-board type gives an STC value of about 35. An STC-35 does not do justice to, say, STC-55 walls. Nevertheless, for doors separated as they are in the case of a sound lock, the TL of one door comes close to adding arithmetically to the loss of the other door. Two doors, well separated, approach doubling the effect of one.

All this implies a perfect seal around the periphery of the door attained only by nailing the door shut and applying a generous bead of acoustical sealant on the crack. A practical operative door must utilize some form of weatherstripping or other means for its seal. Fig. 7-32 illustrates different approaches to sealing a door.[13] Many of these, especially the wiping type, require constant maintenance and frequent replacement. One of the more satisfactory types is the magnetic seal, similar to those on most household refrigerator doors. Zero International manufactures a system of door seals specifically designed for acoustical applications, Fig. 7-33. This type of commercially available acoustical door seal is a good way to get results from a homemade door that approaches the performance of a proprietary door at a fraction of the cost.

Proprietary Acoustical Doors. By far the more satisfactory doors for acoustical isolation in sound rooms are those manufactured especially for the purpose. Such doors offer measured and guaranteed performance over the life of the door with only occasional adjustment of seals. This is in stark contrast to the need for constant seal maintenance in the homemade door shown in Fig. 7-32. Each manufacturer has its own strengths. Some doors like the Overly and the IAC use cam lift hinges, which actually lift the door as it opens.

Manufacturers of building elements that need to be rated for sound transmission use ASTM standards in measuring their products. ASTM e-90 is the appropriate standard for sound transmission measurements. Copies of the standards are available at www.ASTM.com. Most manufacturers build a range of doors to suit specific needs. IAC builds doors ranging from an STC-43 to an impressive STC-64, Fig. 7-34.

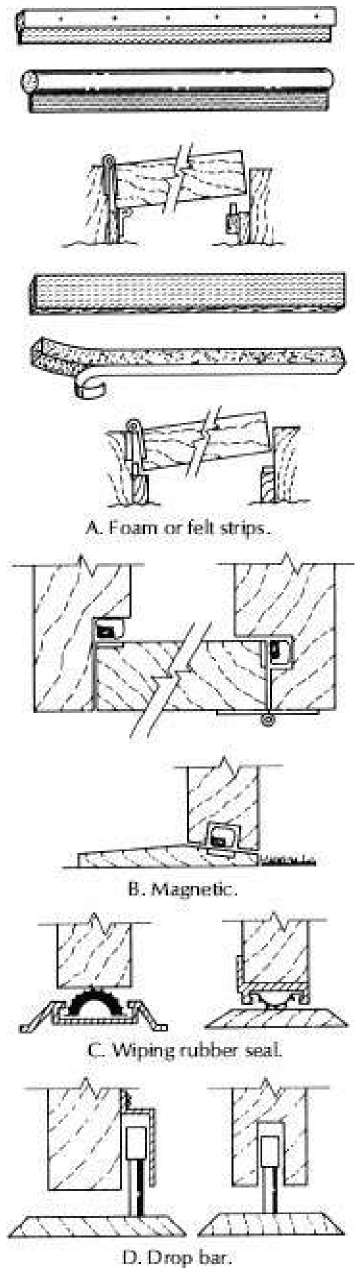


Figure 7-32. Numerous types of weather stripping can be used for sealing doors to audio rooms. Courtesy Tab Books, Inc.

Windows. Occasionally sound rooms require windows. The observation window between control room and studio is an example. It can very easily have a weakening effect on the overall TL of the partition between the two rooms. (See section 7.3.11.) A wall with a rating of STC-60 alone might very well be reduced to STC-50 with even one of the more carefully designed and built windows installed. Just how much the window degrades the overall TL depends on the original loss of the partition, the TL of the window alone, the relative areas of the two, and of course, the care with which the window is installed. To understand the factors going into the design of an effective observation window, a good place to start is to study the effectiveness of glass as a barrier.[14]

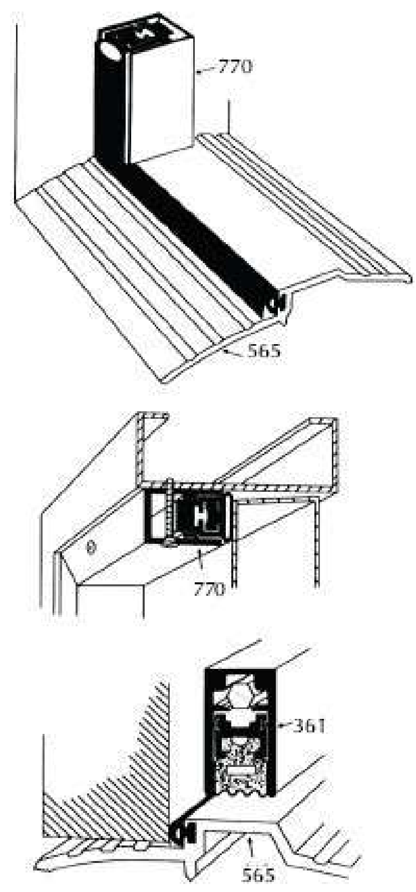


Figure 7-33. Sealing systems from Zero Mfg.

Transmission Loss of Single Glass Plates. The measured transmission loss of ¼in, ½in, and ¾in single-glass plates (or float) 52 in × 76in is shown in Fig. 7-35. As expected, the thicker the glass plate, the higher the general TL except for a coincidence dip in each graph. Although the heavy ¾in plate attains a TL of 40dB or more above 2kHz. It is inappropriate for use in an STC-50 or STC-55 wall. Considering this general lack of sufficient TL and the complication of the coincidence dip, the single-glass approach is insufficient for most observation window needs. Laminated glass is more of a *limp mass* than glass plate of the same thickness and, hence, has certain acoustical advantages in observation windows. The characteristics of ¼in, ½in, and ¾in laminated single-glass plates are shown in Fig. 7-36.

Transmission Loss of Spaced Glass Plates. Fig. 7-37 shows the effect of three different spacings. In all cases the same ½in and ¼in glass plates are used, but the air space is varied from 2in to 6in. The effect of spacing the glass plates is greatest below 1500Hz. There is practically no increase in transmission loss by spacing the two glass plates above 1500Hz. In general, the 2 in increase from 2in to 4in is less effective than the same 2in spacing increase from 4in to 6in. Many observation windows in recording studios utilize spacings of 12in or more to maximize the spacing effect.

When two glass plates are separated only a small amount, such as glass widely used for heat insulation, the sound TL is essentially the same as the glass alone from which it is fabricated. There is practically no acoustical advantage using this type of glass in observation windows. This is one of the few cases where thermal insulation does not correspond to acoustic isolation. The single case of using laminated glass for one of the plates with 6 in separation is included in Fig. 7-37. The superior performance of laminated glass comes with a higher cost.

Managing Cavity Resonance. The TL measurements in Fig. 7-37 were made with no absorbing material around the periphery of the space between the two glass plates. By lining this periphery with absorbent material, the natural cavity resonance of the space is reduced. An average 5 dB increase in TL can be achieved by installing a minimum of 1in absorbent on these *reveals*. The use of 4 in of absorbing material, covered with, perhaps, perforated metal, further improves low-frequency transmission loss.

The practice of using glass plates of different thickness is substantiated by shallower coincidence dips in Fig. 7-37 as compared to Fig. 7-35. Resonance associated with the plates or the cavity tend toward the creation of *acoustical holes*, or the reduction of TL at the resonance frequencies. Hence, distributing these resonance frequencies by the staggering of plate thickness and use of laminated glass is important.

Homemade Acoustical Windows. The essential constructional features of two types of observation windows are shown in Fig. 7-38. Fig. 7-38A is typical of the high TL type commensurate with walls designed for high loss. The high TL of the window is achieved by using heavy laminated glass, maximum practical spacing of the glass plates, absorbent reveals between the glass plates, and other important details such as a generous application of acoustical sealant. It is very important to note that the windowsill and other elements of the frame do not bridge the gap between the two walls and thereby compromise the double wall construction. Bridging the double wall construction at the window is a very common error that must be avoided if the STC of the partition is to be maintained

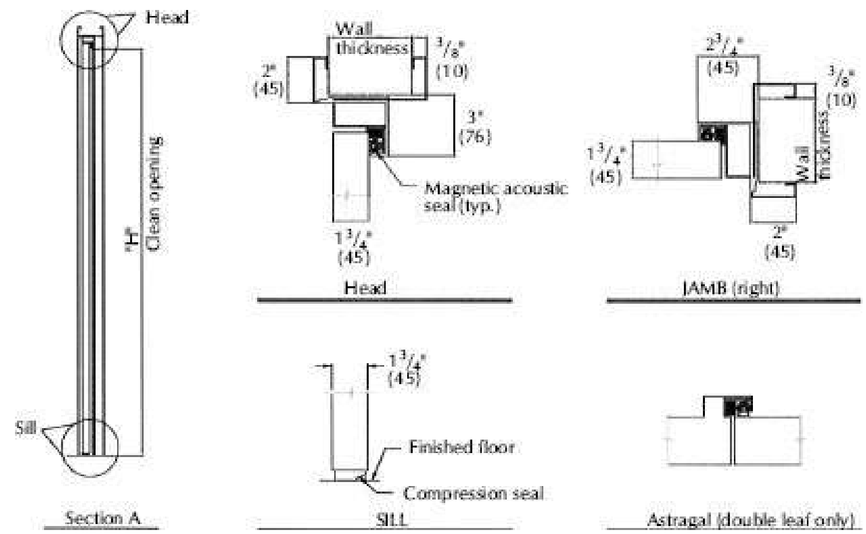


Figure 7-34. IAC STC 43 Door. Courtesy IAC.

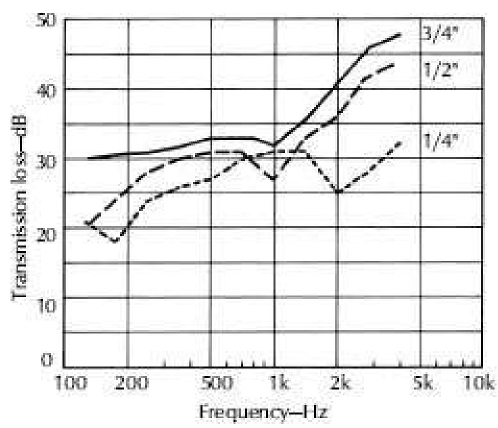


Figure 7-35. Sound TL characteristics of single glass (plate or float) panels. Courtesy Libbey-Owens-Ford Co. (After Reference 15.)

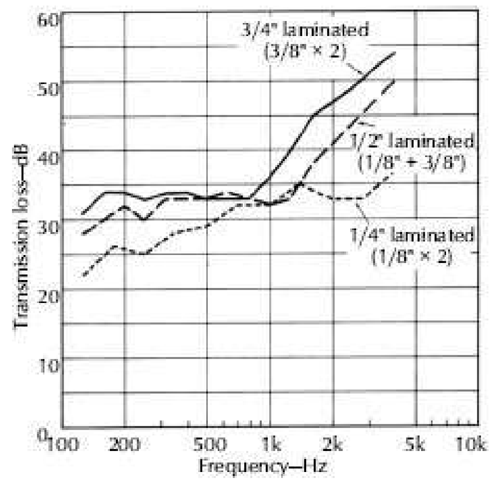


Figure 7-36. Sound TL characteristics of single panels of laminated glass. Courtesy Libbey-Owens-Ford Co. (After Reference 15.)

Fig. 7-38B shows a window for a single stud wall, a more modest TL. The same general demands are placed on this window as on the one in Fig. 7-38A, except that scaled down glass thickness and spacing are appropriate.

Inclining one of the plates, as shown in Fig. 7-38, has advantages and disadvantages. Slanting one pane reduces the average spacing, which slightly reduces the TL. However, slanting one window as shown especially in a studio (as distinct from a control room) will have the beneficial effect of preventing a discrete reflection right back at a performer standing in front of the window. The principal benefit of such plate inclination is really the control of *light* reflections that interfere with visual contact between the rooms.

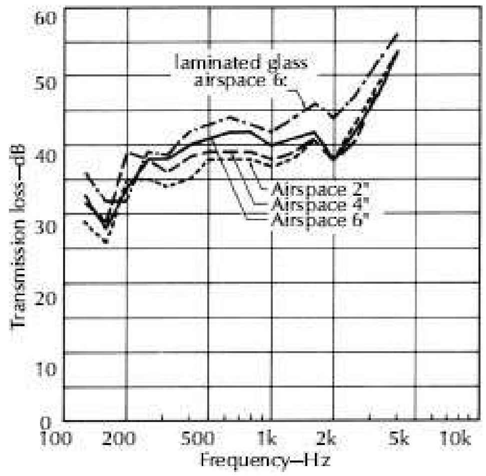


Figure 7-37. Spacing two dissimilar glass plates improves transmission loss. Glass of ½ and ¼ inch thickness used in all cases.

Proprietary Acoustical Windows. Many of the same companies that build proprietary acoustical doors also build acoustical windows. IAC builds a line of windows ranging from STC-35 to STC-58. The STC-53 window from IAC is shown in Fig. 7-39 and Fig. 7-40. It should be noted that the same warning about bridging a double wall construction applies to proprietary windows as well as to home-made ones.

7.4.9 Transmission Loss of a Compound Barrier

We are using the term *compound* to refer to those partitions that are not homogeneous, e.g., those partitions that include areas with differing TL ratings. For example, when an observation window having one TL is set in a wall having another TL, the overall TL is obviously something else, but what is it? It most certainly cannot be obtained by simple manipulation of TLs or STC values. The problem must be referred to as the basics of sound power transmission. Fig. 7-41 illustrates the case of a 4.4 ft × 6.4 ft window set in a 10ft × 15ft partition between control room and studio.

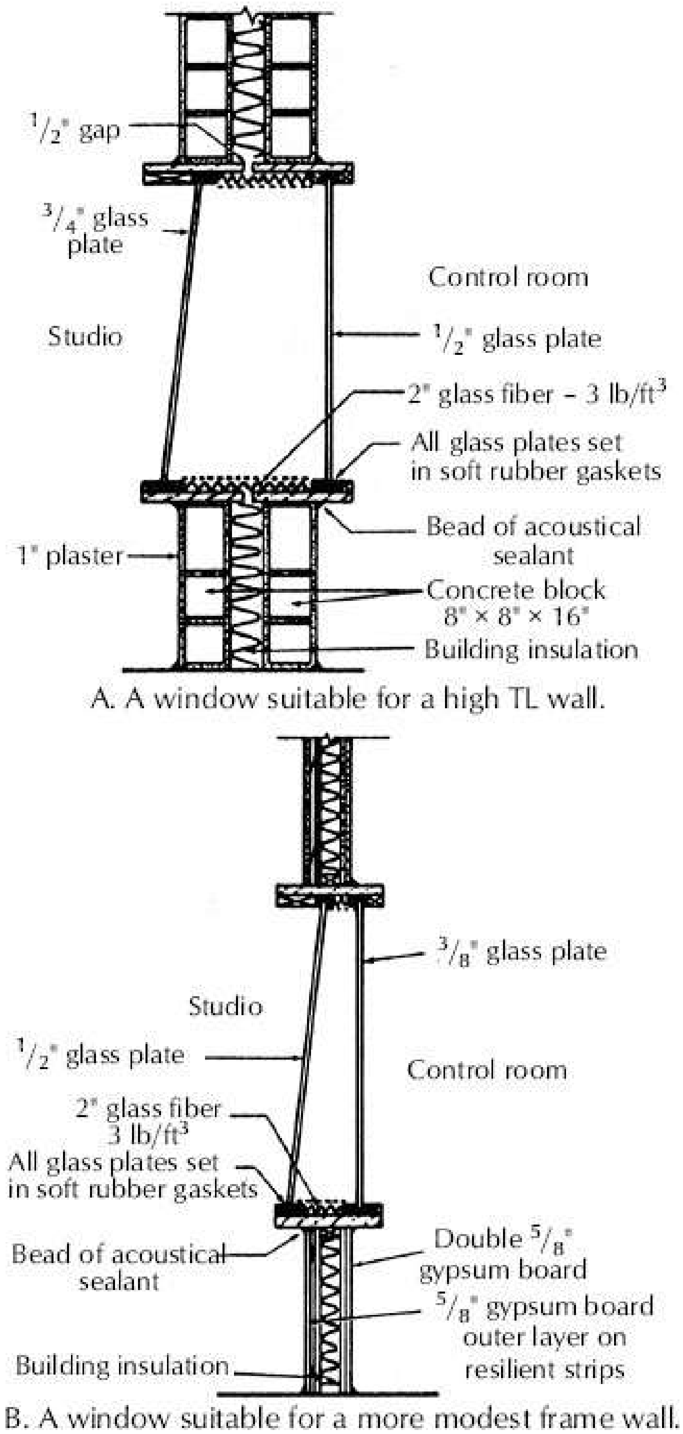


Figure 7-38. Construction details for practical observation windows set in a partition between control room and studio.

The way the transmission loss of the window and the wall affect each other is given by the expression:

$$TL = -10\log\left(\frac{S_1}{10^{\frac{TL_1}{10}}} + \frac{S_2}{10^{\frac{TL_2}{10}}}\right) \qquad (7\text{-}2)$$

where,

TL is the overall transmission loss,

S_1 is the fractional wall surface,

TL_1 is the wall transmission loss in dB,

S_2 is the fractional window surface,

TL_2 is the window transmission loss in dB.

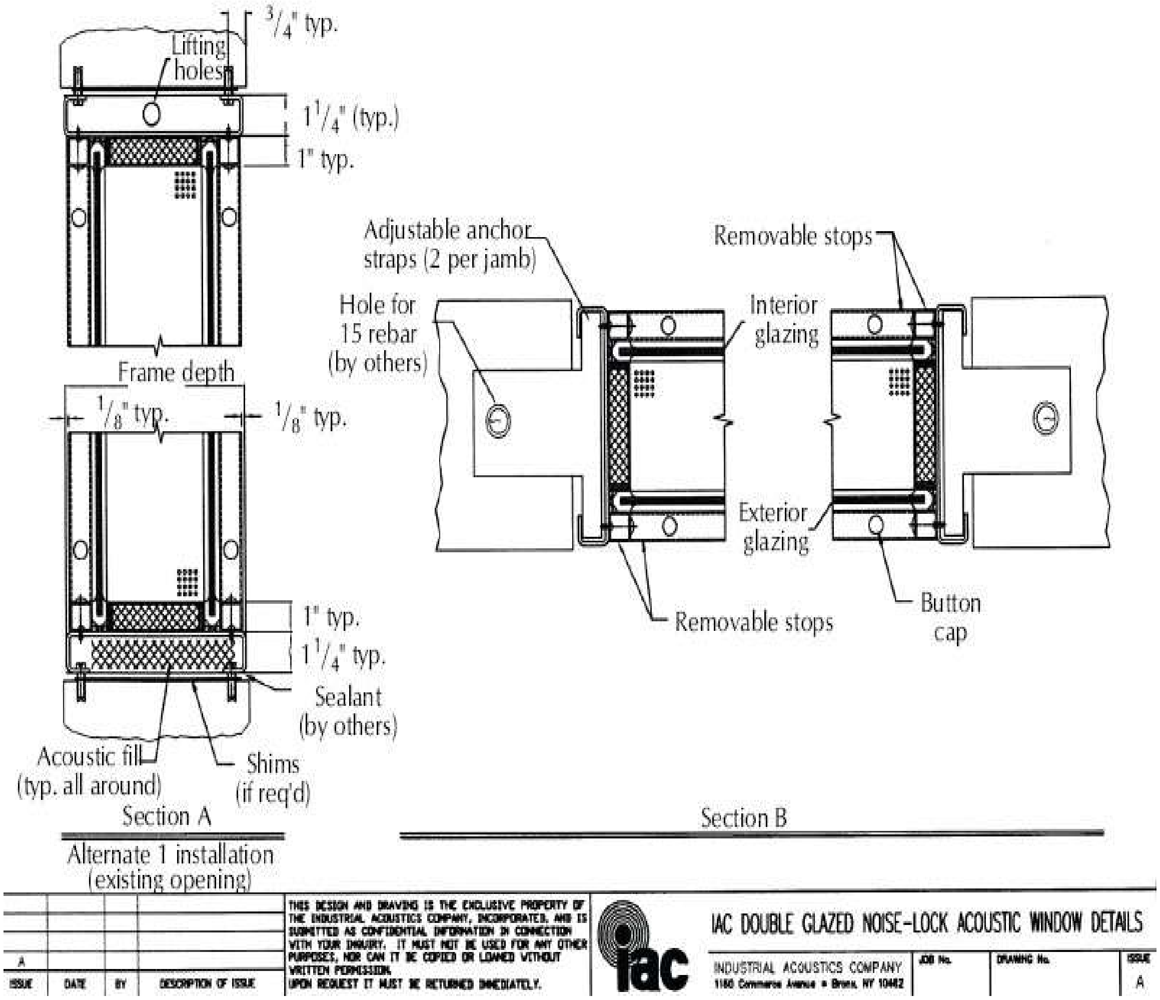


Figure 7-39. Noise-Lock™ window by IAC is rated at STC 53. Courtesy IAC.

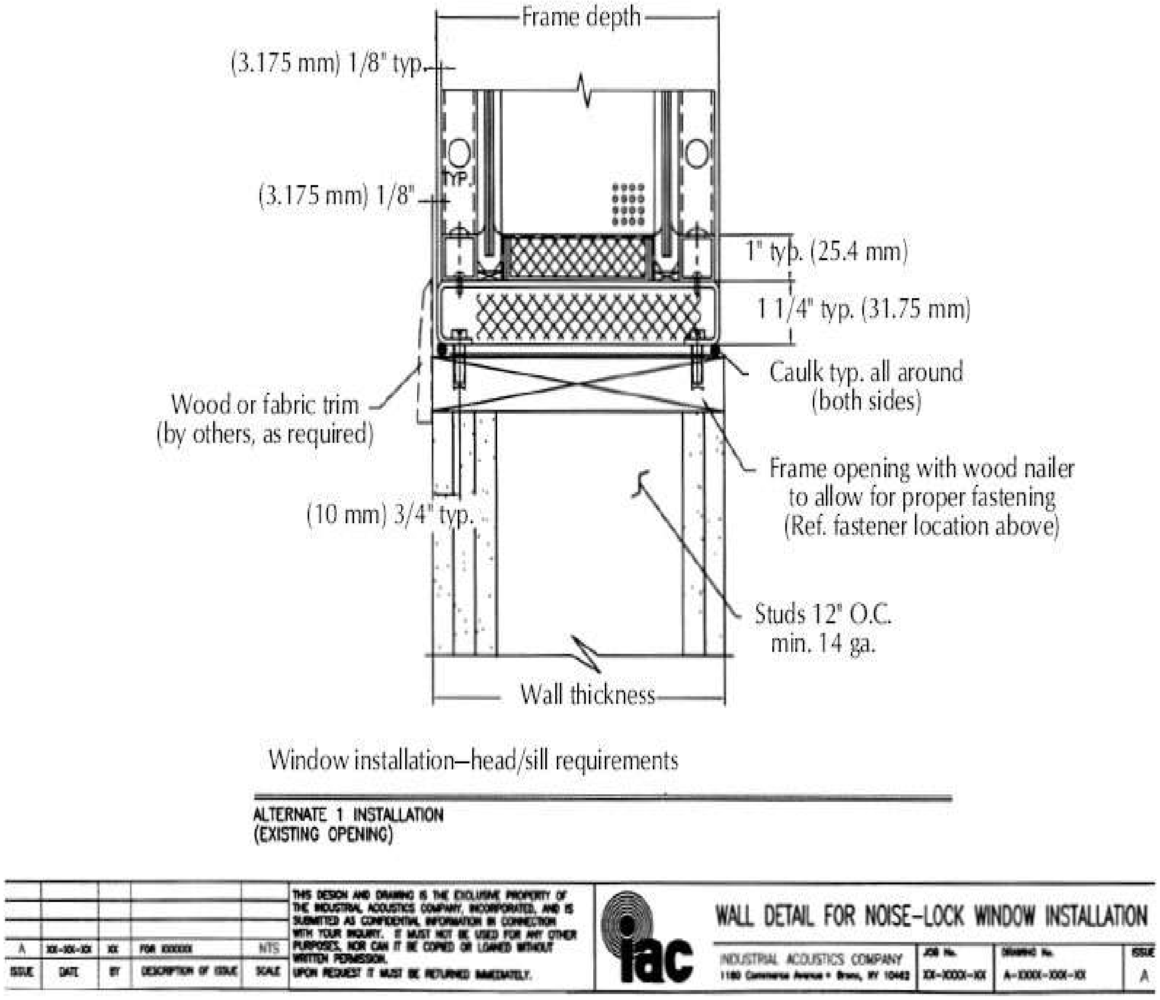


Figure 7-40. Head and sill requirements for IAC Noise-Lock™ windows. Courtesy IAC.

As an example let us say that for a given frequency the wall TL_1 = 50dB and the window TL_2 = 40dB. From Fig. 7-41 we see that S_1 = 0.812 and S_2 = 0.188. The overall TL is

$$TL = -10\log\left(\frac{S_1}{10^{\frac{TL_1}{10}}} + \frac{S_2}{10^{\frac{TL_2}{10}}}\right)$$

$$= 45.7\,\mathrm{dB}$$

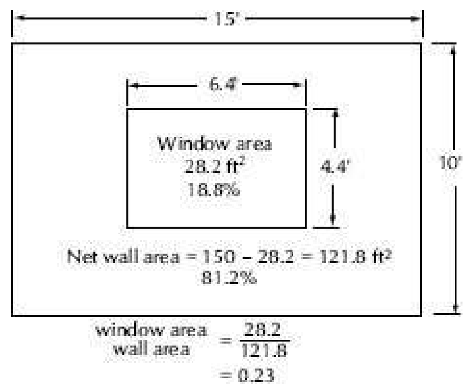


Figure 7-41. Typical observation window in a wall between control room and studio.

The 40dB window has reduced the 50dB wall to a 45.7dB overall effectiveness as a barrier. This is for a given frequency. Fig. 7-42 solves Eq. 7-2 in a graphical form using the following steps:

1. Figure the ratio of glass area to total wall area, and find the number on the X axis.
2. Subtract window TL from wall TL, and find the intersection of this value with the area ratio on the X axis.
3. From the intersection, find the reduction of the wall TL from the left scale.
4. Subtract this figure from the original wall TL.

Using the graph of Fig. 7-42, find the effect of the window on the compound wall. The ratio of the window area to the wall area is 0.23. Locate 0.23 along the bottom axis. The difference in TL between the two is 10dB. Find the intersection between the 10dB line and the ratio of the areas. A reduction of slightly less than 5dB is read off the left scale. Subtracting 5dB from the 50dB wall TL gives the overall TL with a window of 45dB. (Calculated from Eq. 7-2 gives 45.7dB.)

It is usually easier and more economical to get high TL in wall construction than in window construction. The possibility arises of compensating for a deficient window by over designing the wall. For example, recognizing that an STC-70 masonry wall is possible, how far will it lift an STC-45 window? Using Eq. 7-2 again, we find the overall STC to be 52.2 dB, an increase of over 7dB over the STC-45 window. Actually, using Eq. 7-2 with STC values is a gross oversimplification embracing all the inaccuracies of fitting measured TL values with a single-number STC rating. Making the calculations from measured values of TL at each frequency point is much preferred. Of course, all this assumes an airtight seal has been achieved.

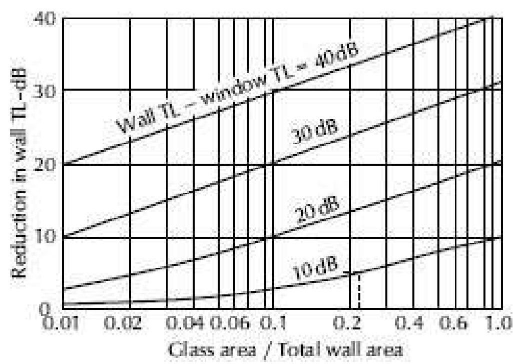


Figure 7-42. Graphical determination of the effect on the overall transmission loss (TL) of a wall by an observation window.

Everything that bridges the isolation system is a potential *short circuit* for noise. Such bridges include, HVAC ducts, electrical conduit, sprinkler systems, plumbing, raceways, and the like.

Now we have the formula for empirically looking at the effect of a crack in the wall (Fig. 7-17 was plotted using Eq. 7-2). Let us assume that an observation window and wall combination have a calculated composite TL of 50dB. The window, installed with less than ideal craftsmanship, developed a ⅛ in (0.125in) crack around the window frame as the mortar dried and pulled from the frame. Since this is the window of Fig. 7-41, the length of the crack is 21.6ft, giving a crack area of 0.225ft^2. What effect will this crack have on the otherwise 50dB wall? Substituting into Eq. 7-4, we find the new TL of the wall with the crack to be 28dB. This is similar to leaving off an entire pane of glass or a layer of gypsum board. If the crack were only $\frac{1}{16}$ in wide, the TL of the wall would be reduced from 50dB to 31.2dB. A crack only 0.001in wide would reduce the TL of 50dB to 40.3dB. Let the builder beware!

7.4.10 Isolation Systems Summary

Noise migrates from one area to another in two ways. It travels through the air and it travels through the structure. To reduce or eliminate airborne noise, one must eliminate all air paths between the spaces. To reduce structure-borne noise one must create isolation systems that eliminate mechanical connections between spaces. It is a rather simple matter to make these statements. Implementing the solutions is obviously much more difficult. The following points should be kept in mind:

- Make seams airtight.
- Analyze all possible flanking paths that noise will take and realize that *all* must be controlled if significant isolation is desired.
- A room built entirely on a floating slab with the ceiling supported entirely by the walls will always be superior to any other method.

7.5 Heating, Ventilating, and Air Conditioning (HVAC) Systems for Low Noise

So far in this chapter we have considered systems that keep unwanted sound out. When we consider HVAC systems we are dealing with systems that:

- Breach the acoustical shell designed to keep noise out.
- Introduce considerable noise of their own.
- Provide a pathway for sound (noise) to easily migrate from one space to another.

HVAC systems can sometimes undermine all the efforts of isolation. Often the cheapest solution to providing HVAC to sound sensitive spaces is to use window units that get shut off when quiet is needed! If this solution is not acceptable, and central distributed systems must be used, the designer must understand that success will require significant expense and engineering. The design of HVAC systems is best left to professional mechanical engineers. No better preparation for this responsibility can be obtained than from carefully studying the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) publications.[16,17,18]

It is important to understand that HVAC systems found in most residences or even in light commercial or office spaces are totally inadequate for use in noise critical spaces. Unlike residential systems that often use high efficiency systems that deliver low volumes of cold air at high velocities, low noise systems require high volume, low velocity delivery. Many commercial systems utilize supply ducts and the return relies on leakage under doors or common ceiling plenums. In order to achieve low noise, both the supply and return must be individually ducted to each room.

7.5.1 Location of HVAC Equipment

From the standpoint of sound room noise, the best location for the

HVAC equipment is in the next county. Short of this, a spot should be selected that isolates the inevitable vibration of such equipment from the sound-sensitive area. A good situation is to have the equipment mounted on a concrete pad completely isolated from the structure. In this way, the noise problem is reduced to handling the noise coming through the ducts, a much simpler task than fighting structure-borne vibration.

7.5.2 Identification of HVAC Noise Producers

The various types and paths of HVAC noise producers are identified in Fig. 7-43. This figure provides an interesting study in flanking paths. It is important to remember that there will be relatively little noise reduction unless all of the paths are controlled. A represents the sound room. B represents the room containing the HVAC system. Looking at the noise sources as numbered, 1 and 2 represent the noise produced by the diffusors themselves. The noise is produced by the air turbulence that is created as the air moves through the diffusor. Many diffusors have a noise rating at a given air flow, and the only element of control in this case is selecting the design with the best rating. Don't forget that this applies to the return grille as well as the supply diffusor. Arrows 3 and 4 represent essentially fan noise, which travels to the room via both supply and return ducts and is quite capable of traveling upstream or downstream. The delivery of fan noise over these two paths can be reduced by silencers and/or duct linings. Sizing the ductwork properly is also a means of combating fan noise since sound power output of a fan is fixed largely by air volume and pressure. Arrow 5 represents a good example of a flanking path that is often missed. Depending on how the ceiling in both of the rooms is constructed, the sound from the HVAC unit can travel up through the ceiling in the HVAC room and comes down into room A. Of course the way to control path 5 is to make sure that the ceilings in both rooms are well built, massive enough to control low frequency vibrations, and of course, airtight. Arrow 6 represents that path where the sound can travel through gaps or holes inadvertently left in the partition. This has already been discussed in section 7.3.11 and in Fig. 7-17. Number 7 represents the sound that can travel straight through a poorly built wall. Numbers 8, 9, and 10 represents the paths that the structure-borne vibrations can take through the structure. We will deal with isolation issues in the next section. Finally, 11 and 12 represent what is called *break-in* noise. This is what happens when sound enters or breaks into a duct and travels down it, radiating into the room.

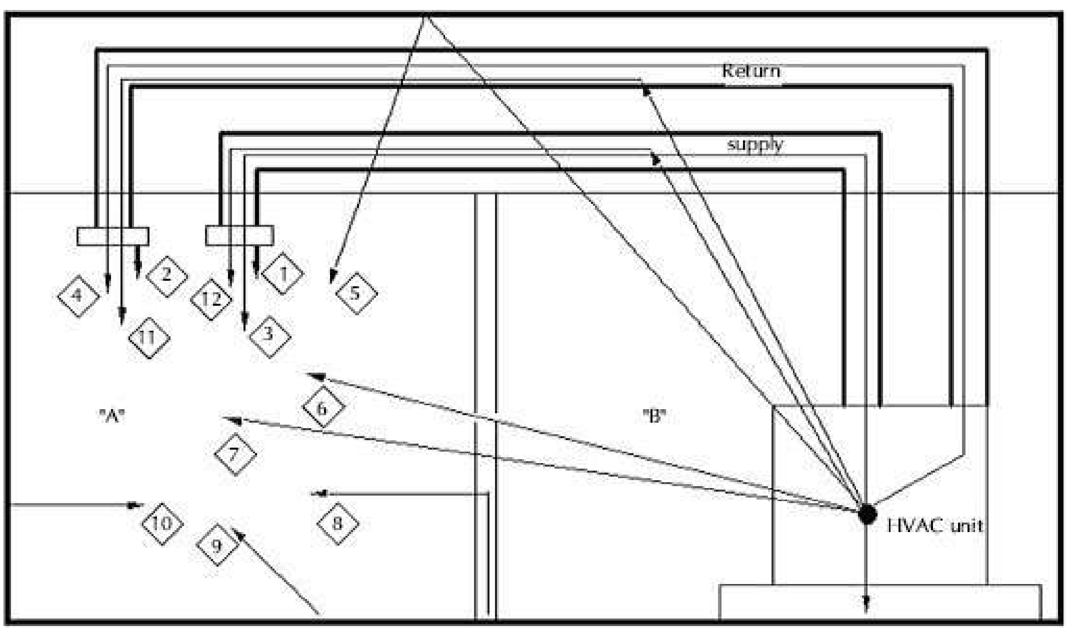


Figure 7-43. Typical paths by which HVAC noise can reach sound-sensitive rooms.

7.5.3 Vibration Isolation

The general rule is first to do all that can reasonably be done at the source of vibration. The simple act of mounting an HVAC equipment unit on four vibration mounts may help reduce transmitted vibration, may be of no effect at all, or may actually amplify the vibrations, depending on the suitability of the mounts for the job. Of course, if it is successful it would drastically reduce or eliminate paths 8, 9, and 10 in Fig. 7-43. The isolation efficiency is purely a function of the relationship between the frequency of the disturbing source f_d to isolator natural frequency f_n, as shown by Fig. 7-44. If $fd = f_n$, a resonance condition exists, and maximum vibration is transmitted. Isolation begins to occur when f_d/f_n is equal to or greater than 2. Once in this isolation range, each time f_d/f_n is doubled, the vibration transmission decreases 4–6 dB. It is beyond the scope of this treatment to go further than to identify the heart of both the problem and the solution, leaving the rest to experts in the field.

7.5.4 Attenuation of Noise in Ducts

Metal ducts with no linings attenuate fan noise to a certain extent. As the duct branches, part of the fan noise energy is guided into each branch. Duct wall vibration absorbs some of the energy, and any discontinuity (such as a bend) reflects some energy back toward the source. A very large discontinuity, such as the outlet of the duct flush with the wall, reflects substantial energy back toward the source. This results in attenuation of noise entering the room, as shown in Fig. 7-45. Unlike many other systems in acoustics this is one attenuation that is greater at low frequencies than at the highs.

Lining a duct increases attenuation primarily in the higher audio frequency range. Fig. 7-46 shows measured duct attenuation with

1inch duct lining on all four sides. The dimensions shown are for the free area inside the duct. This wall effect attenuation is greatest for the smaller ducts. For midband frequencies, a 10ft length of ducting can account for 40dB or 50dB attenuation for ducts 12 in × 24 in or smaller. There is a trade-off, however, as decreasing the cross section of the duct increases the velocity of the air moving through it. Higher air velocities produce greater turbulence noise at the grille/diffusor. Great stress is commonly placed on attenuation contributed by right-angle bends that are lined with duct liner. Fig. 7-47 evaluates attenuation of sound in lined bends. Only lining on the sides is effective, which is the way the elbows of Fig. 7-47 are lined. Here again, attenuation is greater at higher audio frequencies. The indicated duct widths are clear measurements inside the lining. The lining thickness is 10% of the width of the duct and extends two duct widths ahead and two duct widths after the bend. It is apparent that the lining contributes much to attenuation of noise coming down the duct, but less so at lower frequencies. Here too, there is a trade-off. Every bend, lined or not, increases the turbulence and therefore the noise.

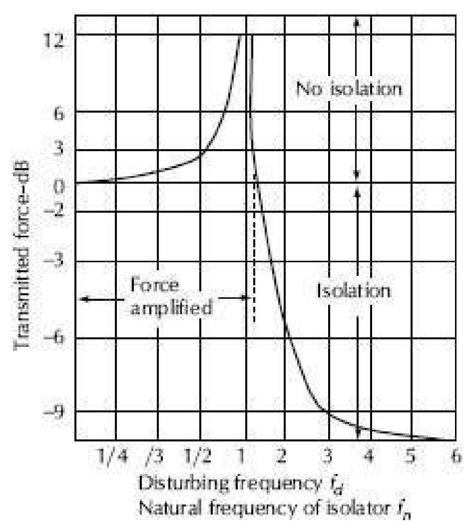


Figure 7-44. Noise from HVAC equipment may be reduced by isolation mounts, or it may actually be amplified. (After ASHRAE, Reference 18.)

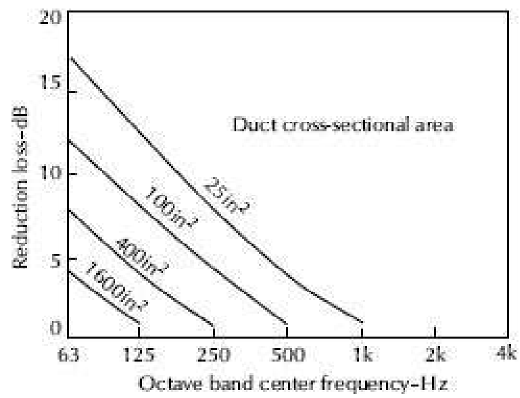


Figure 7-45. The effect of duct cross-sectional area on the attenuation of HVAC noise. (After ASHRAE, Reference 18.)

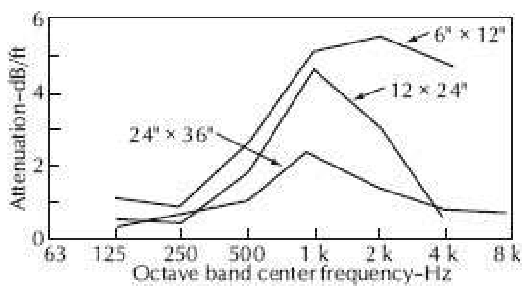


Figure 7-46. Measured noise attenuation in rectangular ducts. (After ASHRAE, Reference 18, which attributes Owens-Corning Fiberglas Corp. Lab Report 32433 and Kodaras Acoustical Laboratories Report KAL-1422-1 submitted to Thermal Insulation Manufacturer's Association.)

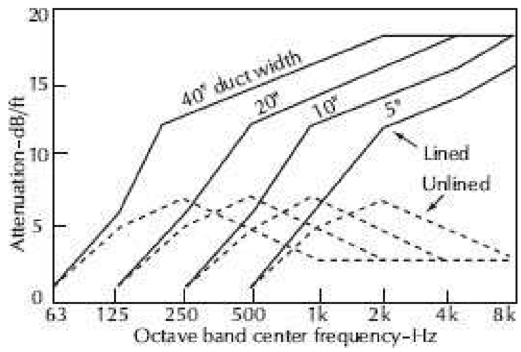


Figure 7-47. Noise attenuation in HVAC square-duct elbows without turning vanes. (After ASHRAE Reference 18.)

7.5.5 Tuned Stub Noise Attenuators

Fan blades can produce line spectra or tonal noise at a blade frequency of

$$\textit{Blade frequency} = \frac{RPM \times \textit{Number of blades}}{60\ \text{Hz}} \quad (7\text{-}3)$$

Usually this noise is kept to a minimum when the HVAC engineer selects the right fan. If such tones continue to be a problem, an effective treatment is to install a tuned stub filter someplace along the duct. These can be very effective in reducing fan tones. A typical stub and its attenuation characteristic are shown in Fig. 7-48A. The comparable characteristic of a reactive muffler is also shown in Fig. 7-48B.

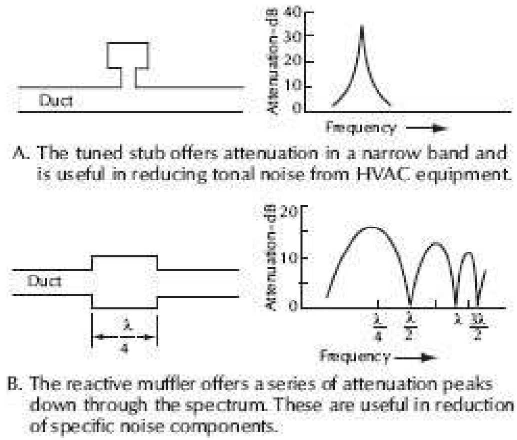


Figure 7-48. The tuned stub and reactive muffler used to attenuate tonal components of noise.

7.5.6 Plenum Noise Attenuators

As previously stated, a most effective procedure in noise reduction is to reduce the noise at, or very close to, the source. If a system produces a noise level that is too high at the sound room end, one possibility is to install a plenum in the supply and another in the return line. Such a plenum is simply a large cavity lined with absorbing material, as shown in Fig. 7-49. Sometimes a nearby room or attic space can be made into a noise-attenuating plenum, usually at the source. The attenuation realized from a plenum can be estimated from the following expression[19]

$$attenuation = 10\log\left[\frac{1}{S_e\left(\frac{\cos\theta}{2\pi d} + \frac{1-a}{S_w a}\right)}\right] \qquad (7\text{-}4)$$

where,

a is the absorption coefficient of the lining,

S_e is the plenum exit area in ft^2,

S_w is the plenum wall area in ft^2,

d is the distance between the entrance and exit in ft,

θ is the angle of incidence at the exit (i.e., the angle that the direction *d* makes with the axis of exit) in degrees.

For those high frequencies where the wavelength is smaller than plenum dimensions, accuracy is within 3dB. At lower frequencies Eq. 7-5 is conservative, and the actual attenuation can be 5dB to 10dB higher than the value it gives.

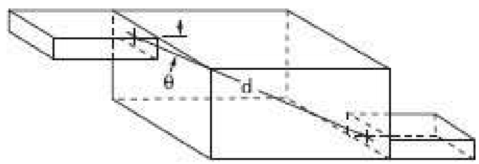


Figure 7-49. A properly designed lined plenum is a very effective attenuator of HVAC noise and is usually located near the equipment. Unused rooms or attic spaces may sometimes be converted to noise-attenuating plenums. (After ASHRAE, Reference 18.)

7.5.7 Proprietary Silencers

When space is at a premium and short runs of duct are necessary, proprietary sound-absorbing units can be installed in the ducts at critical points. There are a number of configurations available, and many attenuation characteristics can be expected. The extra cost of such units may be offset by economies their use would bring in other ways. The user should also be aware that silencers produce a small amount of self-noise and care must be taken to allow the air to return to a laminar flow downstream of the silencer.

The general rule is that the air will require a length equal to 10 times the diameter of the duct to regain a laminar flow.

7.5.8 HVAC Systems Conclusion

The intent of this HVAC section is to emphasize the importance of adequate attention to the design and installation of the heating, ventilating, and air-conditioning system in the construction of studios, control rooms, and listening rooms. HVAC noises commonly dominate in such sound rooms and are often the focus of great disappointment as a beautiful new room is placed into service. The problem is often associated with the lack of appreciation by the architect and the HVAC contractor of the special demands of sound rooms. It is imperative that an NC clause be written into every mechanical (HVAC) contract for sound-sensitive rooms.

Residential HVAC systems commonly employ small ducts and high velocity air delivery systems. Air turbulence noise increases as the sixth power of the velocity; hence, high velocity HVAC systems can easily be the source of excessive turbulence noise at grilles and diffusers. Keeping air velocity below 400ft/min for studios and other professional sound rooms is a basic first requirement. Air flow noise is generated at tees, elbows, and dampers; and it takes from 5 to 10 duct diameters for such turbulence to be smoothed out. This suggests that duct fittings should be spaced adequately. Air flow noise inside a duct causes duct walls to vibrate, tending to radiate into the space outside. Thermal duct wrapping (lagging) helps to dampen such vibrations, but even covered, such ducts should not be exposed in sound-sensitive rooms. This oversimplified treatment of HVAC design is meant to underscore the importance of employing expert design and installation talent, not to create instant experts. The overall HVAC project, however, needs the involvement of the audio engineer at each step.[17,18]

References

1. L. L. Beranek, 1989. Balanced Noise Criterion (NCB) curves. *Journal of the Acoustical Society of America* 86: 650 – 664.

2. M. Rettinger, Noise Level Limits in Recording Studios, *dB the Sound Engineering Magazine*, vol. 12, No. 4, pp. 41 – 43, April 1978.

3. L. L. Beranek, W. E. Blazier, and J. J. Figwer, Preferred Noise Criteria (PNC) Curves and Their Application to Rooms. Presented at the 81st Meeting of the ASA, Washington DC April 1971.

4. B. J. Smith, R. J. Peters, S. Owen, *Acoustics and Noise Control*, Longman Inc. New York 1982.

5. *Classification for Rating Sound Insulation*, American Society for Testing Materials, designation E413–87.

6. Ibid. 5.4.1/5.4.2.

7. M. Rettinger, *Acoustic Design and Noise Control*, New York: Chemical Publishing Co., Inc., 1973.

8. Bruel and Kjaer Instruments Inc. Technical Review # 1–1986.

9. Introduction to Seismic Vibration Control Specifications, Bulletin SVCS-110-3, Mason Industries Inc. 1999.

10. D.W. Green and C. W. Sherry, “Sound Transmission Loss of Gypsum Wallboard Partitions, Report No. 3. 2 X 4 Wood Stud Partitions, *Journal of the Acoustical Society of America*, vol. 71, no. 4, pp. 908 – 914, April 1982. (Report No.1, vol. 71, no. 1 and Report 2, vol. 71, no. 4, also reports extensive measurements of transmission loss.)

11. QuietRock copyright Quiet Solution, 125 Elko Dr., Sunnyvale, CA, 94089.

12. Architectural Specifications for Isolated Floors, Walls, and Ceilings, Bulletin ACS-100, Hollis, NY: Mason Industries, Inc., 1977.

13. F. A. Everest, *How to Build a Small Budget Recording Studio from Scratch*, Blue Ridge Summit, PA: Tab Books, Inc.

14. F. A. Everest, Glass in the Studio, *dB the Sound Engineering Magazine*, Part I, vol. 18, no. 3, pp. 28 – 33, April 1984: Part II vol. 18, no. 4, pp. 41–44, May 1984.

15. Breaking The Sound Barrier, Publication AR-3, Libbey-Owens-Ford Co.

16. *ASHRAE Handbook and Product Directory – 1977*. Fundamentals, Chapter 7, “Sound Control Fundamentals,” American Society of HVAC Engineers, Inc., 345 East 47th St., New York.

17. *ASHRAE Handbook and Product Directory – 1976*. Systems, Chapter 35, “Sound and Vibration Control,” American Society of HVAC Engineers, Inc., 345 East 47th St., New York.

18. *ASHRAE Handbook – 1985*. Fundamentals, Section 2.0, Chapter 7, “Sound and Vibration Fundamentals,” American Society of HVAC Engineers, Inc., 345 East 47th St., New York.

Chapter 8

Acoustical Treatment for Indoor Areas

by Jeff Szymanski

8.1 Acoustical Treatment Overview

It is possible that there is no area in professional audio where there is more confusion, folklore, and just plain misinformation than in the area of acoustical treatment. Everyone, it seems, is an acoustical expert. Of course, like most disciplines, much of acoustics is logical and intuitive if one understands the fundamentals. As Don Davis wrote, “In audio and acoustics the fundamentals are not difficult; the physics are.”[1] The most fundamental of all rules in acoustics is that nothing is large or small in an absolute sense. Everything is large or small relative to the wavelength of the sound under consideration. This is one of the realities that makes the greater field of audio so fascinating. Human ears respond to a range of wavelengths covering approximately 10 octaves, as compared to our eyes, which respond to a range of frequencies spanning about one (1) octave. Even though the bandwidth of visible light is obviously much larger than that of audible sound because of the much higher frequencies involved, the range of the wavelengths in this 10 octave bandwidth poses some unique challenges to the acoustician. We must be able to deal with sounds whose wavelengths are 17m (56ft) and sounds whose wavelengths are 1.7cm (0.6in).

Getting rooms to sound good is an art as much as it is a science. In some situations, concert halls, for example, there is a reasonable agreement as to what constitutes a good hall. In other applications, such as home theaters, recording studios, or houses of worship, there is little agreement among the users, let alone the consultants, as to how these rooms should sound. Considerable research must be done before we are able to trace all of the subjective aspects of room acoustics back to physical parameters. However, some fundamental rules and principles can be noted. The acoustician has very few tools. In fact, there are only two basic methods of manipulating reflective sound. It can either be absorbed or redirected, Fig. 8-1. Every room, from a humble personal listening room to the most elaborate concert hall, consists of surface materials and furnishings that to varying degrees absorb and/or redirect sound. Room acoustics boils down to the management of reflections. In some situations, reflections are problems that must be removed. In other situations, reflections are purposely created to enhance an experience.

This chapter will address general issues of modifying the way rooms sound. Absorption and absorbers will be covered in detail, as well as diffusion and diffusers, and other forms of sound redirection. Additionally, some discussion on the controversial topic of electroacoustical treatments, and brief sections that touch on life safety and the environment as they pertain to acoustical treatments are provided. The information will be thorough, but not exhaustive. There are, after all, entire books dedicated to the subject of acoustical treatments.[2] The intention here is to be able to provide a solid understanding of the fundamentals involved. Specific applications will be dealt with in subsequent chapters.

8.2 Acoustical Absorption

Absorption is the act of turning acoustical energy into some other form of energy, usually heat. The unit of acoustical absorption is the sabin, named after W.C. Sabine (1868–1919), the man considered the father of modern architectural acoustics. It is beyond the scope of this treatment to tell the story of Sabine's pioneering work on room acoustics, but it should be required reading for any serious student of acoustics. Theoretically, 1.0 sabin equates to one square meter (m^2) of complete absorption. Sabine's original work involved determining the sound absorbing power of a material. He posited that comparing the performance of a certain area of material to the same area of open window would yield its absorbing power relative to the ideal.[3] For example, if $1.0m^2$ of a material yielded the same absorbing power as $0.4m^2$ of open window, the relative absorbing power—what we now call the absorption coefficient—would be equal to 0.4.[4]

How absorption is used depends on the application and the desired outcome. Most of the time, absorption is used to make rooms feel less live or reverberant. Absorber performance varies with frequency, with most working well only over a relatively narrow range of frequencies. In addition, absorber performance is not necessarily linear over the effective frequency range.

Measuring or classifying absorbers is not as straightforward as it may seem. There are two main laboratory methods: the impedance tube method and the reverberation chamber method, both of which will be discussed in detail below. Field measurement of absorption will also be discussed below. Absorber performance can also be determined theoretically; discussions of those methods are beyond the scope of this chapter. (The reader is referred to the Bibliography at the end of this chapter for advanced absorber theory texts.)

There are three broad classifications of absorbers: porous, discrete, and resonant. While it is not uncommon for people to design and build their own absorbers (indeed, there has been something of a resurgence in do-it-yourself absorber construction in recent years as a result of the proliferation of how-to guides and Internet discussion forums—this information may or may not be reliable, depending on the reliability of the online resource and the relative expertise of the "experts" offering guidance), many useful porous and resonant absorbers are available commercially. Fundamental information about the design of absorbers is included here for two reasons: there may be those who want to build their own absorbers, and more importantly, these absorbers are sometimes inadvertently constructed in the process of building rooms. This is especially true of resonant absorbers.

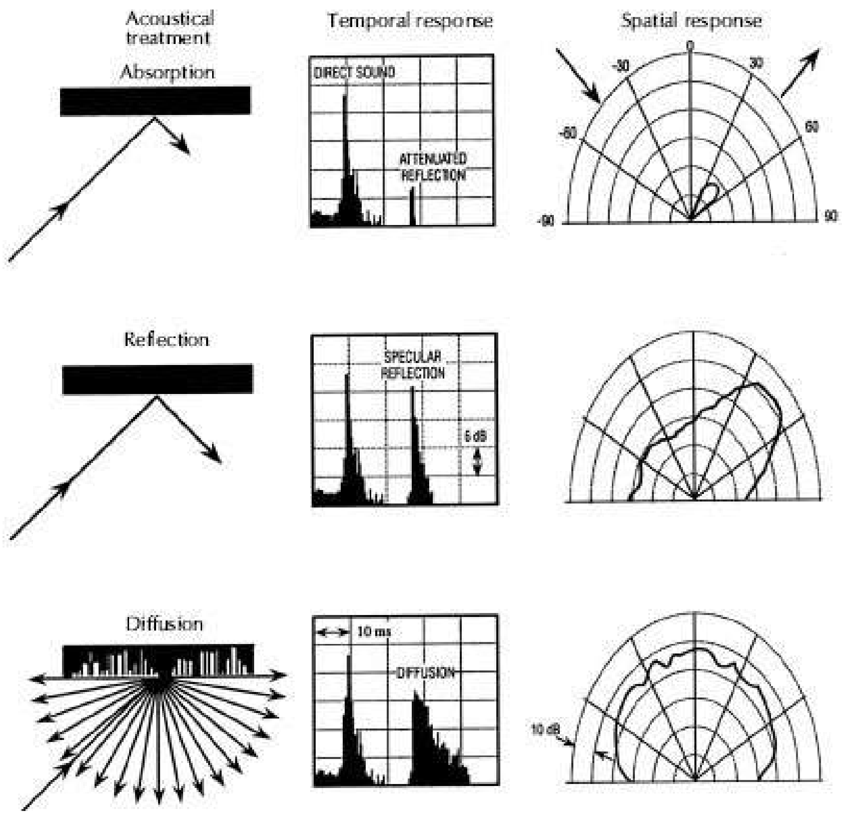


Figure 8-1. Comparison of absorption reflection and diffusion.

8.2.1 Absorption Testing

Standardized testing of absorption began with Sabine and continues to be developed and improved upon in the present day. As mentioned above, the two standardized methods for measuring absorption are the reverberation chamber method and the impedance tube method. One can also measure absorption in the field by using either the standardized methods or the other techniques discussed below.

8.2.1.1 Reverberation Chamber Method

The work of Sabine during the late 19th and early 20th centuries is echoed in the present-day standard methods for measuring absorption in a reverberation chamber: ASTM C423 and ISO 354.[5,6]

In both methods, the general technique involves placing a sample of the material to be tested in a reverberation chamber. This is a chamber that has no absorption whatsoever. The rate of sound decay of the room is measured with the sample in place and compared to the rate of sound decay of the empty room. The absorption of the sample is then calculated.

The method of mounting the sample in the test chamber has an effect on the resulting absorption. Thus, standardized methods for mounting are provided.[6,7] The most common mounting methods employed are Types A, B, and E. Type A simply involves placing the test sample—usually a board-type wall or ceiling absorber—flat against the predefined test area in the chamber (typically on the floor). Type B mounting is typically encountered with acoustical materials that are spray or trowel applied. The material is first applied to a solid backing board and then tested by placing the treated boards over the predefined test area in the chamber. Type E mounting is the standard method employed for absorbers such as acoustical ceiling tiles. This mounting includes a sealed air space of a defined depth behind the absorbers to mimic the real-world installation of acoustical ceiling tiles with an air plenum above. The depth is defined in millimeters and is denoted as a suffix. For example, a test of acoustical ceiling tiles in an E400 mounting means that the tiles were tested over sealed air space that was 400mm (16in) deep.

It should be noted that Type A mounting for board-type wall and ceiling absorbers is so often used as the default method that any mention of mounting method is often carelessly omitted in manufacturer literature. Regardless, it is important to verify the mounting method used when evaluating acoustical performance data. If there is any uncertainty, a complete, independent laboratory test report should be requested and evaluated. Details of the mounting method must be included in the lab report to fulfill the requirements of the test standards.

ASTM C423 is generally used in certified North American laboratories; ISO 354 is generally the adopted standard in European countries. The methods are very similar, but there are some noteworthy differences that can yield different testing results. A main difference that is a frequent subject of criticism is the different minimum sample sizes. The minimum area of material when testing board-type materials in accordance with ASTM C423 is 5.6m^2 (60ft^2)[5] (the recommended test area is 6.7m^2 [72 ft^2]) and that of ISO 354 is 10m^2 (107.6ft^2).[6] In general, this difference in sample size can result in a material having slightly lower absorption coefficients when tested in accordance with ISO 354 relative to the same material tested in accordance with ASTM C423. The ISO method is generally regarded as a more realistic approach when the test results are being applied to spaces that are larger than the test chamber, as is often the case. Nonetheless, ASTM test results have been widely and successfully used in architectural acoustic room design applications for many decades.

The reverberation chamber methods can also be applied to discrete absorbers, such as auditorium seating, highway barriers, office partitions, and even people. The main difference between testing the discrete absorbers and testing panel-type absorbers is how the results are reported. If a material occupies a commensurable area of a test chamber surface, absorption coefficients can be calculated. By contrast, the results of a test of some number of discrete absorbers are generally reported in sabins per unit. (Sometimes referred to as *Type J mounting* in the literature, provided the test met the standard requirements for that mounting). For example, the absorption of acoustical baffles—the type that might be hung from a factory or gymnasium ceiling—is typically reported in sabins per baffle.

When calculating absorption coefficients for board-type absorbers, the number of sabins in each frequency band is divided by the surface area of the test chamber covered by the sample material. The resulting quantity is the Sabine absorption coefficient, abbreviated α_{SAB}. The vast majority of absorption coefficients reported in the literature is Sabine absorption coefficients. Since the material is tested in a reverberant space, the Sabine absorption coefficients are useful for predetermining the acoustical properties of a space, provided that the product is intended for use in a similarly reverberant space (i.e., a space where sound can be considered to be impinging equally on a surface from all angles of incidence).

The frequency range of reverberation chamber measurements is limited. At low frequencies, modal effects can dominate the test chamber, thus making accurate measurements of sound decay difficult. At high frequencies, the chambers are large enough that the absorption of air will start to affect the measurement results. Therefore, the frequency range for a reverberation chamber test is typically limited to the one-third octave bands between 100 and 5000Hz. This is sufficient for most materials and applications as it spans a full six octaves over what is commonly referred to as the *speech* range of frequencies—i.e., the range of frequencies that are important to address design issues related to speech communication.

When acoustical treatments are specifically designed to absorb low frequencies, the reverberation chamber method can fall short. However, D'Antonio has implemented a special application of the ASTM C423 method that utilizes fixed microphone positions (as opposed to the more typical rotating microphone) that measure the decay of the actual modal frequencies of the room. Using this method, D'Antonio has been able to measure low frequency absorption down to the 63Hz octave band.[8,9] The impedance tube method (discussed below) can also be used to measure low frequency absorption, but a large tube with heavy walls (such as poured concrete) is required.

8.2.1.2 Impedance Tube Testing Methods

The laboratory methods generally involve the use of an impedance tube to measure absorption of normally incident sound—i.e., sound arriving perpendicular to the sample. There are two standard methods to measure absorption in an impedance tube: the single-microphone, standing wave method; and a two-microphone, transfer function method.[2] In general, impedance tube measurements are relatively inexpensive, are relatively simple to perform, and can be very useful in the research and development of absorber performance. In the standing wave method, for example, the normal absorption coefficient (α_n) can be calculated from

$$\alpha_n = \frac{I_i - I_r}{I_i} \qquad (8\text{-}1)$$

where,

I_i is the incident sound intensity,

I_r is the reflected sound intensity.

While the cost and time saving benefits of the impedance tube method are obvious, care should be taken since the normal absorption coefficients are not equivalent to the Sabine absorption coefficients discussed in the previous section. In fact, unlike α_{SAB}, α_n can never be greater than 1.0. In one set of experiments, α_{SAB} was as little as 1.2 times and as much as almost 5.0 times greater than α_n.[10] Regardless, there is no established empirical relationship between α_{SAB} and α_n. Normal absorption coefficients should not be used to calculate the properties of a space using standard reverberation time equations.

One main advantage offered by normal absorption coefficients is that they offer an easy way to compare the performance of two absorbers. Reverberation chambers have inherent reproducibility issues (explained in more detail below). The impedance tube can overcome this to some extent. One limitation of the impedance tube is frequency range; large tubes are needed to test low frequencies. Another is that tests of resonant absorbers tend to produce inaccurate results, because of the small sample size.

8.2.1.3 Other Absorption Testing Methods

Many methods can be employed for the measurement of sound absorption outside the confines of a laboratory test chamber or impedance tube.[2] Of course, both the reverberation chamber method and the impedance tube methods can be adopted for use in the field. In fact, Appendix X2 of ASTM C423 provides guidelines for carrying out the reverberation method in the field.[5]

When the sound impinging on an absorber is not totally random—as is the case, more often than not—there may be better methods for describing its performance. One of these methods, described by

Brad Nelson,[11] involves the analysis of a single reflection by means of signal processing techniques. Although Nelson's method describes the measurement of absorption at normal incidence, his method can be extended to determine the in situ absorption coefficients of a material at various angles of incidence, which can be particularly useful for the analysis of absorbers that are being used for reflective control in small rooms. Nelson's method was employed by the author to determine the in situ angular absorption coefficient (α_θ) of two different porous absorbers, the results of which are shown graphically in Fig. 8-2 for reflections in the 2000Hz band. The results at least partly confirm what has often been observed in recording studios: sculpted acoustical foam tends to be more consistent in its control of reflections at oblique angles of incidence relative to flat, fabric-covered, glass fiber panels of higher density. Or, to put it another way, the glass fiber panel offers more off-axis reflections than the acoustical foam panel. Of course, the relative merits of one acoustical treatment over the other are dependent on the application. The important point is that the differences are quantifiable.

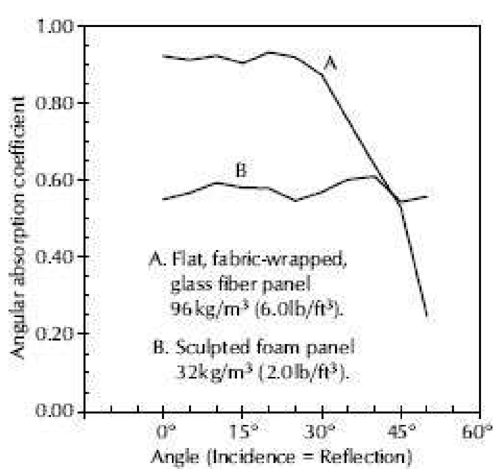


Figure 8-2. Angular absorption coefficients (α_θ) of two absorbers

for the 2000Hz one-third octave band.

8.2.1.4 Absorption Ratings

There are three single number ratings associated with absorption, all of which are calculated using the Sabine absorption coefficients. The first and most common is the *Noise Reduction Coefficient* (NRC). The NRC is the arithmetic average of the 250, 500, 1000, and 2000Hz octave-band Sabine absorption coefficients, rounded to the nearest 0.05.[5] The NRC was originally intended to be a single number rating that gave some indication of the performance of a material in the frequency bands most critical to speech.

The most recent revision of the ASTM C423 standard indicates that the NRC should be reported "in order to provide comparison with values reported in the past."[5] This could indicate a phasing out of the NRC in favor of the more robust *Sound Absorption Average* (SAA). Similar to NRC, the SAA is an arithmetic average, but instead of being limited to four octave bands, the Sabine absorption coefficients of the twelve one-third octave bands from 200 through 2500Hz are averaged and rounded to the nearest 0.01. Table 8-1 provides an example calculation of both NRC and SAA for a set of absorption coefficients.

Table 8-1. Sample Sabine Absorption Coefficient (α_{SAB}) Spectrum with Corresponding Single Number Ratings, NRC, SAA, and α_w.

⅓ Octave Band Center Frequency	α_{SAB}	⅓ Octave Band Center Frequency	α_{SAB}
100Hz	0.54	1250Hz	0.39
125Hz	1.38	1600Hz	0.31
160Hz	1.18	2000Hz	0.30

200Hz	0.88	2500Hz	0.23
250Hz	0.80	3150Hz	0.22
315Hz	0.69	4000Hz	0.22
400Hz	0.73	5000Hz	0.20
500Hz	0.56		
630Hz	0.56	**NRC = 0.55**	
800Hz	0.51	**SAA = 0.53**	
1000Hz	0.47	**α_w = 0.30 (LM)**	

Finally, ISO 11654 provides a single number rating for materials tested in accordance with ISO 354 called the weighted sound absorption coefficient (α_w).[12] A curve matching process is involved to derive the α_w of a material. Additionally, shape indicators can be included in parentheses following the α_w value to indicate areas where absorption has significantly exceeded the reference curve. Table 8-1 shows the α_w for the set of absorption coefficients, with the LM indicating that there may be excess low and mid-frequency absorption offered that is not otherwise apparent from the α_w value. This is useful in that it indicates the actual octave-band or one-third octave band absorption coefficients are probably worth looking into in greater detail.

None of the metrics described above gives an accurate representation of the absorptive behavior (or lack thereof) of a material. NRC averages four bands in the speech frequency range. The problem, of course, is that many different combinations of four numbers can result in the same average, as shown in Table 8-2. The same can be said for SAA. Nonetheless, NRC and SAA can be compared to give a little bit more information than each rating gives on an individual basis. If NRC and SAA are very close, the material probably does not have any extreme deviations in absorption across the speech range of frequencies. If SAA is drastically different from

NRC, it may be indicative of some large variations at certain one-third octave bands. These are, of course, only single number ratings; none of them takes into account the performance of the material below the 200Hz one-third octave band. They can, at most, provide a cursory indication of the relative performance of a material. A full evaluation of the performance of a material should always involve looking at the octave or one-third octave band data in as much detail as possible.

Table 8-2. Two Different Sample Sabine Absorption Coefficient (α_{SAB}) Spectra with Equal NRC and SAA.

	α_{SAB}	
⅓ Octave Band Center Frequency	**Material 1**	**Material 2**
100Hz	0.54	0.01
125Hz	1.38	0.01
160Hz	1.18	0.09
200Hz	0.88	0.18
250Hz	0.80	0.33
315Hz	0.69	0.39
400Hz	0.73	0.42
500Hz	0.56	0.57
630Hz	0.56	0.58
800Hz	0.51	0.67
1000Hz	0.47	0.73
1250Hz	0.39	0.69
1600Hz	0.31	0.60
2000Hz	0.30	0.58
2500Hz	0.23	0.65
3150Hz	0.22	0.67
4000Hz	0.22	0.80
5000Hz	0.20	0.77
NRC =	**0.55**	**0.55**
SAA =	**0.53**	**0.53**

8.2.1.5 Interpreting Test Results

As mentioned at the beginning of this chapter, the acoustical treatment industry is rife with misinformation. Test results, sadly, are no exception. Information in manufacturer literature or on their Web sites is fine for evaluating materials on a cursory basis. This information should eventually be verified, preferably with an independent laboratory test report. If manufacturers cannot supply test reports, any absorption data reported in their literature or on their Web sites should be treated as suspect.

When absorption data is evaluated, the source of the data should be understood, both in terms of which standard method was used and which independent test laboratory was used. Again, the test reports can help clear up any confusion. Close attention should be paid to subtle variations in test results, such as a manufacturer who tested the standard-minimum area of material in lieu of the standard-recommended area of material for an ASTM C423 test. If two materials are otherwise similar, a variation in sample size could explain some of the variation in measured absorption.

Additionally, there are reproducibility issues with the reverberation chamber method. Saha has reported that the absorption coefficients measured in different laboratories vary widely, even when all other factors—e.g., personnel, material sample, test equipment, etc.—are kept constant.[13] Cox and D'Antonio have found absorption coefficient variations between laboratories to be as high as 0.40.[2]

Finally, it is worth noting that Sabine absorption coefficients will often exceed 1.00. This is a source of great confusion since theory states that absorption can only vary between 0.00 (complete reflection) and 1.00 (complete absorption). However, the 0 to 1 rule generally only applies to normal absorption coefficients, which are calculated using the measurement of direct versus reflected sound intensity. Sabine absorption coefficients, however, are calculated using differences in decay rate and by dividing the measured absorption by the sample area. In theory, this should still keep the Sabine absorption coefficients below 1.00. However, edge and diffraction effects are present and are frequently cited (along with some nominal hand-waving) to explain away values greater than 1.00. Edge and diffraction effects are true and valid explanations,[14] but can be confusing in their own right. For example, samples are often tested with the edges covered—i.e., not exposed to sound. Absorption coefficients greater than 1.0 resulting from such a test can therefore be attributed mainly to diffraction effects, which is the process where sound that would not normally be incident on a sample is bent towards the sample and absorbed. The confusion arises when these test results are utilized in applications where the edges of the sample will be exposed to sound.

To better understand the edge and diffractive effects, Sauro et al have investigated the effect of varying the perimeter of an absorption test sample while keeping the area constant, and the effect of varying the area whilst keeping perimeter constant.[15,16] A comparison of the results revealed a probable relationship between measured absorption, area of sample, *and* perimeter length of sample. The inclusion of perimeter length dependence is not currently included in test standards. While other researchers have not yet validated the results of Sauro, et al, they are encouraging. It is apparent that there is still much work to be done in architectural acoustics research!

A better explanation might be simply that Sabine absorption coefficients are not percentages. The variables in the calculation of the Sabine absorption coefficient are rate of decay and test sample area. A change in the former divided by the latter is being determined, which does not strictly conform to the definition of a percentage. Based on this explanation, an α_{SAB} value greater than 1.0 simply indicates a higher absorption than a value lower than 1.00, all other factors being equal. For example, a material with a Sabine absorption coefficient of 1.05 at 500Hz will absorb more sound at 500Hz than the same area of a material having a Sabine absorption coefficient of 0.90, provided that both materials were tested in the same manner.

Regardless of the validity of Sabine absorption coefficients greater than 1.00, they are usually rounded down to 0.99 for the purposes of predictive calculations. This rounding down is especially important if, for example, equations other than the Sabine equation are used to determine reverberation time. Of course, there has been ample debate about this rounding. For example, technically it is not rounding but scaling that is being done. As Saha has pointed out, why only scale the numbers greater than 1.0—what's to be done, if anything, with the other values?[13]

8.2.2 Porous Absorbers

Porous absorbers are the most familiar and commonly available kind. They include natural fibers (e.g., cotton and wood), mineral fibers (e.g., glass fiber and mineral wool), foams, fabrics, carpets, soft plasters, acoustical tile, and so on. The sound wave causes the air particles to vibrate in the porous materials, and frictional losses convert some of the sound energy to heat energy. The amount of loss is a function of the density or how tightly packed the fibers are.

If the fibers are loosely packed, there is little frictional loss. If the fibers are compressed into a dense board, there is little penetration and more reflection from the surface, resulting in less absorption.

Mainly because there is a veritable plethora of extant information with which to work, the Owens Corning 700 Series of semi-rigid glass fiber boards will be discussed in the next section to not only highlight one of the more popular choices for porous absorber, but also to illustrate various trends—such as absorption dependence on thickness and density—that are not uncommon with porous absorbers in general.

8.2.2.1 Mineral and Natural Fibers

Of the varieties of mineral fiber, one of the most popular is the glass fiber panel or board, Fig. 8-3. The absorption of sound for various densities of Owens Corning 700 Series boards is shown in Figs. 8-4 and 8-5.[17] Fig. 8-4 shows the absorption for 2.5cm (1in) thick boards. None of the three densities absorbs well at frequencies below 500Hz. At the higher audio frequencies, the boards of 48kg/m^3 and 96kg/m^3 (3.0lb/ft^3 and 6.0lb/ft^3, respectively) densities are slightly better than the lower density 24kg/m^3 (1.5lb/ft^3) material. Fig. 8-5 shows a comparison between different densities of the 10.2cm (4.0in) thick fiberglass boards. In this case, there is little difference in absorption between the three densities.[15]

Figure 8-3. Glass fiber absorbers.

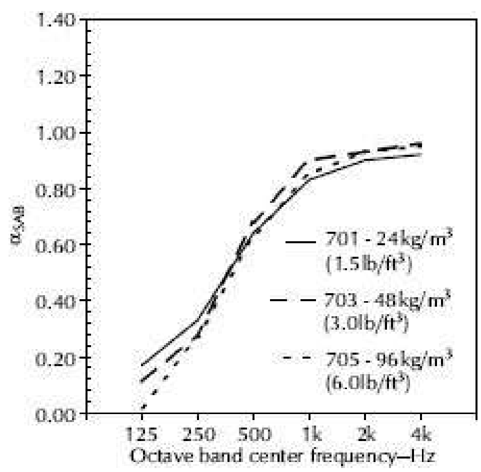


Figure 8-4. The effect of density on the absorption of Owens Corning 700 series glass fiber boards of 2.5cm (1in) thickness, Type A mounting.[15]

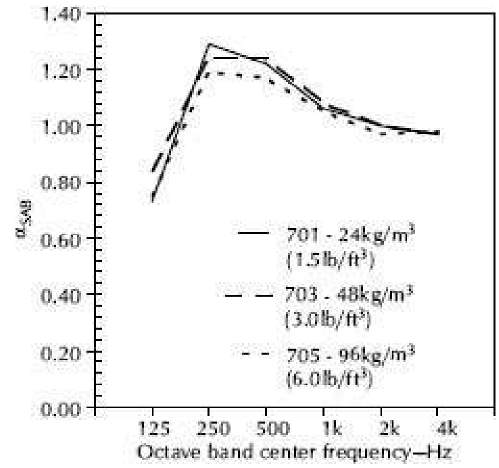


Figure 8-5. The effect of density on the absorption of Owens Corning 700 series glass fiber boards of 10.2cm (4in) thickness.[17]

Boards of medium density have a mechanical advantage in that they can be cut with a knife and press-fitted into place. This is more difficult with materials that have a 24kg/m^3 (1.5lb/ft^3) density and lower, such as building insulation. The denser the board, the greater the cost. Most acoustical purposes are well served by glass fiber of 48kg/m^3 (3.0lb/ft^3) density, although some consultants specify a 96kg/m^3 (6.0lb/ft^3) material. Some consultants regularly specify absorbers that are composed of multiple densities, for example, a combination of Owens Corning 701, 703, and 705. In theory, a multidensity absorber (assuming the least dense material is exposed to the sound source with gradually increasing densities toward the wall) will be as good as or better than a single-density absorber of the same thickness.[18] In practice, this tends to hold true.

Fig. 8-6 explores the effect of thickness of 703 Fiberglas on absorption. The absorption of low-frequency sound energy is much greater with the thicker materials.[17]

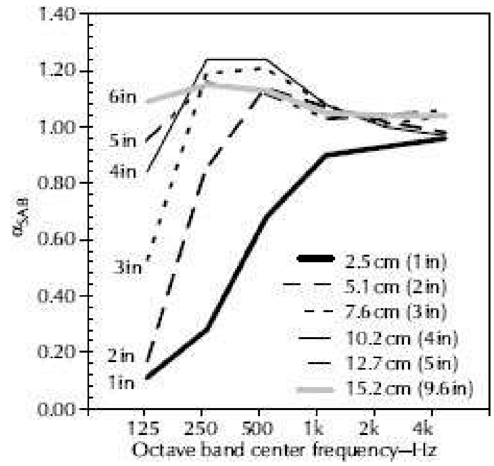


Figure 8-6. The effect of thickness on the absorption of Owens Corning 703 glass fiber boards 48 kg/m^3 (3 lb/ft^3), Type A mounting.[17]

Fig. 8-7 shows the effect of air space behind a 2.5cm (1.0in) thick Owens Corning Linear Glass Board. As the air space is increased in steps from 0.0 to 12.7cm (0.0 to 5.0in), the lower-frequency absorption increases progressively.[17] It is sometimes cost-effective to use thinner glass fiber and arrange for air space behind it; it is sometimes cost-effective to use glass fiber of greater thickness. At other times, the need for low-frequency absorption is so great that both thick material and air space are required.

In acoustical applications, mineral wool (or rock wool) is another popular variation of mineral fiber board, Fig. 8-8. Figs. 8-9 and 8-10 provide an overview of absorption coefficients for materials available from Roxul.[19] The main difference between glass fiber and mineral wool is that mineral wool is generally made from basalt (glass fiber comes from silicates), which leads to a higher heat tolerance.

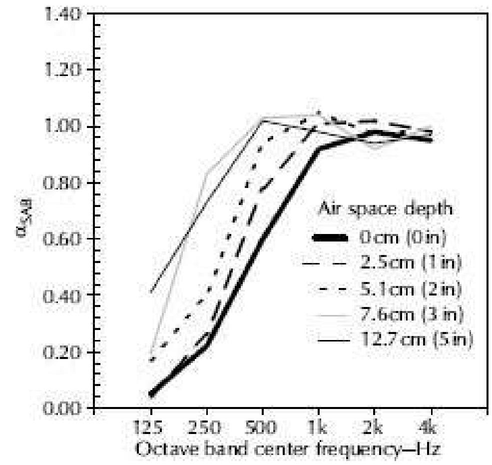


Figure 8-7. The effect of mounting over an air space on the absorption of Owens Corning Linear Glass Cloth faced board of 2.5cm (1in) thickness.[15]

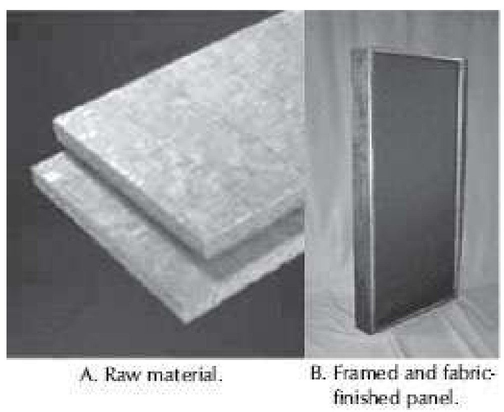


Figure 8-8. Mineral wool absorbers.

Natural fiber materials used in acoustical applications include wood fibers and cotton fibers. Tectum, Inc. manufactures a variety of ceiling and wall panels from aspen wood fibers, which produces a durable acoustical treatment. Absorption coefficients for some Tectum, Inc. materials are shown in Fig. 8-11.[20] There are also an increasing number of suppliers of natural cotton absorption panels. The absorption of natural cotton panels—so far as they have been developed—appears to be comparable to mineral fiber panels of similar density.

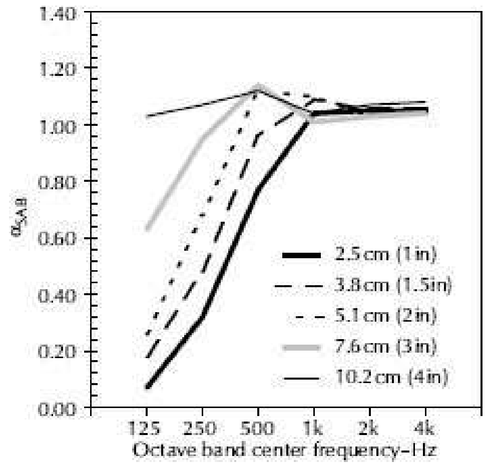


Figure 8-9. The effect of thickness on the absorption of Roxul RockBoard 40 64kg/m^3 (4lb/ft^3) mineral wool boards.[19]

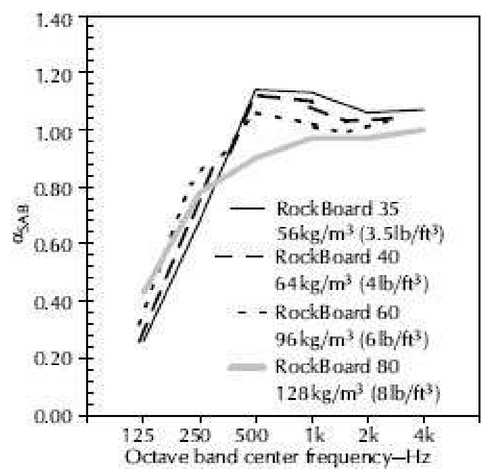


Figure 8-10. The effect of density on the absorption of Roxul RockBoard mineral wool boards of 5.1cm (2in) thickness.[19]

Most fibrous absorbers will be covered with some sort of acoustically transparent fabric finish that is both decorative and practical. The fabric finish is decorative because the natural yellow or green of the glass fiber and mineral wool panels tends to be less than aesthetically pleasing; the finish is practical because airborne fibers from mineral fiber materials can be breathing irritants. Perforated metal (with or without powder-coated finish) and plastic coverings with a high percent of open area (much higher than resonant perforated absorbers discussed below) can also be used with fibrous absorbers. Perforated coverings are typically employed for decorative purposes, maintenance purposes, or to protect the panels from high impacts, such as might occur in a gymnasium. Foil and paper finishes are also sometimes available as low-cost means of containing fibers for glass fiber or mineral wool panels. Because of reflections from the foil or paper, the high-frequency absorption of the faced side of the absorber is significantly lower than that of the unfaced side. (The thin foil or paper used is sometimes referred to as a *membrane*. This has led to confusion with resonant membrane, or diaphragmatic absorbers. For clarity, foil or paper facings as they are applied to fibrous absorber panels are not resonant membranes in the strict sense, but do provide some nominal increases in low-frequency absorption when the foil or paper is exposed to the incident sound.)

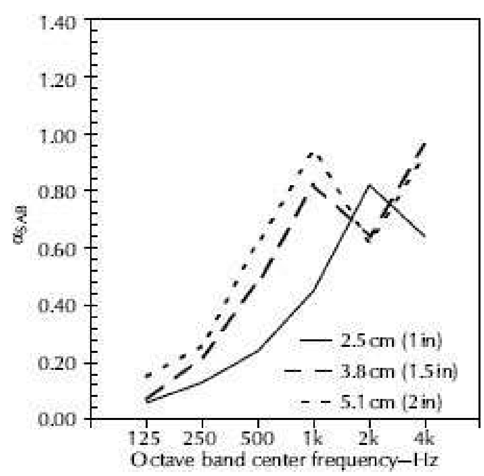


Figure 8-11. The absorption of different thicknesses of Tectum Wall Panels, Type A mounting.[20]

To provide some impact resistance, as well as to provide a surface conducive for some office applications (such as for office partitions), a thin (usually 3.0mm) glass fiber board of high density (usually 160 to 290kg/m3 [10 to 18lb/ft3]) can be applied over the face of a fibrous absorber before the fabric finish is applied. This is often referred to as a tackable surface finish since it can readily accept push pins and thumbtacks.

In terms of installation ease, natural fibers hold some promise since they will offer relief from the itch associated with the handling of mineral fiber boards. Natural fiber products can also be installed without covering, and Tectum, Inc. states that their wood fiber panels can be repainted several times without significant degradation of acoustical performance.

8.2.2.2 Acoustical Foams

There are various types of reticulated open cell foams for acoustical applications, Fig. 8-12. Closed cell foams also find applications in acoustics, but largely as substrates from which acoustical diffusers can be formed. The most common foams used as open cell acoustical absorbers in architectural applications are polyurethane (esters and ethers) and melamine foams. Unlike fibrous boards, foam panels are easy to cut and can be sculpted into shapes and patterns. Besides the ubiquitous wedges and pyramids, acoustical foams have been created with various square, saw tooth, and even curved patterns sculpted into the faces. While removing material generally serves to decrease absorption, creating more exposed surface area tends to increase it. Figs. 8-13 and 8-14 provide absorption coefficients for different patterns of foam and different thicknesses of foam of the same pattern, respectively, for acoustical foam panels available from Auralex Acoustics, Inc.[21]

Figure 8-12. Open cell polyurethane acoustical foam.

In general, acoustical foams are of lower density than fibrous materials; acoustical foam densities are generally in the 8.0 to 40kg/m^3 (0.5 to 2.5lb/ft^3) range. This means that mineral fiber panels tend to provide higher absorption coefficients than foam panels of the same thickness. However, acoustical foams can generally be installed without any decorative covering, which can make them more cost-effective—mineral fiber panels tend to require a fabric finish, or some other cover to contain airborne fibers. Melamine foams, such as the acoustical products offered by Pinta Acoustic, Inc. (formerly Illbruck) are white in color and have a higher resistance to fire relative to polyurethane foams. However, melamine foams generally have lower absorption coefficients (largely due to lower densities) and tend to be less flexible, making them more prone to damage than polyurethane foams. A sampling of the acoustical performance of some melamine foam products available from Pinta Acoustic, Inc. is provided in Fig. 8-15.[22] Melamine foams may be painted (the manufacturer should always be consulted about this), while polyurethane foams should generally not be painted. Because of this, companies offering polyurethane foams generally have a wider variety of colors available.

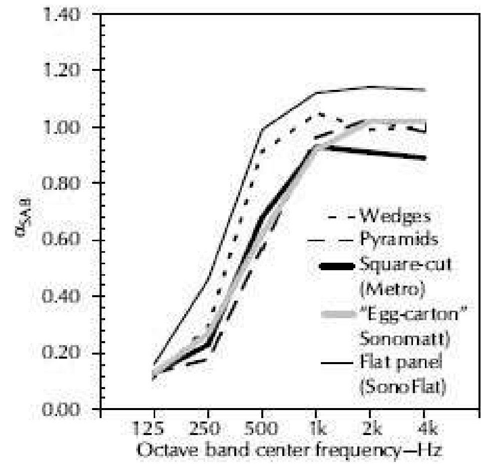


Figure 8-13. The effect of shape on the absorption of Auralex Acoustics polyurethane foam panels of 5.1cm (2in) thickness, Type A mounting.[21]

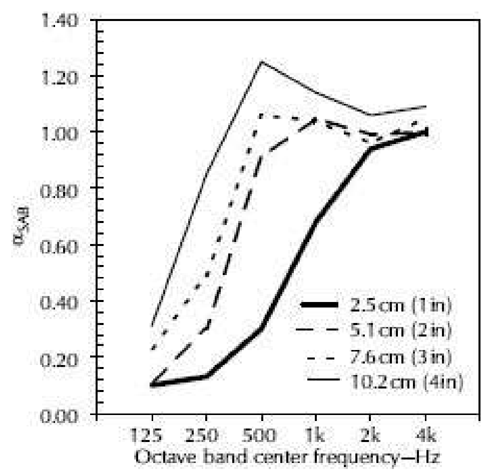


Figure 8-14. The effect of thickness on the absorption of Auralex Acoustics Studiofoam Wedges, Type A mounting.[21]

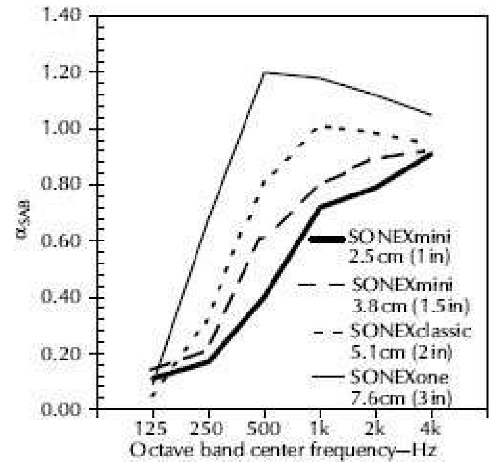


Figure 8-15. The absorption of Pinta Acoustic melamine foam panels of different thicknesses, Type A mountings.[22]

8.2.2.3 Acoustical Tiles

Acoustical tiles have the highest density of the porous absorbers. They are widely used for suspended (lay-in) ceiling treatments. Years ago, it was common to see 30cm × 30cm (12in × 12in) tile mounted directly to a hard plaster (Type A mounting). This is not a very efficient way to use this type of absorber and is no longer popular.

The standard sizes for acoustical tiles are 61cm square (24in × 24in) or 61cm × 122cm (24in × 48in) and the Sabine absorption coefficients are usually given for Type E400 mounting, which mimics a lay-in ceiling with a 400mm (16 in) air space. Fig. 8-16 shows the average absorption coefficients of a sampling of 39 different acoustical tiles. The vertical lines at each frequency point indicate the spread of the coefficients for each frequency. It is interesting to note the wide variance possible with different types of tiles.

8.2.2.4 Spray and Trowel Applied Treatments

Some acoustical treatments can be applied by spray and/or trowel. Many are applied, finished, and detailed much like standard plaster —and are even paintable. Special bonding chemicals and processes give these types of materials their absorptive qualities. Some have a gypsum base, which can provide a look similar to normal plaster or gypsum wallboard walls. Acoustical plasters tend to provide high frequency absorption, with poor low frequency performance, especially when applied thinly (<2.5cm thickness). Acoustical plasters can be an economical option when considering spaces that require large areas of absorption—e.g., a gymnasium ceiling. Some spray applied treatments can provide fireproofing, as well as thermal insulation. They are also popular in historical preservation applications, where the aesthetic appearance of a surface cannot be altered, but the acoustics must be improved to provide better communications in the space.

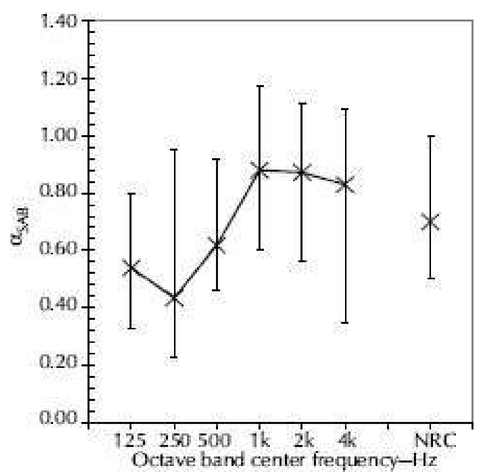


Figure 8-16. The average Sabine absorption coefficients (α_{SAB}) of 39 acoustical ceiling tiles of varying thickness, Type E400 mounting.

8.2.2.5 Carpet and Draperies

Carpet is a visual and comfort asset, and it is a porous absorber of sound, although principally at upper audio frequencies. Carpet is what the electrical engineer might call a low-pass filter. Because it is a high-frequency absorber, carpet should be used cautiously as a room treatment. Carpet can make a well-balanced room bass heavy because of its excessive high frequency absorption. The various types of carpet have different sound absorption characteristics. In general, sound absorption increases with pile weight and height; cut pile has greater absorption than loop pile. Pad material has a significant effect on the absorption of a carpet. Generally, the heavier the carpet pad, the more absorption. Impermeable backing should be used with care as it dramatically reduces the effect of the carpet pad and thereby reduces absorption. Due to the limited thickness of carpet, even the deepest possible pile (with the thickest possible pad) will not absorb much low-frequency sound. Fig. 8-17 shows the absorption coefficient for a typical medium-pile carpet, with and without a carpet pad.[23]

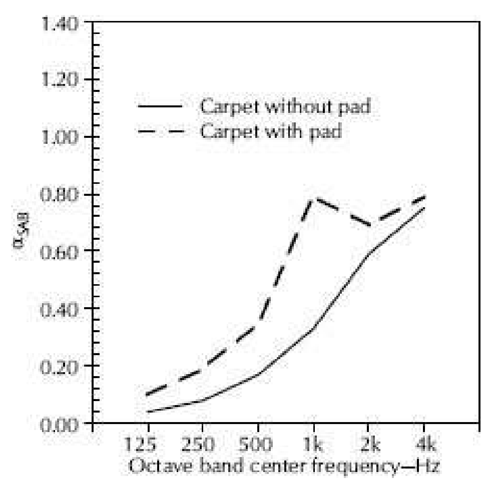


Figure 8-17. The absorption of loop pile tufted carpet (0.7kg/m^2) with and without carpet pad (1.4kg/m^2), Type A mounting.[23]

Draperies are also porous absorbers of sound. Included in the drapery category are drapes, curtains, tapestries, and other fabric wall-hangings, decorative or otherwise. Besides the type and thickness of material, the percent fullness has an effect on how well draperies absorb sound. (The percent fullness is a representation of the amount of extra material in the drapery. For example, 100% fullness would mean that a 3.0m wide curtain actually consists of a 6.0m wide piece of material. Similarly, 150% fullness would indicate that a 3.0m wide drape consists of a 7.5m wide piece of material; a 6.0m wide piece of material being used for a 2.4m wide drape, etc.) Fig. 8-18 shows the absorption coefficients for draperies with different percent fullness.[24] While spacing draperies from the wall does increase absorption slightly, it would not appear to be as significant as percent fullness, as indicated by Fig. 8-19.[24]

8.2.3 Discrete Absorbers

Discrete absorbers can literally be anything. Even an acoustical tile or foam panel is a discrete absorber. The absorption per unit of a tile, panel, board, person, bookshelf, equipment rack, etc., can always be determined. In the context of acoustical treatments, there are two main classes of discrete absorber that should never be ignored: people and furnishings.

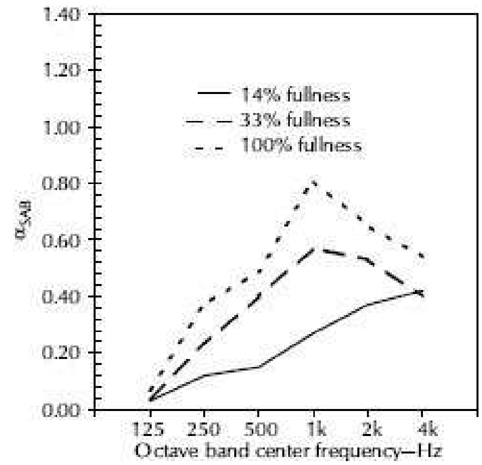


Figure 8-18. The effect of fullness on the absorption of cotton cloth curtain material, 500g/m^2 (14.5oz/yd^2).[24]

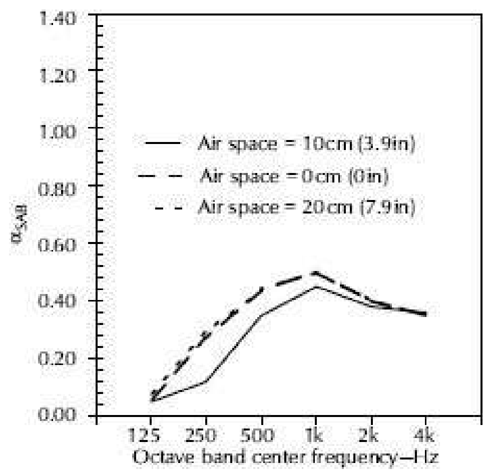


Figure 8-19. The effect of mounting over an air space on the absorption of velvet cloth curtain material, 650g/m^2 (19oz/yd^2).[24]

8.2.3.1 People and Seats

In many large spaces, people and the seats they sit in will be the single largest acoustical treatment in the room. Any acoustical analyses of sufficiently large spaces should include people in the calculations. How the seats behave acoustically when they are empty is another important consideration. Empty wood chairs will not absorb as much as the people sitting in them. A heavily padded seat may absorb just as much sound as a seated individual. A chair that folds up when not in use may have a hard, plastic cover on the underside of the seat that will reflect sound. Perforating the cover to allow sound to pass in through the bottom of the chair when it is folded up may be a worthwhile consideration for some applications. For more information on the absorption of people, seats, and audience areas in general, refer to Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*.

8.2.3.2 Acoustical Baffles and Banners

In very large rooms, such as domed stadiums, arenas, gymnasiums, factories, and even some houses of worship, absorbers need to be placed high on the ceiling to reduce reverberant sound. Installing spray applied acoustical treatments in such spaces is often uneconomical because it would be too labor-intensive. To solve this problem, prefabricated acoustical treatments that hang from the ceiling are often used. Acoustical baffles are typically 61cm × 122cm (24in × 48in)—or some other relatively manageable size—and are often approximately 3.8cm (1.5in) thick. The core material is often a rigid or semi-rigid mineral fiber, such as glass fiber, with a protective covering of polyester fabric, rip-stop nylon, or PVC. Acoustical foam panels and other porous absorption panels are often available as baffles as well. Absorption is reported as the number of sabins per baffle. Acoustical baffles are often hung vertically (perpendicular to the floor), but they can also be hung horizontally, or even at an angle. The pattern of hanging can have an effect on the overall performance of the treatments. For example, some applications will benefit more from baffles hung in two or more directions, versus simply hanging all the baffles parallel to each other in one direction. Hanging is often accomplished via factory- or user-installed grommets or hooks. Acoustical banners are simply scaled-up versions of acoustical baffles. The core absorptive material is sometimes of a slightly lower density to facilitate installing the banners so that they can be allowed to droop from a high ceiling. Sizes for banners tend to be large: 1.2m × 15m (4.0ft × 50ft) (larger sizes are not uncommon).

8.2.3.3 Other Furnishings and Objects

Anyone who has moved into a new home has experienced the absorptive power of furnishings. Rooms simply do not sound the same when they aren't filled with chairs and bookshelves and end tables and knick-knacks and so on. Even in the uncarpeted living spaces in our homes, the addition of even a small number of items can change the acoustical character of the room.

This concept was put to the test when a small room with tile floor and gypsum wallboard walls and ceilings was tested before and after the addition of two couches. The couches in question were fabric—as opposed to leather or leather substitute—and were placed roughly where they eventually wound up staying even after moving in the balance of the room's furnishings. The absorption—in sabins per couch—is shown in Fig. 8-20. (Fig. 8-20 is for illustrative purposes only—i.e., the absorption shown was not measured in a laboratory.)

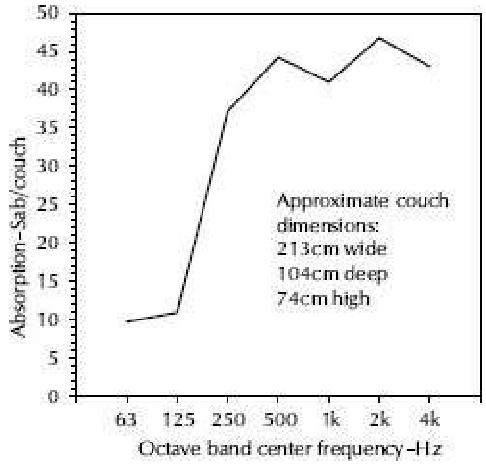


Figure 8-20. Absorption spectrum of a fabric-covered couch in a 5.6m × 4.7m × 2.3m room.

8.2.4 Resonant Absorbers

In the most general sense, a resonant absorber employs the resonant properties of a material or cavity to provide absorption. Resonant absorbers are typically pressure devices, contrasted with porous absorbers that are typically velocity devices. In other words, a porous absorber placed at a point of maximum particle velocity of the sound will provide maximum absorption. A resonant absorber placed at a point of maximum particle pressure will provide maximum absorption. This can become important in applications where maximum low-frequency performance is important. A broadband porous absorber spaced away from a surface will be the most effective method of maximizing low-frequency absorption. In contrast, a resonant absorber placed on or at the surface will provide maximum low-frequency absorption.

Resonant absorbers are often described as having been tuned to address a specific frequency range. The meaning of this will become clear below from the equations involved in determining a resonant absorber's frequency of resonance. It should be noted that many versions of the equations for resonant frequency exist in the literature. Not all of these have been presented accurately and, unfortunately, some equation errors have been perpetuated. Unfortunately, calculating the resonant frequency of a resonant absorber is not straightforward. Careful research and review were undertaken for the sections below. Unless otherwise noted, the Cox and D'Antonio[2] method of utilizing the basic Helmholtz equation as the starting point for resonance calculations was implemented in the following sections.

8.2.4.1 Membrane Absorbers

Membrane absorbers—also called *panel* and *diaphragmatic* absorbers—utilize the resonant properties of a membrane to absorb sound over a narrow frequency range. Unperforated, limp panels of wood, pressed wood fibers, plastic, or other rigid or semi-rigid material are typically employed when constructing a membrane absorber. When mounted on a solid backing, but separated from it by a constrained air space, the panel will respond to incident sound waves by vibrating. This results in a flexing of the fibers, and a certain amount of frictional loss results in absorption of some of the sound energy. The mass of the panel and the springiness of the air constitute a resonant system. In resonant systems, peak absorption occurs at the resonance frequency (f_R). Methods for determining f_R for a membrane absorber are detailed in Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*. It should be emphasized that the equations for f_R yield an approximate result. Errors in calculated versus measured f_R as high as 10% have been measured.[2] Nonetheless, membrane absorbers have been successfully used to control specific resonant modes in small rooms. To control room modes, they must be placed on the appropriate surfaces at points of maximum modal pressure. (For a detailed discussion of room modes see Chapter 6, *Small Room Acoustics*.) Adding porous absorption, such as a mineral fiber panel (typically glass fiber or mineral wool), to the cavity dampens the resonance and effectively broadens the bandwidth or Q factor of the absorber. If the Q factor is broadened, the absorber will be somewhat effective, even if the desired frequency is not precisely attained.

Additionally, care should be taken during design and construction of membrane absorbers. Changes as small as 1 to 2mm to, for example, the cavity depth can alter the performance significantly. Fig. 8-21 shows how the calculated resonant frequency varies with air space for various membranes. Other design tips can be found in Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*.

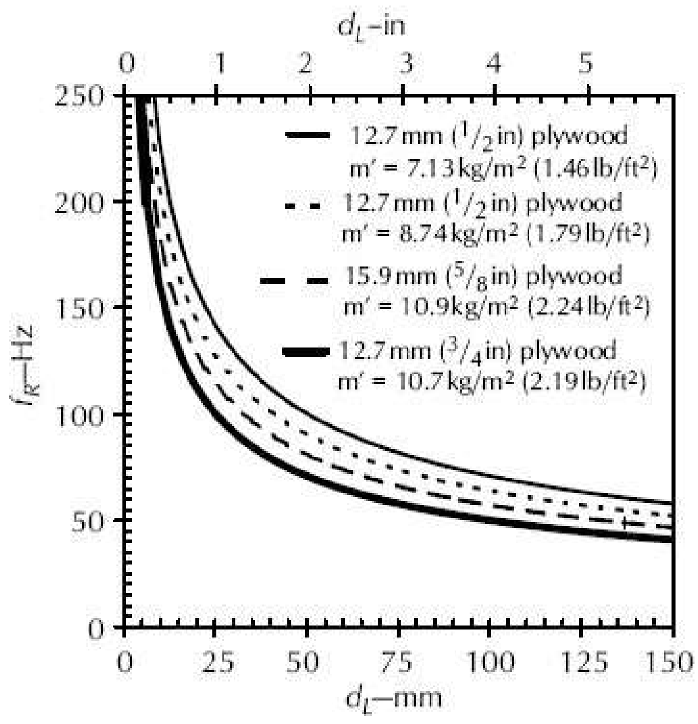


Figure 8-21. Variation of f_R vs. depth of air space for membrane absorbers consisting of common building materials.

Since membrane absorbers require a high level of precision to perform at the desired frequency, they are often customized for a specific application. Mass production is often uneconomical, although some companies offer membrane absorbers, one of which is the Modex Corner Bass Trap from RPG, Inc., with specifiable one-third-octave-band center frequencies between 40 and 100Hz.

Since there have not been many mass-produced membrane absorbers, there is far less empirical test data available on membrane absorbers relative to porous absorbers. Nonetheless, some formal testing of commercially available membrane traps has been undertaken, for example, by Noy et al.[25] Results were mixed; some membrane absorbers performed as designed, others performed well (if not exactly how the designer intended), and some did not work at all.

Putting theory into practice, Fig. 8-22 shows a pair of small room response measurements before and after the addition of a membrane absorber. Frequency is plotted linearly on the x axis (horizontal) with the resonance showing up at about 140Hz. The y axis, going into the page, is the time axis showing the decay of the room coming towards the viewer. The time span on the y axis was about 400ms. A pair of membrane absorbers was built with f_R = 140Hz. One was placed on the ceiling and one on a side wall.

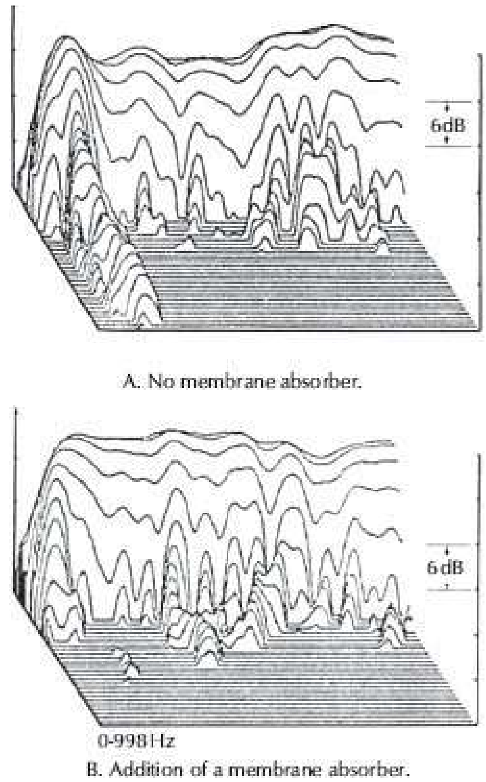


Figure 8-22. The effect of a membrane absorber in a small room.

Membrane absorbers are often inadvertently built into the structure of a room. Wall paneling, ceiling tiles, windows, coverings for orchestra pits, and even elements of furniture and millwork can all be membrane absorbers; the only question is at what frequency they resonate. Remember that everything in a room, including the room itself, has some impact on the acoustics of the room. One of the most common inadvertent membrane absorbers encountered in modern architecture is the gypsum wallboard (GWB) (drywall or sheetrock) cavity. Fortunately, the absorption of the GWB cavity can be calculated, and the calculated results have been shown to be in good agreement with laboratory measurements.[26] Chapter 13, *Acoustical Modeling and Auralization* provides discussion and calculation methods for GWB cavity absorption.

8.2.4.2 Helmholtz Resonators

The ubiquitous cola bottle may be the acoustician's most cherished conversation piece. Bottles and jugs are probably the most common everyday examples of what are referred to in acoustics as *Helmholtz resonators*. As part of his exhaustive and painstakingly detailed work in hearing, sound, and acoustics, Hermann von Helmholtz determined and documented the acoustical properties of an enclosed volume with a relatively small aperture.[27] Helmholtz resonators, as we now know them, have specialized absorptive properties for acoustical applications. At the frequency of resonance, absorption is very high. The frequency range of this absorption is very narrow—only a few Hz wide, typically. When absorptive material, such as loose mineral fiber, is used to partially fill a Helmholtz resonator, the effective frequency range is widened.

Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls* details methods of calculation for the f_R of a Helmholtz resonator. Commercially, one of the most common products utilizing Helmholtz resonator theory is sound absorbing concrete masonry units (CMU). For example, Fig. 8-23 provides the sound absorption data for SoundBlox products available from Proudfoot.[28]

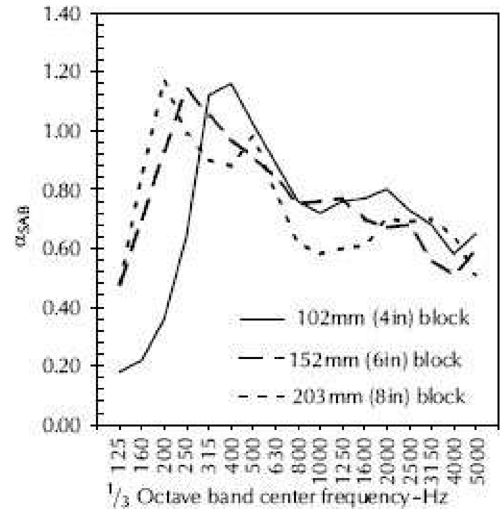


Figure 8-23. The absorption of absorbent concrete blocks from Proudfoot, Inc. Type RSC SoundBlox, utilizing Helmholtz resonance effects.[28]

8.2.4.3 Perforated Membrane Absorbers

Membrane absorbers and Helmholtz resonators are dependent on the size of the air space or cavity they contain. Turning the former into the latter can be accomplished by cutting or drilling openings in the face of the membrane. The tuned cavity of a membrane absorber then becomes the cavity of a Helmholtz resonator. When round holes are used for the openings in the face, a perforated absorber is created. To calculate the f_R for a perforated membrane absorber, first the effective thickness must be calculated. For perforated panels of having holes of diameter d and regular hole spacing S (center-to-center distance between holes), Eq. 8-2 yields the fraction of open area, e

$$\varepsilon = \frac{\pi}{4}\left(\frac{d}{S}\right)^2 \qquad (8\text{-}2)$$

To calculate the effective thickness for a perforated absorber, a correction factor, δ, is required. This factor is often approximated to 0.85, but can be calculated for low values of e (typically <0.16) using

$$\delta = 0.8(1 - 1.4\sqrt{\varepsilon}) \qquad (8\text{-}3)$$

Next, the effective panel thickness t' for a panel of thickness t is calculated from Eq. 8-3 using δ

$$t' = t + \delta d \qquad (8\text{-}4)$$

Finally, f_R for a panel over an air space of depth D is found with

$$f_R = \frac{c}{2\pi}\sqrt{\frac{\varepsilon}{t'D}} \qquad (8\text{-}5)$$

Care should be taken to be consistent with units. For example, if inches are used to calculate e, etc., c (the speed of sound) should be in inches per second.

The f_R of perforated absorbers is generally adjusted by changing e. Increasing e (larger holes, smaller spacing, or both), decreasing D, or using thinner panels will all increase the f_R. The f_R can be lowered by decreasing e, by increasing D, or by using thicker panels. The f_R from Eq. 8-5 is not exact, but is close enough for use in the design stage. The air space is often partially or completely filled with porous absorption. The only drawback to this is that absorptive material in contact with the perforated panel can reduce the absorber's effectiveness.

One of the more obvious perforated membranes that can be used is common pegboard. Standard pegboard tends to create an absorber with an f_R in the 250–500Hz range, as shown in Fig. 8-

24.[17] Since perforated absorbers are often considered for low frequency control, it is not uncommon to fabricate customized perforated boards. For example, a hardboard membrane with d = 6.4mm (¼in), S = 102mm (4in), t = 3.2mm (⅛in), and D = 51mm (2in), a perforated absorber tuned to roughly 125 to 150Hz can be created. The absorption coefficients of such an absorber with 96kg/m^3 (6.0lb/ft^3) of glass fiber filling the air space are shown in Fig. 8-25.

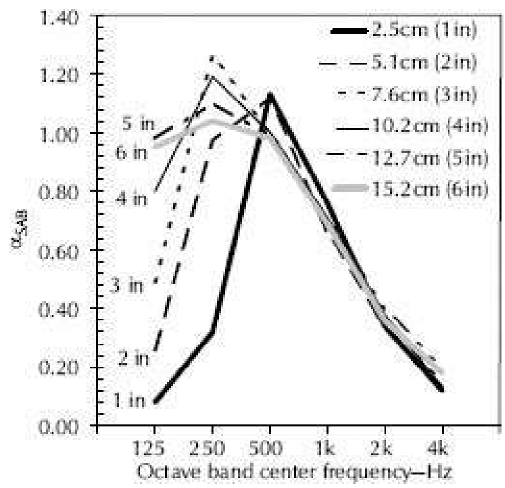


Figure 8-24. The effect of pegboard facing on the absorption of different thicknesses of Owens Corning 703 glass fiber boards, Type A mounting.[17]

Microperforated materials are one of the most recent developments in the area of acoustical treatments. Extremely thin materials with tiny perforations (≪1mm) are stretched over an air space and absorption occurs by means of boundary layer effects.[2] Because they are so thin, microperforated absorbers can be fabricated from visually transparent material. RPG offers the ClearSorber, which can be installed by stretching it over glass without significantly altering the light throughput. The absorption coefficients, dependent on the depth of air space between the microperforated foil and the glass, are shown in Fig. 8-26.[29]

8.2.4.4 Slat Absorbers

Helmholtz resonators can also be constructed by using spaced slats over an air space (with or without absorptive fill). The air mass in the slots between the slats reacts with the springiness of the air in the cavity to form a resonant system, much like that of the perforated panel type. In fact, Eq. 8-5 is again used to calculate the f_R for a slat absorber, but with the following equations for ε, δ, and t':

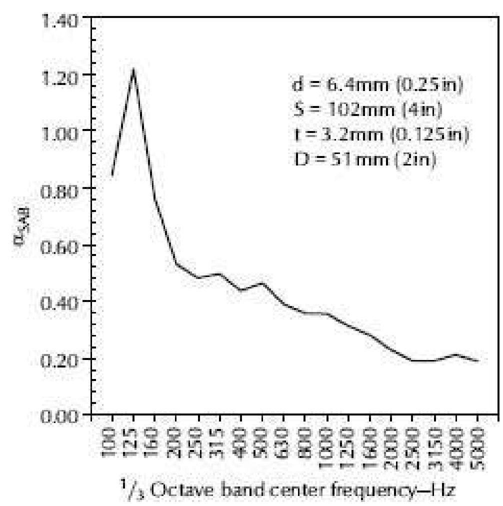


Figure 8-25. The absorption of a perforated membrane absorber “tuned” to 125–150Hz, Type A mounting, cavity filled with 96kg/m^3 (6lb/ft^3) glass fiber.

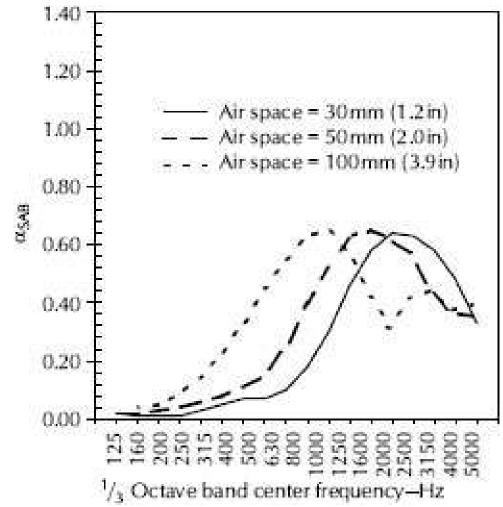


Figure 8-26. The effect of depth of air space on the absorption of a 105μm thick microperforated absorber, the RPG ClearSorber Foil from RPG, Inc.[29]

$$\varepsilon = \frac{r}{w+r} \tag{8-6}$$

$$\delta = -\frac{1}{\pi}\ln\left[\sin\left(\frac{1}{2}\pi\varepsilon\right)\right] \tag{8-7}$$

$$t' = t + 2\delta r \tag{8-8}$$

where,

r is the slot width,

w is the slat width.

While δ is often approximated to a value near 0.6, it is not difficult to calculate. As with perforated absorbers, the above will yield approximate results for the f_R of a slat absorber, which will be fine for most design applications.

In a practical sense, the absorption curve can be broadened by using a variable depth for the air space. Another approach is using slots of different widths. In the structure of Fig. 8-27, both variable air space depth and variable slot width could be used. Porous absorptive fill is shown at the back of the cavity, removed from the slats. This gives a sharper absorption than if the absorbent is in contact with the slats. It should be noted that, all other factors remaining the same, randomly placed slats (yielding randomly sized slots) will lower the overall absorption, while bandwidth is increased.[30]

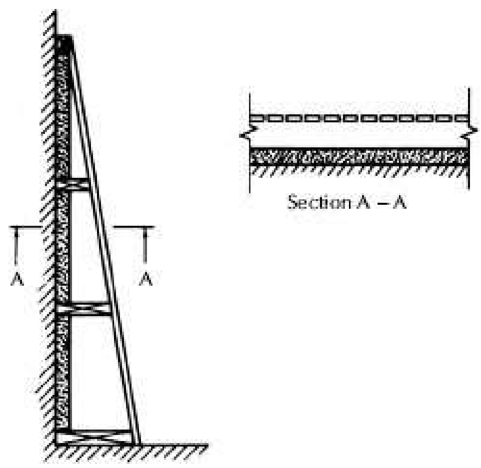


Figure 8-27. A slat absorber with varying depth to widen the effective frequency range of absorption.

8.2.5 Bass Traps

Bass trap lore pervades the professional audio industry, particular in the recording industry. Literature on the subject, however, is scarce. The term has become a catchphrase often used by acoustical product manufacturers to include a large variety of acoustical products, often including what are simply broadband absorbers. Very few bass traps are actually effective at absorbing bass. It is quite difficult to absorb sounds with wavelengths at or approaching 56ft (17m). To most effectively absorb a given frequency at any angle of incidence, including normal incidence the absorbing material should be at least 1⁄10 and, ideally, ¼ of the wavelength of the lowest frequency of interest. For 20Hz, this means a depth between 1.7m (minimum) and 4.3m (ideal)! It is relatively rare to find someone who is willing to build a device that large to trap bass. This may be necessary in the design of very large rooms, like concert halls, but it is probably more interesting to ask what crime has the 20Hz committed that it needs to be trapped? As we shall see in the next chapter, the low end performance of small rooms can be reliably predicted from a study of the distribution of room modes. If the modes are distributed properly, trapping may not be needed. On the other hand, imagine that the modes are not distributed properly and the goal is to fix a problem room. If there is enough space to build a bass trap that is large enough to have an impact on wavelengths that large, most likely one could move a wall, improving, if not optimizing, the modal distribution and obviating the need for trapping.

Nonetheless, the preceding sections have provided many examples of products that could be designed to trap bass without taking up much space. Additionally, there are good broadband absorbers on the market that extend down into the low-frequency region. The practical limits of size and mounting usually lead to a natural cutoff between 50 and 100Hz for many broadband absorbers. These products exhibit performance that is highly dependent on placement, especially in small rooms.

8.2.6 Applications of Absorption

In large rooms (see Chapter 6, *Small Room Acoustics* for the definition of large vs. small rooms) where there is a statistically reverberant sound field, absorption can be used to actually modify the reverberant field, and the results are predictable and fairly straightforward. The whole concept of reverberation time (RT—discussed in detail in Chapter 6) is a statistical one that is based on the assumption of uniform distribution of energy in the room and complete randomness in the direction of sound propagation. In large rooms, both conditions can prevail.

In small rooms—particularly at low frequencies—the direction of propagation is by no means random. Because of this, the propriety of applying the common RT equations to small rooms is questioned. For small rooms (nonreverberant spaces), absorption is useful for control of discrete reflections from surfaces located in the near field of the source and listener. In rooms the size of the average recording studio or home theater, a true reverberant field cannot be found. In such small rooms, the common RT equations cannot be used reliably. Further, predicting or trying to measure RT in small spaces where a reverberant field cannot be supported is typically less useful relative to other analysis techniques. The results of RT measurements in small rooms will neither show RT in the true sense, nor will they reveal much of value regarding the behavior of sound in a small room relative to the time domain. Typically, it is more useful to examine the behavior of sound in the time domain in more detail in small rooms. Determining the presence of reflections (wanted or unwanted), the amplitude of those reflections, and their direction of arrival at listening positions is typically a better approach. For low frequencies, Chapter 6 provides some small room analysis techniques that are more beneficial than the measurement of RT.

8.2.6.1 Use of Absorption in Reverberant Spaces

In large rooms, the common RT equations can be used with reasonable confidence. When absorptive treatment is not uniformly distributed throughout the space, the Sabine formula is typically avoided in lieu of other formulas. RT is covered in detail in Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*.

In reverberant spaces, the selection of absorbers can be based on the absorption data collected in accordance with ASTM C423, as described above in Section 8.2.1.1. Care should be taken, however, to somehow account for effects not directly evident from laboratory measurement methods. As one example, consider a fabric-wrapped, mineral fiber panel that is tested in a Type A mounting configuration. The test specimen is placed in the reverberation chamber directly against a hard (typically solid concrete) surface, often the floor. The absorption coefficients then represent only the absorption provided by the panels. In practice, panels such as these might be directly applied to a GWB surface having absorption characteristics of its own that are significantly different than the solid concrete floor of a reverberation chamber! Applying an absorptive panel to a GWB wall or ceiling not only changes the acoustical behavior of the GWB surface (by changing the mass), but the panel itself will not absorb as measured in the lab, because of the mounting, the size of the room relative to the laboratory test chamber, the proportion of absorptive material relative to the total surface area of the room, etc. This is one example of why the predictive modeling for the acoustics of large spaces can be—like many aspects of acoustics—as much art as science. All acousticians are likely to have methods they use to account for idiosyncrasies that can neither be measured in a laboratory, nor modeled by a computer.

In addition, it is generally agreed among acousticians that reverberation time is no longer considered the single most important parameter in music hall and large auditorium acoustics. Reverberation time is understood to be one of several important criteria of acoustical quality of such halls. Equal or greater stress is now placed on, for example, the ratio of early arriving energy to total sound energy, the presence of lateral reflections, the timing of the arrival of various groups of reflections, and other parameters discussed in detail in Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*.

8.2.6.2 Use of Absorption in Nonreverberant Spaces

In rooms where there is not enough volume or a long enough mean free path to allow a statistical reverberant field to develop, one must view the use of absorption in a somewhat different manner. As alluded to previously, the common RT equations will not work satisfactorily in these spaces.

Absorption is often used in small rooms to control discrete reflections or to change the way the room feels. Contrary to popular belief, the impression of liveness or deadness is not based on the length of the reverberation time. Rather, it is based on the ratio of direct to reflected sound and on the timing of the early reflected sound field, especially in the first 20ms or so. Adjusting the acoustics of nonreverberant spaces (sometimes referred to as *tuning the room*) involves manipulating discrete reflections.

To determine the suitability of a particular absorber, the acoustician needs a direct measurement of the reflected energy from the product. In small rooms, field measurement of absorption, such as the techniques and methods presented in Section 8.2.1.3, might be more appropriate in determining the applicability of a particular absorber to a small room application.

8.2.7 Subjective Impact of Absorption

Sometimes it is useful to consider the extremes. It is interesting to note that rooms with no absorption and rooms with total absorption represent the most acoustically hostile spaces imaginable. At one extreme, there is the absorption-free space, also known as the reverberation chamber. A good real-world example of this is a racquetball court. As anyone who has played racquetball can readily attest to, racquetball courts are not acoustically friendly! At the other extreme is the anechoic chamber. This is a room that is totally absorptive and totally quiet. Since an anechoic chamber has no reflected sound and is isolated from sounds from the outside, a good real-world example of this is the desert. Standing in a part of the desert free of reflective surfaces, such as buildings and mountains, located many kilometers from any noise sources, such as highways and people, at a time when there is no wind, the complete lack of sound would be comparable to what one would experience in an anechoic chamber. It is difficult to describe just how disorienting spending time in either of these chambers can be. Neither the reverberation chamber nor the anechoic chamber is a place where a musician would want to spend much time, let alone perform!

The use of absorption has a powerful impact on the subjective performance of a room. If too much absorption is used, the room will feel too dead—i.e., too much like an anechoic chamber. If too little absorption is used, the room will feel uncomfortably live—i.e., too much like a reverberation chamber. Additionally, the absorption of any material or device is frequency-dependent; absorbers act like filters to the reflected sound. Some energy is turned into heat, but other frequencies are reflected back into the room. Choosing an absorber that has a particularly nonlinear response can result in rooms that sound strange.

More often than not, the best approach is a combination of absorbers. For example, large spaces that already have the seats, people, and carpet as absorbers may benefit from a combination of membrane absorbers and porous absorbers. In a small room, some porous absorbers mixed with some Helmholtz resonators might provide the best sound for the room. Both of these are examples of the artistic (the aural aesthetic) being equally applied with the science (the acoustic physics).

Experience is important when considering the application of absorption and—more importantly—when considering what a particular application will sound like. The savvy acoustician will realize the aural differences between a small room treated with 5.1cm (2.0in) acoustical foam and a room treated with 2.5cm (1.0in) glass fiber panels. On paper, these materials are quite similar (compare Figs. 8-6 and 8-14). However, the knowledge that a room treated with 9.3m^2 (100ft^2) of foam generally sounds darker than a room treated with the same area of 96kg/m^3 fabric-wrapped, glass fiber boards comes only with experience. Likewise, the knowledge that a room treated with a slotted concrete block wall will sound much different than the same room with a GWB wall that is treated with several well-placed perforated absorbers (even though RT predictions for each scenario come out to be approximately the same) comes only with experience.

Table 8-3 gives the absorption coefficients of various popular building materials.

Table 8-3. Absorption Data for Common Building Materials and Acoustical Treatments (All Materials Type A Mounting Unless Noted Other

Material	Fig.	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	Source
Walls & Ceilings								
Brick, unpainted		0.03	0.03	0.03	0.04	0.05	0.07	Ref. 23
Brick, painted		0.01	0.01	0.02	0.02	0.02	0.03	Ref. 23
Concrete block, unpainted		0.36	0.44	0.31	0.29	0.39	0.25	Ref. 23
Concrete block, painted		0.10	0.05	0.06	0.07	0.09	0.08	Ref. 23
One layer 13mm (½in) GWB, Mounted on each side of 90mm (3.5in) metal studs. No cavity insulation		0.26	0.10	0.05	0.07	0.04	0.05	Ref. 26
Two layers 13mm (½in) GWB. Mounted on each side of 90mm (3.5in) metal studs. No cavity insulation		0.15	0.08	0.06	0.07	0.07	0.05	Ref. 26
One layer 13mm (½in) GWB. Mounted on each side of 90mm (3.5in) metal studs. With glass fiber cavity insulation		0.14	0.06	0.09	0.09	0.06	0.05	Ref. 26
One layer 13mm (½in) GWB. Mounted on one side of 90mm (3.5in) metal studs. With or without cavity insulation		0.12	0.10	0.05	0.05	0.04	0.05	Ref. 26
Plaster over tile or brick, smooth finish		0.01	0.02	0.02	0.03	0.04	0.05	Ref. 23
Plaster on lath, rough finish		0.14	0.10	0.06	0.05	0.04	0.03	Ref. 23
Plaster on lath, smooth finish		0.14	0.10	0.06	0.04	0.04	0.03	Ref. 23
Floors								
Heavy carpet without pad		0.02	0.06	0.14	0.37	0.60	0.65	Ref. 23
Heavy carpet with pad		0.08	0.24	0.57	0.69	0.71	0.73	Ref. 23
Concrete or terrazzo		0.01	0.01	0.02	0.02	0.02	0.02	Ref. 23
Linoleum, rubber, cork tile on concrete		0.02	0.03	0.03	0.03	0.03	0.02	Ref. 23
Parquet over concrete		0.04	0.04	0.07	0.06	0.06	0.07	Ref. 23
Marble or glazed tile		0.01	0.01	0.01	0.01	0.02	0.02	Ref. 23
Other								
Ordinary window glass		0.35	0.25	0.18	0.12	0.07	0.04	Ref. 23
Double glazing (1.4–1.6cm thick)		0.10	0.07	0.05	0.03	0.02	0.02	Ref. 2
Water surface		0.01	0.01	0.01	0.02	0.02	0.03	Ref. 23
Acoustical Treatments	**Fig.**							
2.5cm (1in) Owens Corning 701	4-4	0.17	0.33	0.64	0.83	0.90	0.92	Ref. 17
2.5cm (1in) Owens Corning 703	4-4, 4-6	0.11	0.28	0.68	0.90	0.93	0.96	Ref. 17
2.5cm (1in) Owens Corning 705	4-4	0.02	0.27	0.63	0.85	0.93	0.95	Ref. 17
10.2cm (4in) Owens Corning 701	4-5	0.73	1.29	1.22	1.06	1.00	0.97	Ref. 17
10.2cm (4in) Owens Corning 703	4-5, 4-6	0.84	1.24	1.24	1.08	1.00	0.97	Ref. 17
10.2cm (4in) Owens Corning 705	4-5	0.75	1.19	1.17	1.05	0.97	0.98	Ref. 17
5.1cm (2in) Owens Corning 703	4-6	0.17	0.86	1.14	1.07	1.02	0.98	Ref. 17

Material		125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	Source
7.6 cm (3 in) Owens Corning 703	4-6	0.53	1.19	1.21	1.08	1.01	1.04	Ref. 17
12.7 cm (5 in) Owens Corning 703	4-6	0.95	1.16	1.12	1.03	1.04	1.06	Ref. 17
15.2 cm (6ϖ) Owens Corning 703	4-6	1.09	1.15	1.13	1.05	1.04	1.04	Ref. 17
2.5 cm (1 in) Owens Corning. Linear Glass Cloth Board. No airspace	4-7	0.05	0.22	0.60	0.92	0.98	0.95	Ref. 17
2.5 cm (1 in) Owens Corning. Linear Glass Cloth Board. 2.5 cm (1 in) airspace	4-7	0.04	0.26	0.78	1.01	1.02	0.98	Ref. 17
2.5 cm (1 in) Owens Corning. Linear Glass Cloth Board. 5.1 cm (2 in) airspace	4-7	0.17	0.40	0.94	1.05	0.97	0.99	Ref. 17
2.5 cm (1 in) Owens Corning. Linear Glass Cloth Board. 7.6 cm (3 in) airspace	4-7	0.19	0.83	1.03	1.04	0.92	1.00	Ref. 17
2.5 cm (1 in) Owens Corning. Linear Glass Cloth Board. 12.7 cm (5 in)	4-7	0.41	0.73	1.02	0.98	0.94	0.97	Ref. 17
2.5 cm (1 in) Roxul RockBoard 40	4-9	0.07	0.32	0.77	1.04	1.05	1.05	Ref. 19
3.8 cm (1½ in) Roxul RockBoard 40	4-9	0.18	0.48	0.96	1.09	1.05	1.05	Ref. 19
5.1 cm (2 in) Roxul RockBoard 40	4-9, 4-10	0.26	0.68	1.12	1.10	1.03	1.04	Ref. 19
7.6 cm (3 in) Roxul RockBoard 40	4-9	0.63	0.95	1.14	1.01	1.03	1.04	Ref. 19
10.2 cm (4 in) Roxul RockBoard 40	4-9	1.03	1.07	1.12	1.04	1.07	1.08	Ref. 19
5.1 cm (2 in) Roxul RockBoard 35	4-10	0.26	0.68	1.14	1.13	1.06	1.07	Ref. 19
5.1 cm (2 in) Roxul RockBoard 60	4-10	0.32	0.81	1.06	1.02	0.99	1.04	Ref. 19
5.1 cm (2 in) Roxul RockBoard 80	4-10	0.43	0.78	0.90	0.97	0.97	1.00	Ref. 19
2.5 cm (1 in) Tectum Wall Panel	4-11	0.06	0.13	0.24	0.45	0.82	0.64	Ref. 20
3.8 cm (1½ in) Tectum Wall Panel	4-11	0.07	0.22	0.48	0.82	0.64	0.96	Ref. 20
5.1 cm (2 in) Tectum Wall Panel	4-11	0.15	0.26	0.62	0.94	0.62	0.92	Ref. 20
5.1 cm (2 in) Auralex Studiofoam Wedge	4-13, 4-14	0.11	0.30	0.91	1.05	0.99	1.00	Ref. 21
5.1 cm (2 in) Auralex Studiofoam Pyramid	4-13	0.13	0.18	0.57	0.96	1.03	0.98	Ref. 21
5.1 cm (2 in) Auralex Studiofoam Metro	4-13	0.13	0.23	0.68	0.93	0.91	0.89	Ref. 21
5.1 cm (2 in) Auralex Sonomatt	4-13	0.13	0.27	0.62	0.92	1.02	1.02	Ref. 21
5.1 cm (2 in) Auralex Sonoflat	4-13	0.16	0.46	0.99	1.12	1.14	1.13	Ref. 21
5.1 cm (1 in) Auralex Studiofoam Wedge	4-14	0.10	0.13	0.30	0.68	0.94	1.00	Ref. 21
5.1 cm (3 in) Auralex Studiofoam Wedge	4-14	0.23	0.49	1.06	1.04	0.96	1.05	Ref. 21
5.1 cm (4 in) Auralex Studiofoam Wedge	4-14	0.31	0.85	1.25	1.14	1.06	1.09	Ref. 21
2.5 cm (1 in) SONEXmini	4-15	0.11	0.17	0.40	0.72	0.79	0.91	Ref. 22
3.8 cm (1½ in) SONEXmini	4-15	0.14	0.21	0.61	0.80	0.89	0.92	Ref. 22
5.1 cm (2 in) SONEXclassic	4-15	0.05	0.31	0.81	1.01	0.99	0.95	Ref. 22
7.6 cm (3 in) SONEXone	4-15	0.09	0.68	1.20	1.18	1.12	1.05	Ref. 22
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in) centers over 2.5 cm (1 in) thick Owens-Corning 703	4-25	0.08	0.32	1.13	0.76	0.34	0.12	Ref. 17
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in) centers over 5.1 in	4-25	0.26	0.97	1.12	0.66	0.34	0.14	Ref. 17
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in) centers over 7.6 in	4-25	0.49	1.26	1.00	0.69	0.37	0.15	Ref. 17
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in centers over 10.2 in	4-25	0.80	1.19	1.00	0.71	0.38	0.13	Ref. 17
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in) centers over 12.7 in	4-25	0.98	1.10	0.99	0.71	0.40	0.20	Ref. 17
Pegboard with 6.4 mm (¼ in) holes on 2.5 cm (1 in) centers over 15.2 in	4-25	0.95	1.04	0.98	0.69	0.36	0.18	Ref. 17

8.3 Acoustical Diffusion

Compared to acoustical absorption, the science of acoustical diffusion is relatively new. The oft-cited starting point for much of the science of modern diffusion is the work of Manfred Schroeder. In fact, acoustically diffusive treatments that are designed using one of the various numerical methods that will be discussed below are often referred to generically as Schroeder diffusers. In the most basic sense, diffusion can be thought of as a special form of reflection. Materials that have surface irregularities on the order of the wavelengths of the incident sound waves will exhibit diffusive properties. Ideally, a diffuser will redirect the incident acoustical energy equally in all directions and over a wide range of frequencies. However, it is often impractical to construct a device that can diffuse effectively over the entire audible frequency range. Most acoustical diffuser products are designed to work well over a specific range of frequencies, typically between 2 and 4 octaves above roughly 500Hz. Of course, just as with absorbers, one must be concerned with the performance of a diffuser.

8.3.1 Diffuser Testing: Diffusion, Scattering, and Coefficients

The performance of a diffuser can be expressed as the amount of diffusion and as the amount of scattering provided by a surface. While there is still some disagreement as to diffusion nomenclature, Cox and D'Antonio have attempted to establish a distinction between diffusion and scattering, particularly as it relates to the coefficients that are used to quantify diffuser performance.[2] In general, diffusion and the diffusion coefficients relate to the uniformity of the diffuse sound field created by a diffuser. This is most easily explained by looking at polar plots of the diffuse sound field from a diffuser. The less lobing there is in a diffusion polar plot, the more diffusion and the higher the diffusion coefficients.

Scattering and the scattering coefficients relate to the amount of energy that is not reflected in a specular manner. The term specular here denotes the direction of reflection one would expect if the sound were reflecting off a hard flat surface. For example, most high frequency sounds reflect from a hard flat surface at the same angle as the incident sound, Fig. 8-1. This is referred to as specular reflection by Cox and D’Antonio. The more sound that is reflected in a nonspecular manner, the higher the scattering. Therefore, a simple angled wall can provide high scattering but low diffusion, since the reflected sound will still form a lobe, but not in a specular direction. It should be noted that a significant amount of absorption makes it difficult to measure scattering. This makes sense since an absorber does not allow for a high level of specular reflection; absorption can be mistaken for scattering.

The standardized methods for measuring scattering and diffusion coefficients are ISO 17497-1 and ISO 17497-2, respectively.[31,32] The latter provides guidelines for measuring the performance of diffusive surfaces and reporting the diffusion coefficients. The results allow diffusers of different designs to be objectively compared. The results cannot, however, be incorporated into acoustical modeling programs. For that, the scattering coefficients must be used when measured in accordance with ISO 17497-1.

Diffusion and scattering measurement methods are relatively new; ISO 17497-1 was published in 2004 and ISO 17497-2 in 2012. Therefore, very few independent acoustical test laboratories in

North America are currently equipped to perform the standardized scattering and diffusion tests. Because of this, diffusion and scattering coefficients for surfaces and treatments (diffusers or otherwise) are difficult to find. Indeed, because so little testing is being performed on diffusers, there is some degree of confusion in the industry as to what diffusion and scattering coefficients actually mean in subjective terms. For example, what does a diffusion (or scattering) coefficient of 0.84 at 2500Hz sound like? There is no denying that the information is useful and that objective quantification of diffusers is necessary. However, comparing diffusion or scattering coefficients for different materials would be a theoretical exercise at best. The process is further complicated by the fact that commercial diffusers vary dramatically in shape and style; each manufacturer claims some degree of superiority because of some unique application of some innovative mathematics.

Nonetheless, there are diffusion and scattering coefficients available in the literature (Cox and D'Antonio offer a significant amount of laboratory measured diffusion and scattering coefficients)[2], and some manufacturers have begun pursuing independent tests of their diffusive offerings. Fig. 8-28 pictures examples of various commercial diffusers. The next few decades will be a very exciting time for diffusion, particularly if more independent acoustical laboratories begin to offer AES and/or ISO testing services. Additionally, some labs have performed enough testing of diffusion and scattering to begin suggesting improvements to the standard tests.[33]

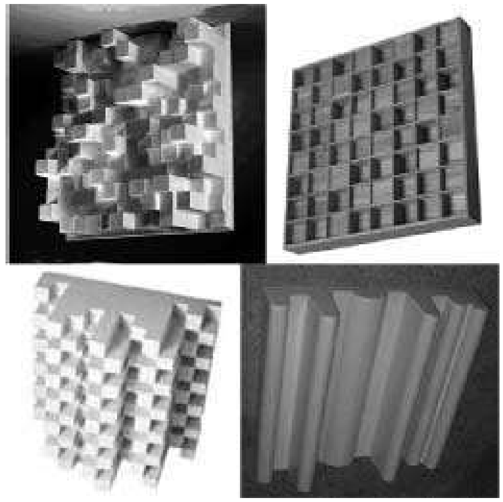


Figure 8-28. Various commercial diffusers.

8.3.2 Mathematical (Numerical) Diffusers

The quadratic residue diffuser (QRD) is one form of a family of diffusers known as reflection phase gratings, or more generally, mathematical or numerical diffusers. Numerical diffusers, such as the QRD, are based on the pioneering work of Manfred Schroeder.[34] Numerical diffusers consist of a periodic series of slots or wells of equal width, with the depth determined by a number theory sequence. The depth sequence is developed via Eq. 8-9

$$\text{well depth factor} = n^2 \bmod p \qquad (8\text{-}9)$$

where,

p is a prime number,

n is an integer ≥ 0.

The "mod" in Eq. 8-9 refers to modulo, which is a number theory mathematical process whereby the first number, in this case the square of n, is divided by the second number, in this case p, and the remainder is equal to the well depth factor. For example, if $n = 5$

and $p = 7$, Eq. 8-9 becomes

$$Well\ depth\ factor = 25 mod\ 7$$

25 divided by 7 = 3 with a remainder of 4.
Therefore:

$$\begin{aligned} well\ depth\ factor &= 25 mod\ 7 \\ &= 4 \end{aligned}$$

In a similar manner, the well depth factors for all the other wells are obtained, as shown in Fig. 8-29. Two complete periods—plus an extra well added to maintain symmetry—are shown in Fig. 8-29 for a $p = 7$QRD. Usually the wells are separated by thin, rigid separators (but not always). An important feature of QRD is symmetry. This allows them to be manufactured and utilized in multiple modular forms.

D'Antonio and Konnert have outlined the theory and application of reflection phase-grating diffusers.[35] The maximum frequency for effective diffusion is determined by the width of the wells; the minimum frequency for effective diffusion is determined by the depth of the well. Commercial diffusers built on these principles are available from a variety of manufacturers. RPG, Inc. and its founder, Dr. Peter D'Antonio, have done pioneering work in the area of diffusion, particularly with respect to Schroeder diffusion and, more recently, with state-of-the-art diffusive surfaces that are customized for an application through the use of special computer models and algorithms.

Numerical diffusers such as QRDs can be 1D or 2D in terms of the diffused sound pattern. QRDs consisting of a series of vertical or horizontal wells will diffuse sound in the horizontal or vertical directions, respectively. In other words, if the wells run in the ceiling-floor direction, diffusion will occur laterally, from side to side, and the resulting diffusion pattern would resemble a cylinder. (Incident sound parallel to the wells will be reflected more than diffused.) More complex numerical diffusers employ sequences of wells—often elevated blocks or square-shaped depressions—that vary in depth (or height) both horizontally and vertically. Incident sound striking these devices will be diffused in a spherical pattern.

8.3.3 Random Diffusion

Besides numerical diffusers, diffusion can also result from the randomization of a surface. In theory, these surfaces cannot provide ideal diffusion. However, listening to the results after treating the surfaces of a room with random diffusers would indicate that, subjectively, they perform quite well. Since any randomization of a surface breaks up specular reflection to some degree, this is not unexpected. The only limitation will be the frequency range of significant diffusion. The rules for well width and depth discussed previously would still apply, albeit in a general sense since the diffusers will not have been designed using a formal number theory algorithm. The benefit of random diffusion is that everyday materials and objects can provide significant diffusion. For example, bookshelves, media storage shelves, decorative trim or plasterwork, furnishings, fixtures, and other decorations can all provide some diffusion. This can be particularly helpful when budget is a concern since diffusive treatments tend to cost twice to ten times more than absorptive treatments on a per square foot basis.

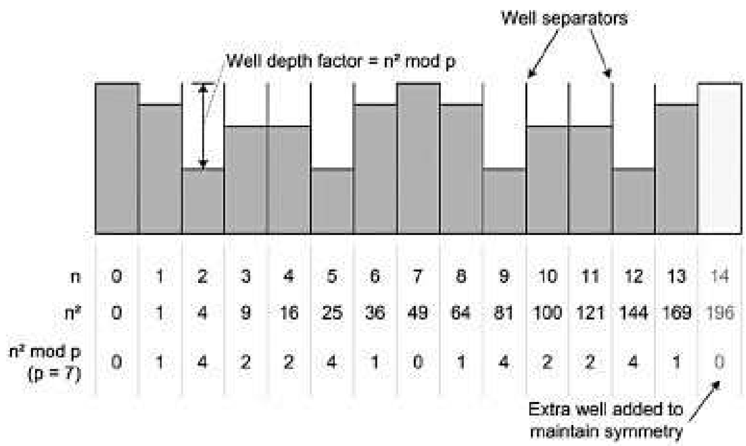


Figure 8-29. A profile of two periods (with one extra well to maintain symmetry) of a QRD of prime number $p = 7$.

8.3.4 Applications of Diffusion

Like absorption, major differences in the size of the space being treated need to be considered when applying diffusion principles. Diffusion tends to provide the most per-dollar-spent benefit in large spaces. Diffusion tends to happen in the far field. In the near field of a diffuser, the scattering effects are typically less pronounced and can actually be less subjectively pleasing than a flat reflective surface in some applications.

For large spaces, diffusion is often employed on the ceiling or rear wall of a space. This provides distribution of the sound energy in the room that envelops a listener. There will be a noticeable (and measurable) reduction in RT, but the reduction is not nearly as severe as it would be with a similar area of absorptive treatment in the same space. Additionally, a typical listener tends not to notice that RT has been reduced, but rather that the decay has been smoothed out, intelligibility has improved, and the room generally sounds better after diffusion has been appropriately applied.

For small spaces, the decision to use diffusion is more challenging. D'Antonio and Cox suggest that the full benefit of diffusers is realized when the listener is a minimum working distance of about 3m (10ft) away from the diffusive surface.[36] This tends to be a good rule-of-thumb from which to start. This rule provides something of a size threshold that must be exceeded to get the most value from the application of diffusers. When small rooms are being treated, it is not uncommon to be able to get far more value from absorption relative to the same area of diffusion.

The most common applications of diffusion in small rooms, such as recording studios and residential theaters or listening rooms, is (like larger spaces) on the rear wall and ceiling. The application of diffusion to the rear wall was particularly popular in the heyday of Live End/Dead End (LEDE) recording studio design. While LEDE is still a popular approach, it has been replaced with the more general reflection-free zone (RFZ) approach to control room design. Regardless of the design method, diffusion can be useful for removing reflective artifacts from a small room without making the room too dead. It can also be used to great effect, for example, on the ceiling of a home theater. A diffusive, rather than absorptive, ceiling can sometimes provide a better feel to a home theater, regardless of the ceiling height.

8.4 Reflection and Other Forms of Sound Redirection

In addition to diffusion, sound can be redirected by controlled reflection, diffraction, or refraction. Diffraction takes place when sound bends around an object, such as when a passing train is audible behind a wall. The low frequencies from the train rumble have large wavelengths relative to the height of the wall, allowing them to bend over the top of it. This can come into play indoors when sound bends around office partitions, podiums, or other common obstacles.

Refraction is the only form of acoustic redirection that does not involve some sort of object. Acoustic refraction is the bending of a sound wave caused by changes in the velocity of sound through a medium. Refraction is often thought of as an optical phenomenon; however, acoustic refraction occurs when there are temperature gradients in a room. Because the speed of sound is dependent on the temperature of the air, when an acoustic wave passes through a temperature gradient, it will bend toward the cooler air. This can occur indoors in large rooms when cooler air from air-conditioning vents located on the ceiling blows into a room with warmer air below; the sound will bend upwards until the temperature reaches equilibrium. Even in recording studios, a heating vent blowing warm air over one loudspeaker in a stereo pair can skew the sound towards the other loudspeaker and wreak havoc on the stereo image!

Finally, everything in the room, including the room itself, reflects sound in some way, even the absorbers. One could look at an absorber as an inefficient reflector. When an item is small with respect to the wavelength of sound impinging upon it, it will have little effect on that sound. A foot stool placed in front of a woofer will have little effect on the 1.7m wavelength of 200Hz, the 4.4m wavelength of 100Hz, or the 6.9m wavelength of 50Hz. The wave will diffract around it and continue on its merry way. The wavelength of 8kHz is 4.3 cm. Just about anything that is placed in front of a tweeter that is reproducing 8kHz and up can effectively block or redirect the sound.

Although it may seem odd, reflection is a very useful form of acoustic treatment, especially in concert halls. Adding reflections to the direct sound is what makes concert halls what they are. However, adding reflections to a monitoring environment can impede critical analysis by coloring the sound coming from the speakers. It is not a question of reflections being good or bad; rather, the designer must decide to include or exclude reflections based on the use of the facility and the desired outcome.

For more information on the design of usefully reflective surfaces for large spaces, see Chapter 9, *Room-acoustical Fundamentals for Auditoriums and Concert Halls*.

8.5 Electronic Treatments

Absorptive acoustical treatments effectively add damping to the room. The reduction of decay is generally the goal. It is a myth that electronics can be used in place of absorptive acoustical treatments. There is no electronic device that can be inserted into the signal path that will prevent sound from a loudspeaker from reflecting off the surfaces of the room. Nonetheless, since the beginning of the electroacoustic era, devices such as electronic absorbers and room equalizers have been proposed. Not all of these are without merit. As early as 1953, Olson and May proposed an electronic sound absorber consisting of a microphone, amplifier, and loudspeaker.[37] Over a short distance from the microphone, the device could be tuned to achieve as much as 10 to 20dB of practical attenuation over a 1 to 2 octave range of low frequencies. Olson and May proposed that their electronic sound absorbers could be used to reduce noise at the ears of airline passengers and factory workers. Unfortunately, the ineffectiveness of this type of absorber over larger distances made it impractical for use in architectural applications. The concept, however, paved the way for future developments.

The invention of the parametric equalizer (PEQ) brought a new wave of hope for electroacoustical treatments. Unfortunately, the insertion of a PEQ into the signal chain, even to reduce narrowband problems in small rooms, usually caused more harm than good. Because of the variability of the sound pressure distribution in a small room, the desired effect of the PEQ was usually limited to a small area of the room. Additionally, phase anomalies usually made the treatment sound unnatural. The use of a PEQ to tune a recording studio control room, for example, came and went quickly and for good reason.

The age of digital signal processing, combined with the availability of high-quality audio equipment to a wider range of users, such as home theater owners, ushered in a new hope for electroacoustical treatments. The most recent devices, while sometimes referred to as room equalization (as in previous decades), are often referred to as digital room correction, or DRC. The most important improvement of these devices over their analog ancestors is their ability to address sound problems occurring in the time/phase domains. The latest in DRC systems are able to address minimum-phase problems, such as axial room modes (see Chapter 6, *Small Room Acoustics*). These problems often manifest themselves not as amplitude problems (which are what would be addressed in the use of analog equalizers), but as decay problems. More modern DRC systems, such as those developed by Wilson et al, that incorporate the latest in digital signal processing, can now actually add the damping that is required to address minimum-phase low-frequency problems.[38] Additionally, many DRC systems require that the room response be measured at multiple listening locations in the room so that algorithms can be used to determine corrections that can benefit a larger area of the room.

The same advances in signal processing have also brought about wider applications for the original electronic sound absorber of Olson and May. Bag End has developed the E-Trap, an electronic bass trap that offers the ability to add significant and measurable damping at two different low frequencies.[39]

While DRC devices and electronic traps offer much in the way of being able to actually address the problems with the loudspeaker-room interface, they cannot be expected to be more than electronic tweaks. They cannot replace a good acoustical room design with proper incorporation of nonelectronic treatments. They can provide some damping, particularly in the lowest octave or two where in many rooms it is often impractical—if not impossible—to incorporate porous or resonant absorbers.

8.6 Acoustical Treatments and Life Safety

The most important consideration when selecting acoustical treatments is safety. Most often, common sense should prevail. For example, asbestos acoustical treatments—which were quite popular several decades ago—should be avoided because of the inherent health risks associated with handling asbestos materials and breathing its fibers. Acoustical treatments will have to meet any applicable building codes and safety standards to be used in a particular facility. Specific installations may also dictate that specific materials be avoided because of allergies or special use of the facility—e.g., health care or correctional facilities. Since many acoustical treatments will be hung from walls and ceilings, only the manufacturer-approved mounting methods should be used to prevent injury from falling objects. The two most common health and safety concerns for acoustical treatment materials are flammability and breathability.

Acoustical treatments must not only meet the applicable fire safety codes, but, in general, should not be flammable. The flammability of an interior finish such as an acoustical treatment is typically tested in accordance with the ASTM E84 standard to measure flammability.[40] The results of the ASTM E84 test are a flame spread index and a smoke developed index. Building codes further classify materials according to the test results. International Building Code (IBC) classifications are as follows:[41]

- Class A: Flame spread index = 0–25,
  smoke developed index = 0–450.
- Class B: Flame spread index = 26–75,
  smoke developed index = 0–450.
- Class C: Flame spread index = 76–200,
  smoke developed index = 0–450.

Materials to be used as interior finishes, such as acoustical treatments, are often tested in accordance with ASTM E84, with test results provided in manufacturer literature. The ASTM E84 test results and corresponding IBC classifications for some typical acoustical materials are summarized in Table 8-4.

Table 8-4. Typical ASTM E84 Test Results and Corresponding IBC Classifications for Common Acoustical Treatments

Acoustical Treatment	Flame Spread Index	Smoke Devel-oped Index	IBC Class	Comments
Glass Fiber panels	15	0	A	Unfaced material
Mineral wool panels	5	10	A	Unfaced material
Wood fiber panels	0	0	A	Unfaced, treated material
Cotton fiber panels	10	20	A	Unfaced, treated material
Acoustical foam panels Polyurethane	35	350	B	Unfaced, treated material NFPA 286 test may also be required
Acoustical foam panels Melamine	5	50	A	Unfaced material NFPA 286 test may also be required
Acoustical plaster	0	0	A	Unfaced material
Acoustical Diffusers Polystyrene	15	145	A	Treated material NFPA 286 test may also be required
Acoustical Diffusers Wood	25	450	A	Treated material

In general, most acoustical materials are Class A materials. Some acoustical foam treatments, as well as some acoustical treatments made of wood, are Class B. Care should be taken that any acoustical treatments made of foam or wood have been tested and that the manufacturer can provide proof of testing. It should also be noted that some jurisdictions require that acoustical foam materials be subjected to more stringent flammability requirements, such as the NFPA 286 test method.[42]

With regards to breathability, precautions should be taken if the acoustical treatment material contains fibers that could be respiratory or skin irritants. The fibers of many common acoustical treatments, such as glass fiber and mineral wool panels, are respiratory and skin irritants, but are harmless once the treatments have been installed in their final configuration, usually with a fabric or other material encasing the fibrous board. Nonetheless, precautions such as wearing gloves and breathing masks should be taken when handling the raw materials or when installing the panels. Additionally, damaged panels should be repaired or replaced in order to minimize the possibility of fibers becoming airborne.

Some facilities may have additional safety requirements. Some health care facilities may disallow porous materials of any kind to minimize the possibility of, for example, mold or bacterial growth. Clean room facilities may also prohibit the use of porous materials on the grounds of minimizing the introduction of airborne particles. Correctional facilities will often prohibit any materials that can be burned (including some fire-resistant materials) and securing acoustical treatment panels to walls or ceilings without any removable mechanical fasteners, such as screwed, rivets, bolts, etc. Still other facilities may have safety requirements based on, for example, the heat produced by a piece of machinery, the chemicals involved in a manufacturing process, and so on. The applicable laws, codes, and regulations—including rules imposed by the end user—should always be consulted prior to the purchase, construction, and installation of acoustical treatments.

8.7 Acoustical Treatments and the Environment

Acoustical treatments should be selected with an appropriate level of environmental awareness. Depending on the application, selection could include not only what the material itself is made of, but also how it is made, how it is transported to the facility, and how it will be disposed of should it be replaced sometime in the future. Many acoustical treatments, such as those consisting of natural wood or cotton fibers, can contribute to Leadership in Energy and Environmental Design (LEED) certification. Unlike audio electronics where overseas manufacturing has become the norm, acoustical treatments are often manufactured and fabricated locally, thus saving on the financial and environmental costs of transportation.

Even acoustical treatments such as polyurethane foam panels, which are a byproduct of the petroleum refining process and can involve the use of carbon dioxide (a greenhouse gas) in the manufacturing process, are becoming more environmentally friendly. For example, one manufacturer of acoustical foam products, Auralex Acoustics, Inc., has begun using soy components in their polyurethane products, thereby reducing the use of carbon-rich petroleum components by as much as 60%.

The best possible environmentally friendly approach to the use of acoustical treatments is to limit their use. The better a facility can be designed from the beginning, the fewer specialty acoustical treatment materials will be required. Rooms from recording studios to cathedrals that are designed with acoustics in mind from the beginning generally require relatively fewer specialty acoustical treatments. Acoustical treatments are difficult to avoid completely; almost every space where the production or reproduction of sound takes place, or where the ability to communicate is tantamount, will require some acoustical treatment. Nonetheless, the most conservative approach to facility design should ensure that only those acoustical treatments that are absolutely necessary are implemented in the final construction.

References

1. Don Davis, *If Bad Sound Were Fatal, Audio Would Be the Leading Cause of Death*, Bloomington, IN: 1st Books, 2004.
2. Trevor J. Cox and Peter D’Antonio, *Acoustic Absorbers and Diffusers:*

Theory, Design and Application, London and New York: Spon Press, 2004.

3. W.C. Sabine, *Collected Papers in Acoustics*, Cambridge: Harvard University Press, 1922.

4. Leo L. Beranek, *Acoustics*, Woodbury, NY: Acoustical Society of America, 1993.

5. *Standard Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method*. ASTM C423-09a.

6. *Acoustics — Measurement of Sound Absorption in a Reverberation Room*. ISO 354:2003.

7. *Standard Practices for Mounting Test Specimens during Sound Absorption Tests*. ASTM E795-05.

8. Peter D'Antonio, "Bass Management," *Diffuse Reflections* newsletter, Vol. 10, issue 1, 2005.

9. Peter D'Antonio, "Bass Management," *Diffuse Reflections* newsletter, Vol. 10, issue 2, 2005.

10. Albert London, "The Determination of Reverberant Sound Absorption Coefficients from Acoustic Impedance Measurements," *Journal of the Acoustical Society of America*, Vol. 22, no. 2, pp. 263–269, March 1950.

11. Brad Nelson, "Single Reflection Analysis in Determining Absorption Coefficients," *Synergetic Audio Concepts Newsletter*, Vol. 8, no. 1, p. 14, Fall 1980. Other comments in Vol. 26, no. 1, Winter 1998.

12. *Acoustics — Sound Absorbers for Use in Buildings — Rating of Sound Absorption*. ISO 11654:1997.

13. Pranab Saha, “Do We Know What We Are Doing?” *Sound and Vibration*, Vol. 35, no. 4, April 2001.

14. John S. Parkinson, “Area and Pattern Effects in the Measurement of Sound Absorption,” *Journal of the Acoustical Society of America*, Vol. 2, no. 1, pp. 112–122, July 1930.

15. Ron Sauro, Michael Vargas, and Gary Mange, “Absorption coefficients-part 1: is square area enough?,” *Proceedings of INTER-NOISE 2009*, October 2009.

16. Ron Sauro, Michael Vargas, and Gary Mange, “Absorption coefficients-part 2: is edge effect more important than expected?,” *Proceedings of INTER-NOISE 2009*, October 2009.

17. *Noise Control Manual*, sixth edition, Publication no. 5-BMG-8277-G, Toledo, OH: Owens Corning Fiberglass Corp. April 1986.

18. F.P. Mechel, “Design Charts for Sound Absorber Layers,” *Journal of the Acoustical Society of America*, Vol. 83, no. 3, pp. 1002–1013, March 1988.

19. Various data sheets available from Roxul®, Inc., http://www.roxul.com/sw34127.asp, accessed March 9, 2008.

20. Data available from Tectum, Inc., http://www.tectum.com/technicaldata/Wall Panel Test Data.pdf, accessed December 23, 2013.

21. Data available from Auralex Acoustics, Inc., http://www.auralex.com/testdata/test/mastertable.pdf, accessed December 23, 2013.

22. Various data sheets available from Pinta Acoustic, Inc., http://www.pinta-acoustic.com/literature.php, accessed December 23, 2013.

23. Cyril M. Harris (Editor in Chief), *Handbook of Acoustical Measurements and Noise Control*, Woodbury, NY: Acoustical Society of America, 1998.

24. V. S. Mankovsky, *Acoustics of Studios and Auditoria*, New York: Focal Press, 1971.

25. Dirk Noy, Gabriel Hauser, and John Storyk, "Low Frequency Absorbers—Applications and Comparisons," *Presented at the 115th Convention of the Audio Engineering Society*, Preprint No. 5944, October 2003.

26. J.S. Bradley, "Sound Absorption of Gypsum Board Cavity Walls," *Journal of the Audio Engineering Society*, Vol. 45, no. 4, pp. 253–259, April 1997.

27. Hermann L.F. Helmholtz, *On the Sensations of Tone*, Mineola, NY: Dover, 1954. (An unaltered, unabridged re-publication of the Ellis translation of the second edition, originally published by Longmans & Co. in 1885.)

28. Data available from Proudfoot, Inc., http://theproudfootcompany.com/Stores/Soundblox/Sound_Absorption_Coefficients accessed December 23, 2013.

29. Data available from RPG, Inc., http://www.rpginc.com/product_Clearsorber_Foil.cfm, accessed December 23, 2013.

30. M. Rettinger, "Low-Frequency Slot Absorbers," *db The Sound Engineering Magazine*, Vol. 10, no. 6, pp. 40–43, June 1976.

31. *Acoustics — Sound-Scattering Properties of Surfaces — Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room*. ISO 17497-1:2004.

32. *Acoustics — Sound-Scattering Properties of Surfaces — Part 2: Measurement of the directional diffusion coefficient in a free field*. ISO 17497-2:2012.

33. Ron Sauro and Michael Vargas, “A suggested method to be used to measure scattering coefficients of full scale samples,” *Proceedings of INTER-NOISE 2009*, October 2009.

34. M.R. Schroeder, “Diffuse Sound Reflection by Maximum-Length Sequences,” *Journal of the Acoustical Society of America*, Vol. 57, no. 1, pp. 149–150, January 1975.

35. Peter D. Antonio and John H. Konnert, “The Reflection Phase Grating Diffusor: Design Theory and Application,” *Journal of the Audio Engineering Society*, Vol. 32, no. 4, pp 228–238, April 1984.

36. Peter D’Antonio and Trevor Cox, “Technical Bulletin on the Application of Diffusion in Critical Listening Rooms,” Upper Marlboro, MD: RPG, Inc., 2003.

37. Harry F. Olson and Everett G. May, “Electronic Sound Absorber,” *Journal of the Acoustical Society of America*, Vol. 25, no. 6, pp. 1130–1136, November 1953.

38. Rhonda J. Wilson, Michael D. Capp, and J. Robert Stuart, “The Loudspeaker-Room Interface—Controlling Excitation of Room Modes,” *Presented at the 23rd International Conference of the Audio Engineering Society*, May 2003.

39. Reza Kashani and James Wischmeyer, “Electronic Bass Trap,” *Presented at the 117th Convention of the Audio Engineering Society*, Preprint No. 6277, October 2004.

40. *Standard Test Method for Surface Burning Characteristics of Building Materials*. ASTM E84-01.

41. *2006 International Building Code®*, International Code Council, 2006.

42. *Standard Methods of Fire Tests for Evaluating Contribution of Wall and Ceiling Interior Finish to Room Fire Growth*. NFPA 286, 2006 edition.

Bibliography

Trevor J. Cox and Peter D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application*, London and New York: Spon Press, 2004.

W.C. Sabine, *Collected Papers in Acoustics*, Cambridge: Harvard University Press, 1922.

Heinrich Kuttruff, *Room Acoustics*, New York: Spon Press, 2000.

C. Zwikker and C.W. Kosten, *Sound Absorbing Materials*, New York: Elsevier Publishing Company, Inc., 1949.

Uno Ingard, *Noise Reduction Technology, Part 1, ABSORBERS*, Kittery Point, ME: http://www.lulu.com/, 1999.

Chapter 9

Room-Acoustical Fundamentals for Auditoriums and Concert Halls

by Dr. Wolfgang Ahnert and Hans-Peter Tennhardt

9.1 Introduction

With musical or spoken performances in auditoriums and concert halls, acoustic evaluation is mainly based on the subjective perception of audience and performers. These judgments are generally not based on defined criteria, but characterize the sensed tonal perception. Besides the secondary factors influencing the overall acoustic impression, like for instance comfortableness of the seats, air conditioning, interference level, optical, architectural and stylistic impression, it is especially the expectation of the listener that plays a significant role for the acoustic evaluation. If a listener in a classical concert is sitting next to the woodwind instruments but hears the brass instruments much louder, even though he cannot see them, his expectations as a listener and thus the acoustics are "off." Numerous subjective and objective room-acoustical criteria were defined and their correlation determined in order to objectify these judgments. However, these individual criteria are closely linked with each other and their acoustic effects can neither be exchanged nor individually altered. They become effective for judgment only in their weighted totality. The judgment of the performers, on the other hand, can be regarded as a kind of "workplace evaluation." Only the musician, singer or speaker who feels completely at ease with all "fringe" factors will also praise the acoustical quality. The main factors judged here are the sound volume and the mutual listening, which is also responsible for the intonation. An acoustically adequate response from the auditorium has to be realized for the performers so that this positive correspondence supports the overall artistic experience. The overall acoustic impression of his own work as it is perceived in the reception area plays a very subordinate role for the performer. What is important for him, however, are rehearsal conditions where the acoustics are as close as possible to those of the actual performance and acoustical criteria that depend as little as possible on the occupation density both in the audience area as well as in the platform area. Generally, a performance room must not show any disturbing reflection characteristics like echo effects or flutter echoes. All seats have to guarantee a good audibility that is in good conformity with the auditory expectation. This requires a balanced sound of high clarity and an adequate spaciousness. Localization shifts or deviations between acoustical and visual directional impression must not occur. If the room is used as a concert hall, the spatial unity between the auditorium and the platform areas has to be maintained in order to avoid sound distortions.

Based on these considerations and well-founded, objective measurement technical examinations and subjective tests, partially in reverberation-free rooms within artificially generated sound fields, it was possible to define room-acoustical quality criteria that enable an optimum listening and acoustical experience in dependence on the usage function of the room. The wider the spectrum of usage is, the broader is the limit of the desirable reference value ranges of these criteria. Without extensive variable acoustical measures, also electronic ones, only a compromise brings about a somewhat satisfactory solution. It stands to reason that this compromise can only be as good as the degree in which the room-acoustical requirements coincide with it.

A precondition for an optimum room-acoustical design of auditoriums and concert halls is the very early coordination in the planning phase. The basis here is the establishment of the room's primary structure according to its intended use (room shape, volume, topography of the spectators and the platform areas). The secondary structure that decides the design of the structures on the walls and ceilings as well as their acoustic effectiveness has to be worked out on this basis. A planning methodology for guaranteeing the room-acoustical functional and quality assurance of first-class concert halls and auditoriums as well as rooms with a complicated primary structure is reflected in the application of simulation tests by means of mathematical and physical models (see section 9.3 and Chapter 39 *Computer aided Sound System Design*).

9.2 Room-Acoustical Criteria, Requirements

The acoustical evaluation by listeners and actors of the acoustical playback-quality of a signal that is emitted from a natural acoustic source or via electroacoustical devices, is mostly very imprecise. This evaluation is influenced by existing objective causes like disturbing climatic, seating, and visibility conditions as well as by subjective circumstances like for instance the subjective attitude and receptiveness towards the content and the antecedents of the performance. Very differentiated is the subjective rating of music, where the term “good acoustics” is defined, depending on the genre, as a sufficient sound volume, a good time and register clarity of the sound, and a spaciousness that meets the composition. Timbre changes that deviate from the “natural” timbre of the acoustic sources and from the usual distance dependence (high frequency sounds are less effective at a larger distance from the place of performance than at closer range) are judged as being unnatural, if traditional music is concerned. These experiences determine also the listening expectation for a very spatial and reverberating sound in a large cathedral, whereas one expects a “dry” sound in the open. Thus deviations from this experience are regarded as being bothersome. A listener seated in the front section of a concert hall expects a clearer sound than one seated in a rear section. On the other hand, however, he wants to enjoy an optimally balanced acoustic pattern on all seats, as he has grown up with the media and mainly post-processed sound productions that are independent of the room, and thus acquired auditory expectations, which do not allow an evaluation of the objectively existing room.

The evaluation of speech is generally a bit easier, since optimum audibility and clear intelligibility are desired herein an atmosphere that is not influenced by the room or electro-acoustical means. Perhaps with the exception of sacral rooms, the spaciousness generally does not play such an outstanding role in this regard, whereas sound volume and intelligibility are all the more important.

Numerous room-acoustical criteria were defined in order to clarify the terms applied for the subjective and objective assessment of a spoken or musical performance. In the following we have listed a relevant selection of them, in which context one should note that there is a close correlation between the individual criteria. One single optimally determined parameter may not at all be acoustically satisfactory, because another parameter influences the judgment in a negative way. For example, the optimum value range of center time and definition can only be evaluated with a subjectively correct estimated reverberation time. The guide values of the reverberance measure are valid only if the clarity measure C_{80} is in the optimal range.

On principle, the room-acoustical quality criteria can be subdivided into time and energy criteria. The main type of use, speech or music, then determines the recommendations for the guide values to be targeted. With multi-purpose halls (without available variable measures for changing the acoustics), a compromise is required that should orient itself on the main type of use.

9.2.1 Time Criteria

9.2.1.1 Reverberation Time (RT)

The reverberation time (RT) is not only the oldest, but also the most best-known room-acoustical quantity. It is the time that passes after an acoustic source in a room has been turned off until the mean steady-state sound-energy density $w(t)$ has decreased to 1/1,000,000 of the initial value w_0 or until the sound pressure has decayed to 1/1.000—i. e., by 60dB

$$w(RT) = 10^{-6} w_0 \qquad (9\text{-}1)$$

Thus the time response of the sound energy density in reverberation[1] results as

$$w(t) = w_0^{-6\log\frac{t}{RT}} = w_0^{-13.82\frac{t}{RT}} \qquad (9\text{-}2)$$

The steady-state condition is reached only after the starting time t_{st} of the even sound distribution in a room (approximately 20 sound reflections within 10ms) after the starting time of the excitation[2]

$$t_{st} = 1\ldots\, 2(0.17\ldots\, 0.34)\sqrt{V} \qquad (9\text{-}3)$$

where,

t_{st} is in ms,

V is in m^3 (ft^3).

The defined drop of the sound pressure level of 60dB corresponds roughly to the dynamic range of a large orchestra.[3] The listener, however, can follow the decay process only until the noise level in the room becomes perceptible. This subjectively assessed parameter *reverberation time duration* thus depends on the excitation level as well as on the noise level.

The required evaluation dynamic range is difficult to achieve even with objective measuring, especially in the low frequency range. Therefore, the reverberation time is determined by measuring the sound level decay in a range from −5dB to −35dB and then defined as $RT_{30\ dB}$ (also RT_{30}). The initial reverberation time (IRT according to Atal,[2] $RT_{15\ dB}$ between −5dB and −20dB) and the early decay time (EDT according to Jordan,[2] $RT_{10\ dB}$ between 0dB and −10dB) are mostly more in conformity with the subjective assessment of the duration of reverberation, especially at low level volumes. This also explains the fact that the reverberation time subjectively perceived in the room may vary, while the values measured objectively according to the classical definition with a dynamic range of 60dB or 30dB are, except admissible fluctuations, generally independent of the location.

Serving as a single indicator for the principal characterization of the room in an occupied or unoccupied state, the reverberation time is used as the mean value between the two octave bandwidths 500Hz and 1000Hz or the four 1/3-octave bandwidths 500Hz, 630Hz, 800Hz, and 1000Hz, and referred to as the *mean reverberation time*.

The desirable convenient value of the reverberation time *RT* depends on the kind of performance (speech or music) and the size of the room. For auditoriums and concert halls, the desired values of the mean reverberation time for between 500Hz and 1000Hz with a room occupation of between 80% and 100% are given in Fig. 9-1 and the admissible frequency tolerance ranges are shown in Figs. 9-2 and 9-3. This shows that in order to guarantee a specific warmth of sound with musical performances, an increase of the reverberation time in the low frequency range is admissible (see section 9.2.1.2), while with spoken performances a decrease of the reverberation time is desirable in this frequency range (see section 9.2.2.9).

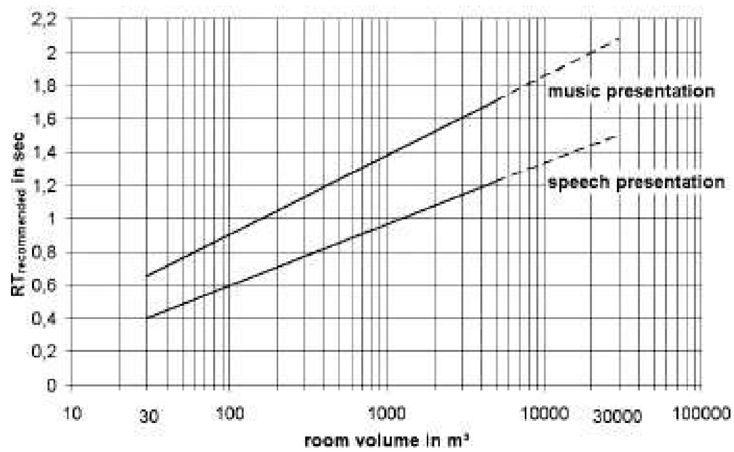


Figure 9-1. Recommended value of the mean reverberation time $RT_{recommended}$ between 500Hz and 1000Hz for speech and music presentations as a function of room volume *V*.

The reverberation time of a room as defined by EYRING mainly depends on the size of the room and on the sound absorbing properties of the boundary surfaces and non-surface forming furnishings:

$$RT_{60} = 0.163^{*} \frac{V}{-\ln(1+\bar{a})S_{tot} + 4mV} \qquad (9\text{-}4)$$

$^{*}0.049$ for U.S. units

where,

RT is the reverberation time in s,

V is the room volume in m^3 (ft^3),

$\bar{a}$ is A_{tot}/S_{tot} which is the Room-averaged coefficient of absorption,

A_{tot} is the total absorption surface in m^2 (ft^2),

S_{tot} is the total room surface in m^2 (ft^2),

m is the energy attenuation factor of the air in m^{-1}, see Fig. 9-4.

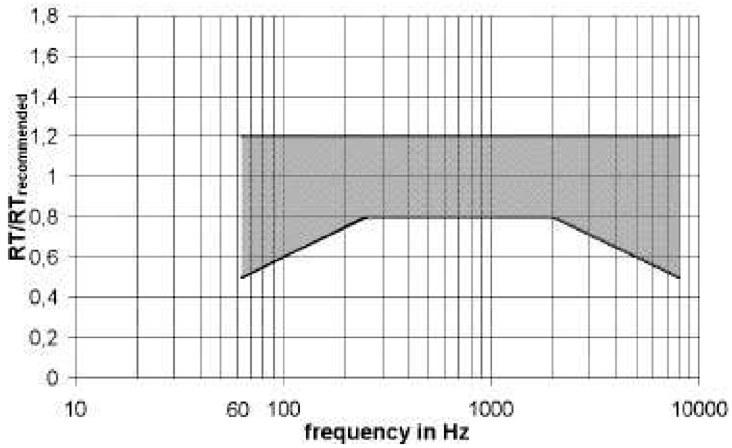


Figure 9-2. Frequency dependent tolerance range of reverberation time *RT* referred to $RT_{recommeded}$ for speech presentations.

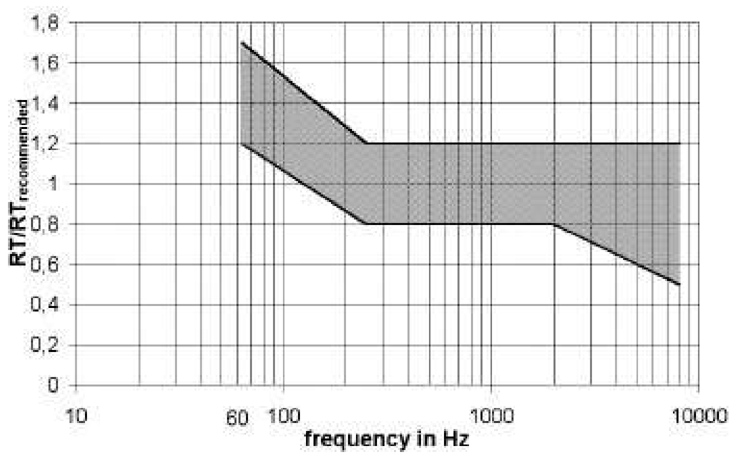


Figure 9-3. Frequency-dependent tolerance range of reverberation time *RT* referred to $RT_{recommeded}$ for music presentations.

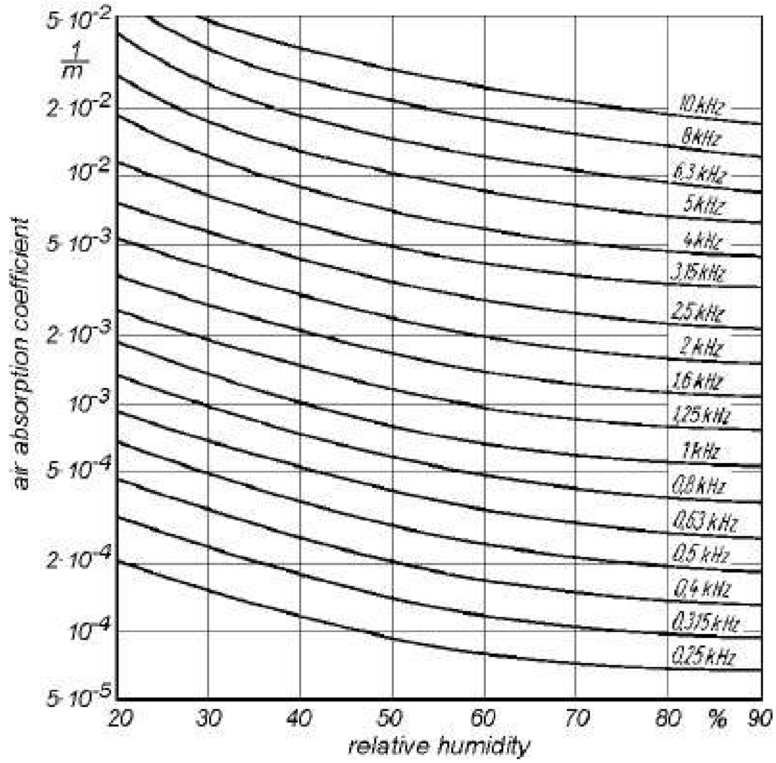


Figure 9-4. Air absorption coefficient m as a function of relative humidity F.

The correlation between the mean sound absorption coefficient and the reverberation time for different relations between room volume V and total surface S_{tot} is graphically shown in Fig. 9-5.

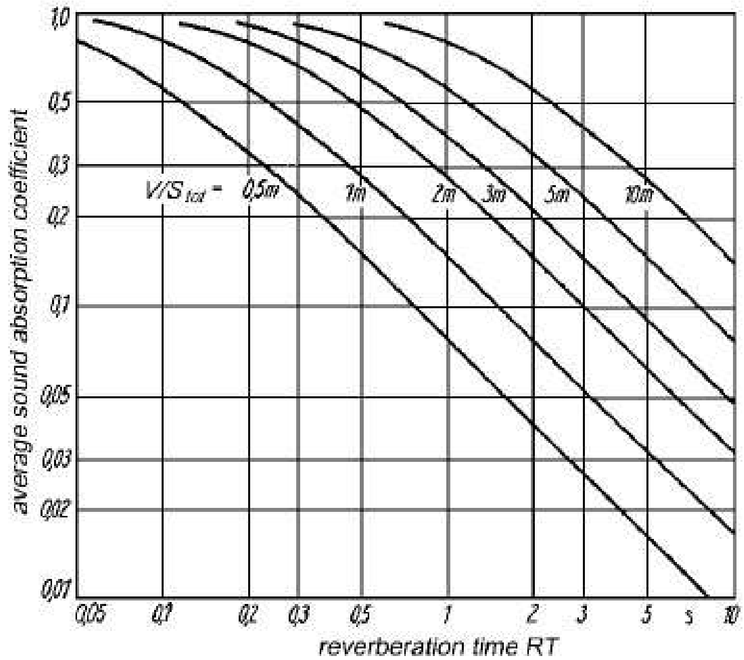


Figure 9-5. Correlation between average sound absorption coefficient and reverberation time for various ratios of room volume V and room surface S_{tot}.

The total sound absorption surface of the room A_{tot} consists of the planar absorption surfaces with the corresponding partial surfaces S_n and the corresponding frequency-depending coefficient of sound absorption α_n plus the non-surface forming absorption surfaces A_k consisting, e.g., of the audience and the furnishings.

$$A_{tot} = \sum_n \alpha_n S_n + \sum_k A_k \qquad (9\text{-}5)$$

For an average sound absorption coefficient of up to $\bar{a}$ = 0.25, the Eq. 9-4 by EYRING[2] can be simplified by means of series expansion according to Sabine[4] to

$$RT = 0.163^* \frac{V}{A_{tot} + 4mV} \qquad (9\text{-}6)$$

*0.049 for U.S. units

where,

RT is the reverberation time in s,

V is the room volume in m^3 (ft^3),

A_{tot} is the total absorption surface in m^2 (ft^2),

m is the energy attenuation factor of the air in m^{-1}, see Fig. 9-4.

The correlation between the reverberation time RT, the room volume V, the equivalent sound absorption surface A_{tot}, and the unavoidable air damping m is graphically shown in Fig. 9-6.

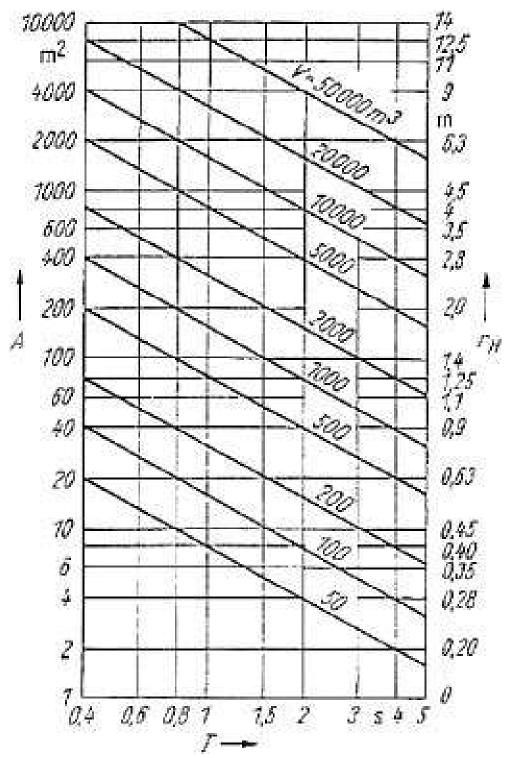


Figure 9-6. Correlation between reverberation time RT, room volume V, and equivalent sound absorption area A, according to Eq. 9-7.

The above stated frequency-dependent sound absorption coefficient has to be determined by measuring or calculation of the diffuse all-round sound incidence. Measurement is generally done in the reverberation room by using Eq. 9-6. If the sound absorption coefficient is measured by using an impedance tube (or Kundt's tube) with vertical sound incidence, the results can only be converted to the diffuse sound incidence by means of the diagrams of Morse and Bolt,[2] if one can assume that the complex input impedance of the absorber is independent of the angle, i.e., if the lateral sound propagation is inhibited in the absorber (e.g., porous material with a high specific flow resistance).

Properly speaking, the above-mentioned derivatives of the reverberation time from the sound absorption in the room are only valid for approximately cube-shaped rooms with an even distribution of the sound absorbing surfaces within the room. With room shapes deviating heavily from a square or a rectangle, or in case of a necessary one-sided layout of the absorbing audience area, these factors also have a decisive effect on the reverberation time. With the same room volume and the same equivalent sound absorption surface in the room, inclining the side wall surfaces towards the room's ceiling or towards the sound absorbing audience area results in deviations of the measured reverberation time of up to 100%. For numerous room shapes there exist calculating methods with different degrees of exactness, so for example, for cylinder-shaped rooms.[5] The cause of these differences lies mainly with the geometrical conditions of the room and their influence on the resulting path length of the sound rays determining the reverberation.

The absorbed sound power P_{ab} of a room can be derived from the ratio energy density W = sound energy W/volume V under consideration of the differential coefficient $P_a b = dW/dt$ representing the rate of energy decay in the room and taken from

Eqs. 9-5 and 9-6.

$$P_{ab} = \frac{1}{4}cwA \tag{9-7}$$

where,

c is the sound velocity.

In steady-state, the absorbed sound power is equal to the Power P fed into the room. This results in the *average sound energy density* w_r in the diffuse sound field of the room as

$$w_r = \frac{4P}{cA} \tag{9-8}$$

While the sound energy density w_r in the diffuse sound field is approximately constant, the direct sound energy and thus also its density w_d decreases at close range to the source with the square of the distance r from the source, according to

$$w_d = \frac{P}{c}\frac{1}{4\pi r^2} \tag{9-9}$$

Strictly speaking, this is valid only for spherical acoustic sources[6] given a sufficient distance it can be applied, however, to most practically effective acoustic sources.

For the sound pressure in this range of predominantly direct sound, this results in a decline with $1/r$. (Strictly speaking, this decline sets in only outside of an interference zone, the near field. The range of this near field is of the order of the dimensions of the source and 0.4m away from its center.)

If the direct sound and the diffuse sound energy densities are equal ($w_d = w_r$), Eqs. 9-8 and 9-9 can be equated, that means it is possible to determine a specific distance from the source, the reverberation radius (critical distance for omnidirectional sources) r_H. With a spherical acoustic source there is

$$
\begin{aligned}
r_H &= (0.3^*)\sqrt{\frac{A}{16\pi}} \\
&\approx (0.3^*)\sqrt{\frac{A}{50}} \\
&\approx 0.041(0.043^*)\sqrt{A} \\
&\approx 0.057(0.01^*)\sqrt{\frac{V}{RT}}
\end{aligned}
\qquad (9\text{-}10)
$$

* for U. S. units

where,

r_H is in m (ft),

A is in m^2 (ft^2),

V is in cubic m^3 (ft^3),

RT is in s.

With a directional acoustic source (speaker, sound transducer), this distance is replaced by the critical distance r_R

$$r_R = \Gamma(\vartheta)\sqrt{\gamma(r_H)} \qquad (9\text{-}11)$$

where,

$\Gamma(\vartheta)$ is the angular directivity ratio of the acoustic source (the ratio between the sound pressure that is radiated at the angle ϑ against the reference axis and the sound pressure that is generated on the reference axis at the same distance, in other words, the polars, γ is the front-to-random factor of the acoustic source.)

9.2.1.2 Bass Ratio (BR) (Beranek)

Besides the reverberation time *RT* at medium frequencies, the frequency response of the reverberation time is of great importance, especially at low frequencies, as compared to the medium ones. The bass ratio, i. e., the ratio between the reverberation times at octave center frequencies of 125Hz and 250Hz and octave center frequencies of 500Hz and 1000Hz (average reverberation time) is calculated basing on the following relation:[7]

$$BR = \frac{RT_{125\,\mathrm{Hz}} + RT_{250\,\mathrm{Hz}}}{RT_{500\,\mathrm{Hz}} + RT_{1000\,\mathrm{Hz}}} \qquad (9\text{-}12)$$

For music, the desirable bass ratio is $BR \approx 1.0–1.3$. For speech, on the other hand, the bass ratio should at most have a value of $BR \approx$ 0.9... 1.0.

9.2.2 Energy Criteria

According to the laws of system theory, a room can be acoustically regarded as a linear transmission system that can be fully described through its impulse response h(t) in the time domain. If the unit impulse δ(t) is used as an input signal, the impulse response is linked with the transmission function in the frequency domain through the Fourier transform

$$\underline{G}(\omega) = F\{h(t)\} \quad (9\text{-}13)$$

where,

$$h(t) = F^{-1}\{\underline{G}(\omega)\}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \underline{G}(\omega)^{(j\omega t)d\omega}$$

As regards the measuring technique, the room to be examined is excited with a very short impulse (delta unit impulse) and the impulse response *h(t)* is determined at defined locations in the room, see Fig. 9-7.

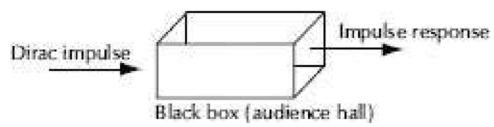


Figure 9-7. Basic solution of signal theory for identification of an unknown room.

Here, the impulse response contains the same information as a quasi-statically measured transmission frequency response.

Generally, the time responses of the following sound-field-proportionate factors (so-called reflectograms) are derived from measured or calculated room impulse responses $h(t)$

Sound pressure: $p(t) \approx h(t)$ (9-14)

Sound energy density: $w(t) \approx h^2(t)$ (9-15)

Ear-inertia weighted: $J_{\tau_0}(t) \approx \int_0^t h^2(t')^{\left(\frac{t'-t}{\tau_0}\right)} \mathrm{d}t'$ (9-16)

where,

τ_0 is 35ms.

$$\text{Sound energy: } W(t) \approx \int_0^{\tau} h^2(t')\mathrm{d}t' \qquad (9\text{-}17)$$

Basic reflectogram figures are graphically shown in Fig. 9-8.

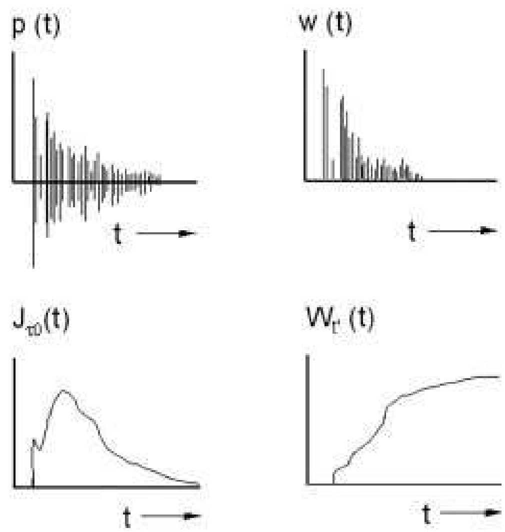


Figure 9-8. Behavior of sound field quantity versus time (reflectograms) for sound pressure $p(t)$, sound energy density $w(t)$, ear inertia weighted sound intensity $J_{\tau0}(t)$ and sound energy $W_{t'}$(t).

In order to simplify the mathematical and measuring-technical correlations, a sound-energy-proportional factor is defined as *sound energy component* E_f. Being a proportionality factor of the sound energy, this factor shows the dimension of an acoustical impedance and is calculated from the sound pressure response $p(t)$.

Sound energy component:

$$E_{t'} = \int_0^{t'} p^2(t)\,dt \qquad (9\text{-}18)$$

where,

t' is in ms.

For determining a sound-volume-equivalent energy component, t' has to be set to equal ∞. In practical rooms of medium size, $t' \approx 800$ms is sufficient.

For measuring of all speech-relevant room-acoustical criteria, an acoustic source with the frequency-dependent directivity pattern of a human speaker has on principle to be used for exciting the sound field, while with musical performances it is sufficient to use a non-directional acoustic source for the first approximation.

The majority of the room-acoustical quality criteria is based on the monaural, directionally unweighted assessment of the impulse response. Head-related binaural criteria are still an exception. The influence of the sound-incidence direction of early initial reflections on room-acoustical quality criteria is principally known. Since subjective evaluation criteria are still missing to a large extent however, this may also be generally disregarded when measuring or calculating room impulse responses. For the determination of most of the relevant criteria, the energetic addition of the two ear signals of an artificial head is sufficient here.

Just like the directional dependence, the frequency dependence of the room-acoustical energy criteria has also not been researched in depth, so that it is generally sufficient at the moment to evaluate the octave with the center frequency of 1000Hz.

9.2.2.1 Strength Measure (G) (P. Lehmann)

The strength measure G is the ten-fold logarithmic ratio between the sound energy components at the measuring location and those measured at 10m distance from the same acoustic source in the free field. It characterizes the volume level

$$G = 10\log\left(\frac{E_{\infty}}{E_{\infty,10\,\mathrm{m}}}\right)\,\mathrm{dB} \qquad (9\text{-}19)$$

Here, $E_{\infty,10m}$ is the reference sound energy component existing at 10 m (32.8ft) distance with the free sound transmission of the acoustic source.

According to reference 8 or 9, the optimum values for musical and speech performance rooms are located between +1dB ≤ G ≤ +10 dB which means that the loudness at any given listener's seat in real rooms should be roughly equal to or twice as high as in the open at 10m (32.8ft) distance from the sound source.

9.2.2.2 Sound Pressure Distribution (Δ L)

The decrease of sound pressure level ΔL in dB describes the distribution of the volume at different reception positions in comparison with a reference measuring position or also for a specific measuring position on the stage in comparison with others. If the sound energy component at the reference measuring position or for the reference measuring position on stage is labeled with E_0 and at the reception measuring position or for the measuring position on stage with E, one calculates a sound pressure level distribution ΔL

$$\Delta L = 10\log\left(\frac{E_{\infty}}{E_{\infty,0}}\right)\,\mathrm{dB} \qquad (9\text{-}20)$$

It is of advantage for a room if ΔL for speech and music is in a range of 0dB ≥ ΔL ≥ −5dB.

9.2.2.3 Interaural Cross-Correlation Coefficient IACC

The IACC is a binaural, head-related criterion and serves for describing the equality of the two ear signals between two freely selectable temporal limits t_1 and t_2. In this respect, however, the selection of these temporal limits, the frequency evaluation as well as the subjective statement are not clarified yet. In general, one can examine the signal identity for the initial reflections (t_1 = 0ms, t_2 = 80ms) or for the reverberation component ($t_1 \geq t_{st}$, $t_2 \geq RT$ [see section 9.2.1.1]). The frequency filtration should generally take place in octave bandwidths of between 125Hz and 4000Hz.

According to reference 10 or 11 the standard interaural cross-correlation function IACF is defined as

$$IACF_{t_1 t_2}(\tau) = \frac{\int_{t_1}^{t_2} p_L(t) \times p_R(t+\tau)\,dt}{\left[\int_{t_1}^{t_2} p_L^{\,2}(t)\,dt \times \int_{t_1}^{t_2} p_R^{\,2}(t)\,dt\right]^{1/2}} \qquad (9\text{-}21)$$

where,

$p_L(t)$ is the impulse response at the entrance to the left auditory canal,

$p_R\,(t)$ is the impulse response at the entrance to the right auditory canal.

Then the interaural cross-correlation coefficient IACC is

$$IACC_{t_1 t_2} = max\left|IACF_{t_1 t_2}(\tau)\right|$$
$$\text{for } -1 \text{ ms} < \tau < +1 \text{ ms}$$

9.2.2.4 Center Time (t_s) (Kürer)

For music and speech performances, the center time t_s is a reference value for spatial impression and clarity and results at a measuring position from the ratio between the summed-up products of the energy components of the arriving sound reflections and the corresponding delay times and the total energy component. It corresponds to the instant of the first moment in the squared impulse response and is thus determined according to the following ratio:

$$t_s = \frac{\sum_i t_i E_i}{E_{ges}} \qquad (9\text{-}22)$$

The higher the center time t_s is, the more spatial is the acoustic impression at the listener's position. The maximum achievable center time t_s is based on the optimum reverberation time. According to Hoffmeier[12], there is a good correlation between center time and intelligibility of speech with a frequency evaluation of four octaves between 500Hz, 1000Hz, 2000Hz, and 4000Hz.
For music, the desirable center time t_s is:
$t_s \approx 70...150$ms with a 1000Hz octave,
For speech:
$t_s \approx 60...80$ms with four octaves between 500Hz and 4000Hz.

9.2.2.5 Echo Criterion (EK) (Dietsch)

If we look at the build-up function of the center time $t_s(\tau)$:

$$t_s(\tau) = \frac{\int_0^\tau |p(t)|^n t \cdot dt}{\int_0^\tau |p(t)|^n \cdot dt} \qquad (9\text{-}23)$$

using as exponents for the incoming sound reflection $n = 0.67$ with speech and $n = 1$ with music and comparing it with the difference quotient

$$EK(\tau) = \frac{\Delta t_s(\tau)}{\Delta \tau_E} \qquad (9\text{-}24)$$

we can discern echo distortions for music or speech when applying values of $\Delta\tau_E = 14$ms for music and $\Delta\tau_E = 9$ms for speech, ascertained by subjective tests.[13] The echo criterion depends on the motif. With fast and accentuated speech or music, the limit values are lower.

For 50% (EK_{50}%) and 10% (EK_{10}%), respectively, of the listeners perceiving this echo, the limit values of the echo criterion amount to:

- Echo perceptible with music for $EK_{50}\% \geq 1.8$; $EK_{10}\% \geq 1.5$ for two octave bands 1kHz and 2kHz mid frequencies.
- Echo perceptible with speech for $EK_{50\%} \geq 1.0$; $EK_{10\%} \geq 0.9$ for one octave band 1kHz.

9.2.2.6 Definition Measure C_{50} for Speech (Ahnert)

The *definition measure* C_{50} describes the intelligibility of speech and also of singing. It is generally calculated in a bandwidth of four octaves between 500Hz and 4000Hz from the tenfold logarithm of the ratio between the sound energy arriving at a reception measuring position up to a delay time of 50ms after the arrival of the direct sound and the following energy:

$$C_{50} = 10\lg\left(\frac{E_{50}}{E_\infty - E_{50}}\right)\,\mathrm{dB} \qquad (9\text{-}25)$$

A good intelligibility of speech is generally given when $C_{50} \geq 0$dB.

The frequency-dependent definition measure C_{50} should increase by approximately 5dB with octave center frequencies over 1000Hz (starting with the octave center frequencies 2000Hz, 4000Hz, and 8000Hz), and decrease by this value with octave center frequencies below 1000Hz (octave center frequencies 500Hz, 250Hz, and 12Hz).

According to Höhne and Schroth[14] the limits of the perception of the difference of the definition measure are at $\Delta C50 \approx \pm 2.5$dB.

An equivalent, albeit less used criterion is the degree of definition D, also called D_{50}, that results from the ratio between the sound energy arriving at the reception measuring position up to a delay time of 50ms after the arrival of the direct sound and the entire energy (given in %) is

$$D = \frac{E_{50}}{E_\infty} \qquad (9\text{-}26)$$

The correlation with the definition measure C_{50} is determined by the equation

$$C_{50} = 10\lg\left(\frac{D_{50}}{1 - D_{50}}\right)\,\mathrm{dB} \qquad (9\text{-}27)$$

One should thus strive for an intelligibility of syllables of at least 85%, $D = D_{50} \geq 0.5$, or 50%.

9.2.2.7 *Speech Transmission Index (STI) (Houtgast, Steeneken)*

The determination of the STI-values is based on measuring the reduction of the signal modulation between the location of the sound source, e.g., on stage—and the reception measuring position with octave center frequencies of 125Hz up to 8000Hz. Here STEENEKEN and Houtgast[15] have proposed to excite the room or open space to be measured with a special modulated noise and then to determine the reduced modulation depth.

The authors proceeded on the assumption that not only reverberation and noise reduce the intelligibility of speech, but generally all external signals or signal changes that occur on the path from source to listener. For ascertaining this influence they employ the *modulation transmission function* (MTF) for acoustical purposes. The available useful signal S (signal) is put into relation with the prevailing interfering signal N (noise). Thus the determined *modulation reduction factor* $m(F)$ is a factor that characterizes the interference with speech intelligibility

$$m(F) = \frac{1}{\sqrt{1 + \left(\frac{2\pi F \cdot RT}{13.8}\right)^2}} \cdot \frac{1}{1 + 10^{-\left(\frac{S/N}{10\ \mathrm{dB}}\right)}} \qquad (9\text{-}28)$$

where,

F is the modulation frequency in Hz,

RT_{60} is the reverberation time in s,

S/N is the signal/noise ratio in dB.

To this effect one uses modulation frequencies from 0.63 Hz to 12.5Hz in third octaves. In addition, the modulation transmission function is subjected to a frequency weighting (WMTF - weighted modulation transmission function), in order to achieve a complete correlation to speech intelligibility. In doing so, the modulation transmission function is divided into 7 octave bands, which are each modulated with the modulation frequency (14). This results in a matrix of 7 × 14 = 98 modulation reduction factors, m_i.

The (apparent) effective signal-noise ratio X can be calculated from the modulation reduction factors m_i

$$X_i = 10\log\left(\frac{m_i}{1-m_i}\right)\,\text{dB} \qquad (9\text{-}29)$$

These values will be averaged and for the seven octave bands the Modulation Transfer Indices $MTI = (X_{average} + 15)/30$, are calculated. After a frequency weighting in the seven bands (partially separated for male or female speech) you obtain the Speech Transmission Index *STI*.

The excitation of the sound field is done by means of a sound source having the directivity behavior of a human speaker's mouth.

In order to render 20 years ago this relatively time consuming procedure in real-time operation, the RASTI-procedure (rapid speech transmission index) was developed from it in cooperation with the company Brüel & Kjaer.[16] The modulation transmission function is calculated here only for two octave bands (500Hz and 2kHz) that are especially important for the intelligibility of speech and for select modulation frequencies, i.e., in all for only nine modulation reduction factors *mi*. However, this measure is used increasingly less.

Note: Schroeder could show that the 98 modulation reduction factors $m(F)$ may also be derived from a measured impulse response

$$m(F) = \frac{\int_0^\infty h^2(t) e^{-j2\pi F t} dt}{\int_0^\infty h^2(t) dt} \qquad (9\text{-}30)$$

This is done now with modern computer-based measurement routines like MLSSA, EASERA, or Win-MLS.

A new method to estimate the speech intelligibility measures an impulse response and derives STI values with the excitation with a modulated noise. The frequency spectrum of this excitation noise is shown in Fig. 9-9.

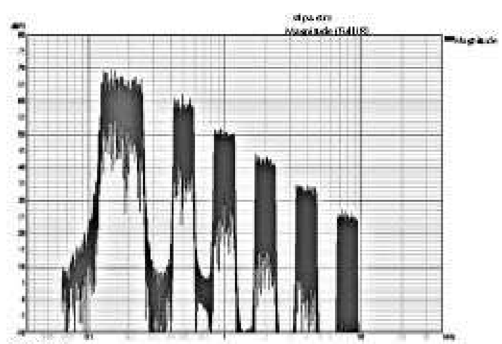


Figure 9-9. STIPa signal in frequency presentation.

You recognize 1/2-octave band noise, radiated through the sound system into the room. By means of a mobile receiver at any receiver location the STIPa values can be determined.[8,9] Any layman may use this method and no special knowledge is needed. It is used more and more to verify the quality of emergency call systems (EN 60849),[33] especially in airports, stations or large malls.

According to definition the STI-value is calculated by using the results of Eq. 9-29

$$STI = \frac{X + 15}{30} \qquad (9\text{-}31)$$

Based on the comparison of subjective examination results with a maximum possible intelligibility of syllables of 96%, the STI values are graded in subjective values for syllable intelligibility according to Table 9-1 (EN ISO 9921: Feb. 2004).

Table 9-1. Subjective Weighting for STI

Subjective Intelligibility	STI Value
unsatisfactory	0.00-0.30
poor	0.30-0.45
satisfactory	0.45-0.60
good	0.60-0.75
excellent	0.75-1.00

9.2.2.8 Articulation Loss, Alcons, with Speech (Peutz, Klein)

Peutz[17] and Klein[18] have ascertained that the articulation loss of spoken consonants *Alcons* is decisive for the evaluation of speech intelligibility in rooms. Starting from this discovery they developed a criterion for the determination of intelligibility:

$$Alcons \approx 0.652\left(\frac{r_{LH}}{r_H}\right)^2 RT\ \% \qquad (9\text{-}32)$$

where,

r_{LH} is the distance sound source-listener,

r_H is the reverberation radius or, in case of directional sound sources, critical distance r_R,

RT_{60} is the reverberation time in s.

From the measured room impulse response one can determine *Alcons* according to Peutz,[17] if for the direct sound energy one applies the energy after about 25ms to 40ms (default 35ms), and for the reverberation energy the residual energy after 35ms

$$Alcons \approx 0.652\left(\frac{E_\infty - E_{35}}{E_{35}}\right) \cdot RT\ \% \qquad (9\text{-}33)$$

Assigning the results to speech intelligibility yields Table 9-2.

Table 9-2. Subjective Weighting for *Alcons*

Subjective Intelligibility	Alcons
Ideal intelligibility	≤3%
Good intelligibility	3-8%
Satisfactory intelligibility	8-11%
Poor intelligibility	>11%
Worthless intelligibility	>20% (limit value 15%)

Long reverberation times entail an increased articulation loss. With the corresponding duration, this reverberation acts like noise on the following signals and thus reduces the intelligibility.

Fig. 9-10 shows the articulation loss *Alcons* as a function of the *S/N* ratio and the reverberation time RT. The top diagram allows us to ascertain the influence of the difference L_R (diffuse sound level) – L_N (noise level) and of the reverberation time *RT* on the *Alcons* value, which gives $ALcons_{R/N}$. Depending on how large the signal to noise ratio ($L_D - L_{RN}$) is, the value is then corrected in the bottom diagram in order to obtain $Alcons_{D/R/N}$. The noise and the signal level have to be entered as dB(A) values.

The illustration shows also that with an increase of the S/N ratio to more than 25 dB, it is practically no longer possible to achieve an improved intelligibility. (In praxis, this value is often even considerably lower, since with high volumes, for example above 90dB, and due to the heavy impedance changes in the middle ear that set on here as well as through the strong bass emphasis that occurs owing to the frequency-dependent ear sensitivity.)

9.2.2.9 Subjective Intelligibility Tests

A subjective evaluation method for speech intelligibility consists in the recognizability of clearly spoken pronounced words (so-called test words) chosen on the basis of the word-frequency dictionary and a language-relevant phoneme distribution. In German intelligibility test *logatoms* (monosyllable consonant-vowel-groups that do not readily make sense, so that a logical supplementation of logatoms that were not clearly understood during the test is not possible, e.g., "grirk," "spres") are used for exciting the room. In English-speaking countries, however, test words as shown in Table 9-3 are used.[19] Per test there are between 200 and 1000 words to be used. The ratio between correctly understood words (or logatoms or sentences) and the total number read yields the word or syllable or sentence intelligibility V rated in percentages. The intelligibility of words V_W and the intelligibility of sentences V_S can be derived from Fig. 9-11.

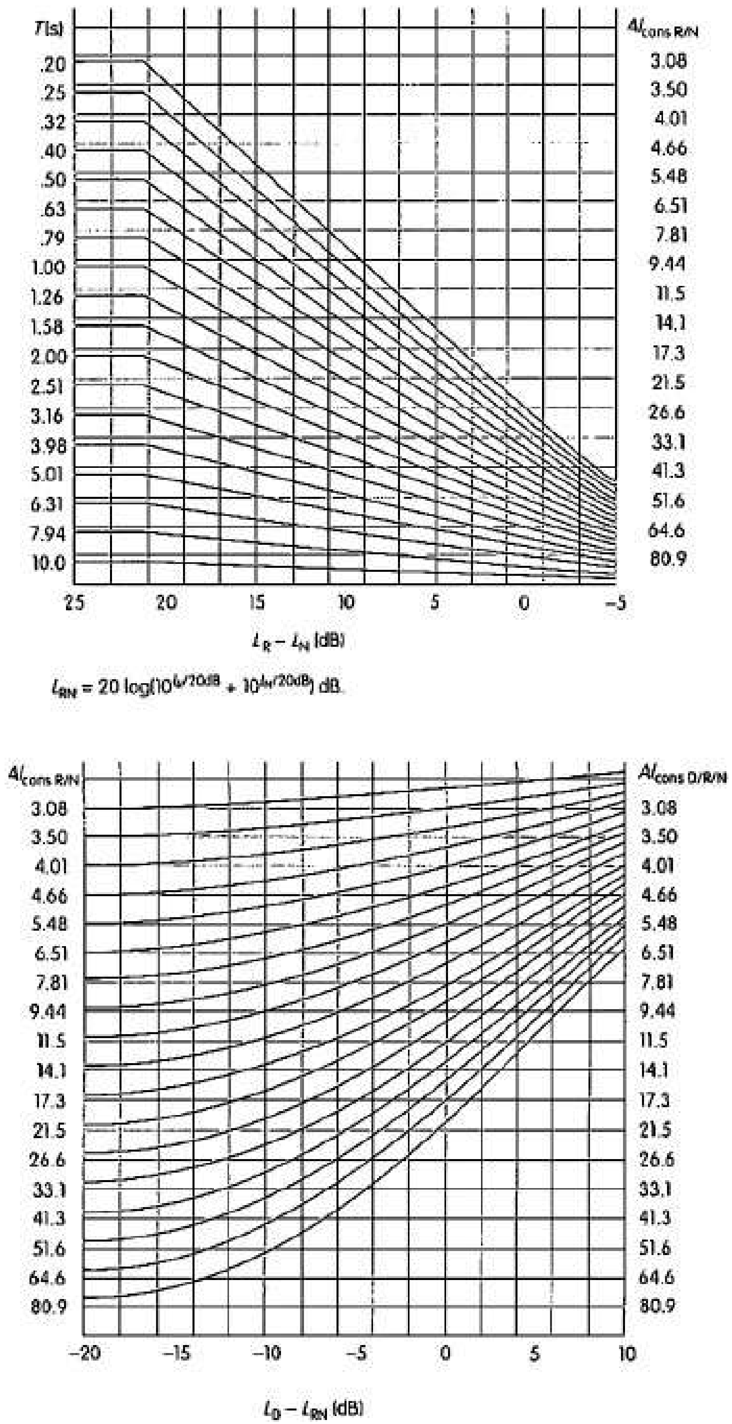


Figure 9-10. Articulation loss Alcons as a function of the level ratio between diffuse sound L_R and direct-sound level L_D, reverberation time R_T and noise level L_N.

Table 9-4 shows the correlation between the intelligibility values and the ratings.

The results of the subjective intelligibility test are greatly influenced by speech velocity which includes the number of spoken syllables or words within the articulation time (articulation velocity) and the break time. Therefore so-called predictor sentences are often used to precede the words or logatoms that are not part of the test. These sentences consist of three to four syllables each, for example: “Mark the word...”, “Please write down...,” “We’re going to write....” Additionally to providing a continuous flow of speech, this also serves for guaranteeing that the evaluation takes place in an approximately steady-state condition of the room.

Table 9-3. Examples of English Words Used in Intelligibility Tests

aisle	done	jam	ram	tame
barb	dub	law	ring	toil
barge	feed	lawn	rip	ton
bark	feet	lisle	rub	trill
baste	file	live	run	tub
bead	five	loon	sale	vouch
beige	foil	loop	same	vow
boil	fume	mess	shod	whack
choke	fuse	met	shop	wham
chore	get	neat	should	woe
cod	good	need	shrill	woke
coil	guess	oil	sip	would
coon	hews	ouch	skill	yaw
coop	hive	paw	soil	yawn
cop	hod	pawn	soon	yes
couch	hood	pews	soot	yet
could	hop	poke	soup	zing
cow	how	pour	spill	zip
dale	huge	pure	still	
dame	jack	rack	tale	

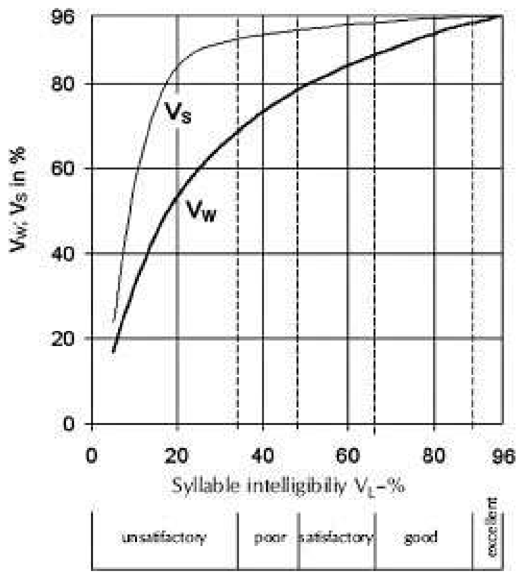


Figure 9-11. Assessment of the quality of speech intelligibility as a function of syllable intelligibility V_I, word intelligibility V_W, and sentence intelligibility V_S.

Table 9-4. Correlation Between the Intelligibility Values and the Ratings

Rating	Syllable Intelligibility V_L in %	Sentence Intelligibility V_S in %	Word Intelligibility V_W in %
Excellent	90–96	96	94–96
Good	67–90	95–96	87–94
Satisfactory	48–67	92–95	78–87
Poor	34–48	89–92	67 –78
Unsatisfactory	0–34	0–89	0–67

There is a close correlation between the subjectively ascertained syllable intelligibility and room-acoustical criteria. According to reference 20 for example, a long reverberation time reduces the syllable intelligibility, Fig. 9-12, owing to the occurrence of masking effects, despite an increase in loudness, see Eq. 9-8.

Quite recently, comprehensive examinations concerning the frequency-dependence of speech-weighting room-acoustical criteria were conducted in order to find the influence of spatial sound coloration.[12] It was ascertained that with broadband frequency weighting between 20Hz and 20kHz the definition measure C_{50} (see section 9.2.2.6) correlates very insufficiently with the syllable intelligibility. Through a frequency evaluation across three to four octaves around a center frequency of 1000Hz, however, the influence of the sound coloration can sufficiently be taken into account. Even better results regarding the subjective weightings are provided by the frequency analysis, if the following frequency responses occur, Fig. 9-13.

As the definition declines with rising frequency due to sound coloration, the intelligibility of speech is also low (bad intelligibility → 3). This includes also the definition responses versus frequency with a maximum value at 1000Hz, poor intelligibility → 4 in Fig. 9-13.

The definition responses versus rising frequency to be aimed at for room-acoustical planning should either be constant (good intelligibility → 1) or increasing (very good intelligibility → 2). With regard to auditory psychology, this result is supported by the importance for speech intelligibility of the consonants situated in this higher frequency range.

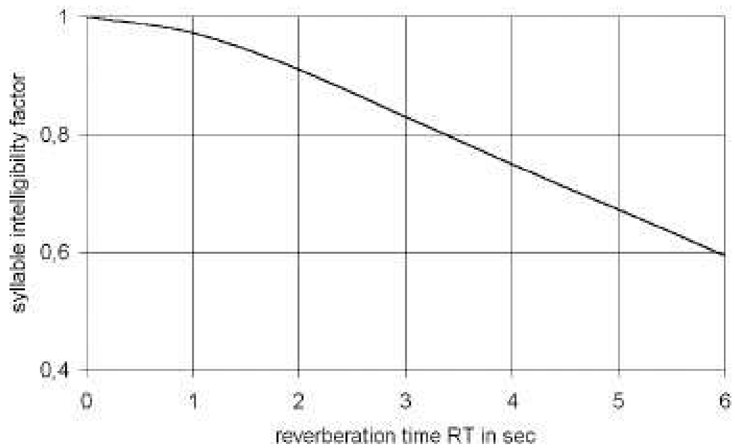


Figure 9-12. Syllable intelligibility factor as a function of reverberation time.

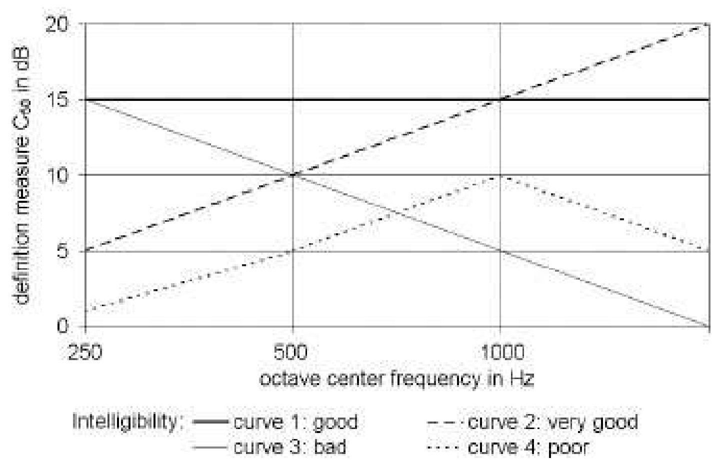


Figure 9-13. Correlation between attainable intelligibility and frequency dependence of the definition measure C_{50}.

The determination of speech intelligibility through the definition measure C_{50} can easily lead to faulty results as the mathematical integration limit of 50ms is not a jump function with regards to intelligibility without knowledge of the surrounding sound reflection distribution.

The best correlation with the influence of the spatial sound coloration exists between the subjective speech intelligibility and the center time t_S (see section 9.2.2.4) with a frequency weighting between the octave of 500Hz to the octave of 4000Hz. According to Hoffmeier,[12] the syllable intelligibility V measured at the point of detection is then calculated as

$$V = 0.96 \cdot V_{sp} \cdot V_{S/R} \cdot V_R \qquad (9\text{-}34)$$

where,

V_{sp} is the influence factor of the sound source (trained speaker V_{sp} = 1, untrained speaker $V_{sp} \approx 0.9$),

$V_{S/N}$ is the influence factor of the useful level (speech level) L_x and of the disturbance level L_{st} according to Fig. 9-14,[6]

$$V_R = -\left(6 \cdot 10^{-6}\left(\frac{t_s}{ms}\right)^2\right) - 0.0012\left(\frac{t_s}{ms}\right) + 1.0488$$

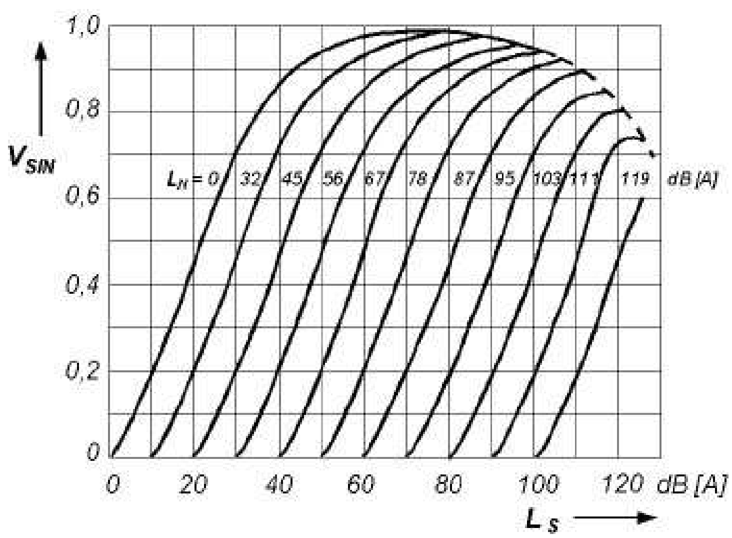


Figure 9-14. Syllable intelligibility factor $V_{S/N}$ as a function of speech sound pressure level L_S and noise pressure level L_N.

The correlation shown in Fig. 9-15 can also be derived between articulation loss and syllable intelligibility. For low reverberation times, syllable intelligibility is almost independent of the articulation loss. An inverse correlation behavior sets in only with increasing reverberation time. It is evident that with the usual *Alcons* values between 1% and 50%, syllable intelligibility can take on values between 68% and 93% (meaning a variation of 25%) and that for an articulation loss < 15% (the limit value of acceptable intelligibility) the syllable intelligibility V_S reaches always, independently of the reverberation time, values over 75% which corresponds roughly to a definition measure of C_{50}> −4dB.

This correlation can also be seen in Fig. 9-16, which shows the correlation between measured RASTI-values and articulation loss *Alcons*. One sees that acceptable articulation losses of *Alcons* <15% require RASTI values in the range from 0.4 to 1 (meaning between satisfactory and excellent intelligibility). Via the equation

$$RASTI = 0.9482 - 0.1845 \ln(Alcons) \qquad (9\text{-}35)$$

it is also possible to establish an analytical correlation between the two quantities. In good approximation this relationship may be used not only for RASTI but for STI as well.

9.2.2.10 Clarity Measure (C_{80}) for Music (Abdel Alim)

The clarity measure (C_{80}) describes the temporal transparency of musical performances (defined for an octave center frequency of 1000Hz) and is calculated from the tenfold logarithm of the ratio between the sound energy arriving at a reception measuring position up to 80ms after the arrival of the direct sound and the following sound energy.[21]

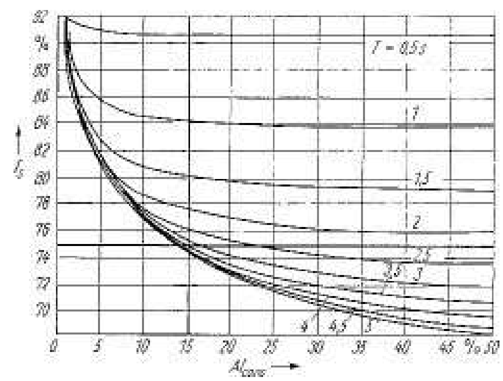


Figure 9-15. Syllable intelligibility I_S as a function of the articulation loss *Alcons*. Parameter: reverberation time *RT*. Preconditions: approximate statistical reverberation behavior; signal-to-noise ratio (*S/N*) ≥ 25dB.

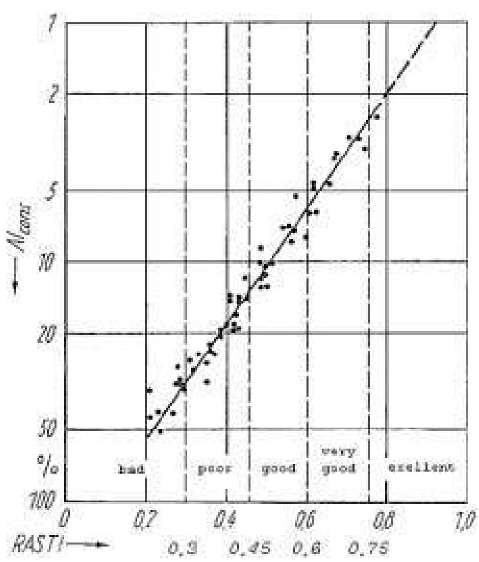


Figure 9-16. Relationship between Alcons values and *RASTI* values.

$$C_{80} = 10\log\left(\frac{E_{80}}{E_{\infty} - E_{80}}\right)\text{dB} \qquad (9\text{-}36)$$

The value for a good clarity measure C_{80} depends strongly on the musical genre. For romantic music, a range of approximately −3dB ≤ C80 ≤ + 4dB is regarded as being good, whereas classic and modern music will allow values up to + 6 to +8dB. According to Höhne and Schroth,[14] the perception limit of clarity measure differences is about $\Delta C80 \approx \pm 3.0$dB.

According to Reichardt et al.,[22] there is an analytical correlation between the clarity measure C_{80} and the center time t_S, as given by

$$\begin{aligned} C_{80} &= 10.83 - (0.95 \cdot t_S) \\ t_S &= 114 - 10.53 \cdot C_{80} \end{aligned} \qquad (9\text{-}37)$$

where,

C_{80} is in dB,

t_S is in ms.

This correlation is graphically depicted in Fig. 9-17.

Figure 9-17. Center time t_S as a function of the clarity measure C80.

9.2.2.11 Sound Coloration Measures (KT) and (KH) for Music (Schmidt)

The sound coloration measures[23] evaluate the volume-equivalent energy fractions of the room impulse response of low and high frequency components (K_T, octave around 100Hz and K_H, octave around 3150Hz, respectively) related to a medium frequency range in an octave bandwidth of 500Hz.

$$K_T = 10\log\left(\frac{E_{\infty,\ 100\ \mathrm{Hz}}}{E_{\infty,\ 500\ \mathrm{Hz}}}\right)\ \mathrm{dB} \qquad (9\text{-}38)$$

$$K_H = 10\log\left(\frac{E_{\infty,\ 3150\ \mathrm{Hz}}}{E_{\infty,\ 500\ \mathrm{Hz}}}\right)\ \mathrm{dB} \qquad (9\text{-}39)$$

The measures correlate with the subjective impression of the spectral sound coloration conditioned by the acoustical room characteristics. Optimum values are $K_{T;H}$ = −3 to +3dB.

9.2.2.12 Spatial Impression Measure (R) for Music (U. Lehmann)

The spatial impression measure *R*[24], [25] consists of the two components spaciousness and reverberance. The spaciousness is based on the ability of the listener to ascertain through more or less defined localization that a part of the arriving direct sound reaches him not only as direct sound from the sound source, but also as reflected sound from the room's boundary surfaces (the perception of "envelopment" in music). The reverberance is generated by the non-stationary character of the music that constantly generates build-up and decaying processes in the room. As regards auditory perception, it is mainly the decaying process that becomes effective as reverberation. Both components are not consciously perceived separately, their mutual influencing of the room is very differentiated.[26] Among the energy fractions of the sound field that increase the spatial impression are the sound reflections arriving after 80 ms from all directions of the room as well as sound reflections between 25ms and 80ms, that are geometrically situated outside a conical window of ±40°, whose axis is formed between the location of the listener and the center of the sound source. Thus all sound reflections up to 25ms and the ones from the front of the above-mentioned conical window have a diminishing effect on the spatial impression of the room. The tenfold logarithm of this relation is then defined as the spatial impression measure R in dB.

$$R = 10\log\left[\frac{(E_\infty - E_{25}) - (E_{80R} - E_{25R})}{E_{25} + (E_{80R} - E_{25R})}\right]\text{dB} \qquad (9\text{-}40)$$

where,

E_R is the sound energy fraction measured with a directional microphone (beaming angle ±40° at 500Hz to 1000Hz, aimed at the sound source).

One achieves a mean (favorable) room impression if the spatial impression measure R is within a range of approximately $-5\text{dB} \le R \le +1\text{dB}$.

Spatial impression measures below −5dB down to −10dB are referred to as being less spatial, others between +1dB up to +7dB as very spatial.

9.2.2.13 Lateral Efficiency (LE) for Music (Jordan), (LF) (Barron) and (LFC) (Kleiner)

For the subjective assessment of the apparent extension of a musical sound source, e.g., on stage, the early sound reflections arriving at a listener's seat from the side are of eminent importance, as compared with all other directions. Therefore the ratio between the laterally arriving sound energy components and those arriving from all sides, each within a time of up to 80ms, is determined and its tenfold logarithm calculated therefrom.

If one multiplies the arriving sound reflections with $\cos^2\sigma$, being the angle between the direction of the sound source and that of the arriving sound wave, one achieves the more important evaluation of the lateral reflections. With measurements this angle-dependent evaluation is achieved by employing a microphone with bi-directional characteristics. Lateral Efficiency, *LE*, is

$$LE = \frac{E_{80Bi} - E_{25Bi}}{E_{80}} \qquad (9\text{-}41)$$

where,

E_{Bi} is the sound energy component measured with a bidirectional microphone (gradient microphone).

The higher the lateral efficiency, the acoustically broader the sound source appears. It is of advantage if the lateral efficiency is within the range of $0.3 \leq LE \leq 0.8$.

For obtaining a uniform representation of the energy measures in room acoustics, these can also be defined as lateral efficiency measure $10\log LE$. Then the favorable range is between $-5\text{dB} \leq 10\log LE \leq -1\text{dB}$.

According to Barron it is the sound reflections arriving from the side at a listener’s position within a time window from 5ms to 80ms that are responsible for the acoustically perceived extension of the musical sound source (contrary to Jordan who considers a time window from 25 ms to 80 ms). This is caused by a different evaluation of the effect of the lateral reflections between 5ms and 25ms.

The ratio between these sound energy components is then a measure for the lateral fraction *LF*:

$$LF = \frac{E_{80Bi} - E_{5Bi}}{E_{80}} \qquad (9\text{-}42)$$

where,

E_{Bi} is the sound energy component measured with a bidirectional microphone (gradient microphone).

It is an advantage if LF is within the range of $0.10 \leq LF \leq 0.25$, or, with the logarithmic representation of the lateral fraction measure $10\log LF$, within $-10\text{dB} \leq 10\log LF \leq -6\text{dB}$.

Both lateral efficiencies LE and LF have in common that, thanks to using a gradient microphone, the resulting contribution of a single sound reflection to the lateral sound energy behaves like the square of the cosine of the reflection incidence angle, referred to the axis of the highest microphone sensibility.[27] Kleiner defines, therefore, the lateral efficiency coefficient LFC in better accordance with the subjective evaluation, whereby the contributions of the sound reflections vary like the cosine of the angle.

$$LFC = \frac{\int_{5}^{80} \left| p_{Bi}(t) \cdot p(t) \right| \mathrm{dt}}{E_{80}} \qquad (9\text{-}43)$$

9.2.2.14 Reverberance Measure (H) (Beranek)

The reverberance measure describes the reverberance and the spatial impression of musical performances. It is calculated for the octave of 1000Hz from the tenfold logarithm of the ratio between the sound energy component arriving at the reception measuring position as from 50ms after the arrival of the direct sound and the energy component that arrives at the reception position within

50ms.

$$H = 10\log\left(\frac{E_\infty - E_{50}}{E_{50}}\right) \text{dB} \qquad (9\text{-}44)$$

In contrast to the definition measure C_{50} an omnisource is used during the measurements of the reverberance measure *H*.

Under the prerequisite that the clarity measure is within the optimum range, one can define a guide value range of 0 dB ≤ *H* ≤ + 4dB for concert halls, and of −2dB ≤ *H* ≤ + 4dB for musical theaters with optional use for concerts. A mean spatial impression is achieved if the reverberation factor *H* is within a range of −5dB ≤ *H* ≤ + 2dB.

Schmidt[2] examined the correlation between the reverberance measure *H* and the subjectively perceived reverberation time RT_{sub}, see Fig. 9-18. For a reverberance measure *H* = 0dB, the subjectively perceived reverberation time coincides with the objectively measured reverberation time.

9.2.2.15 Register Balance Measure (B_R) (Tennhardt)

With musical performances, the relation of the partial volumes of individual orchestra instrument groups between each other and to the singer is an important quality criterion for the balance (register balance) and is defined by the frequency-dependent time structure of the sound field.[28] The register balance measure *BR* between two orchestra instrument groups *x* and *y* is calculated from the A-frequency weighted volume-equivalent sound energy components of these two groups, corrected by the reference balance measure B_{xy} of optimum balance.

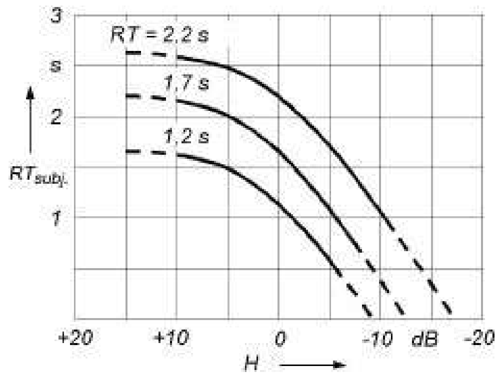


Figure 9-18. Subjectively perceived reverberation time RT_{subj} as a function of reverberance measure H and objective reverberation time RT as a parameter.

$$B_{Rxy} = 10\log\left(\frac{E_{\infty x}}{E_{\infty y}}\right) \mathrm{dB(A)} + B_{xy} \qquad (9\text{-}45)$$

where,

B_{xy}, is in dBA.

		Group x				
		A	B	C	D	S
	A	–	−5.8	1.5	0	−2.8
Group y	B	5.8	–	7.3	5.8	3.0
	C	−1.5	−7.3	–	−1.5	−4.3
	D	0	−5.8	1.5	–	−2.8
	S	2.8	−3.0	4.3	2.8	–

Group A: String instruments.

Group B: Woodwind instruments.

Group C: Brass instruments.

Group D: Bass instruments.

Group S: Singers.

Significant differences in balance do not occur if 4dB(A) < B_R <

−4dB(A) and if this tendency occurs binaurally.

9.3 Planning Fundamentals

9.3.1 Introduction

When planning acoustical projects one has to start out from the fundamental utilization concept envisaged for the rooms. In this respect one distinguishes between rooms intended merely for speech presentation, rooms to be used exclusively for music performances and a wide range of multi-purpose rooms.

In the following we are going to point out the most important design criteria with the most important parameters placed in front. Whenever necessary the special features of the different utilization profiles will be particularly referred to.

Strictly speaking, acoustical planning is required for all rooms as well as for open-air facilities, only the scope and the nature of the measures to be taken vary from case to case. The primordial task of the acoustician should, therefore, consist in discussing the utilization profile of the room with the building owner and the architect, but not without taking into consideration that this profile may change in the course of utilization, so that an experienced acoustician should by no means fail to pay due attention to the modern trends as well as to the utilization purposes which may arise from or have already arisen from within the environs of the new building or the facility to be refurbished, respectively.

On the one hand it is certainly not sensible for a small town to try to style the acoustical quality of a hall to that of a pure concert hall, if this type of event will perhaps take place no more than ten times a year in the hall to be built. In this case a multi-purpose hall whose acoustical properties enable symphonic concerts to be performed in high quality is certainly a reasonable solution, all the more if measures of variable acoustics and of the so-called “electronic architecture” are included in the project.

In rooms lacking any acoustical conditioning whatsoever, on the other hand, many types of events can be performed only with certain reservations, which have to be declined from the acoustical point of view.

Table 9-5 shows the interrelation between utilization profile and effort in acoustical measures. These measures can be characterized as follows:

Auditoriums, congress centers are mainly used for speech presentation. They are mostly equipped with a sound reinforcement system, but may sometimes also do without it. Music performances without a sound reinforcement system take place in a reduced style as a setting for ceremonial acts and festivities. Owing to the short reverberation time abiding their utilization concept, larger concert performances mostly require in such rooms the room-acoustical support of electro-acoustical equipment (see Chapter 39 *Computer aided Sound System Design*, section 39.1).

Table 9-5. Interrelation Between Utilization Profile and Acoustical Measures

Utilization Profile	Scope and Quality of the Acoustical Measures				
	Very High	High	Medium	Low	Very Low
Pure concert hall	x				
Pure opera house	x				
Multi-genre theater		x			
Multi-functional hall, also for modern music		x			
Open air theater				x	
Club and bar areas, jazz clubs				x	
Auditoriums for speech			x		
Lecture and classrooms					x

Spoken-drama theaters serve in their classical form for speech presentation with occasional accompaniment by natural music instruments and vocalists. Apart from serving as a support for solo instruments in a music performance, utilization of electroacoustical systems is reserved almost exclusively for playing-in effects or for mutual hearing.

Multi-genre theaters are gaining in the theater scene an ever-growing importance as against the pure music or spoken-drama theater. The presentation of speech or music from natural sources must be possible here without compromise. While the classical music or spoken-drama theater got along with an average reverberation time of about 1s, the trend in the planning of modern multi-genre theaters tends to a somewhat longer reverberation time of up to 1.7s with a strong portion of definition-enhancing initial sound energy and a reduced reverberance measure (less energy at the listener seat after 50ms than within the first 50 ms). Here it may also be appropriate to make use of variable acoustics for reverberation time reduction, if, e.g., electro-acoustical performances (shows, pop concerts, etc.) are presented. This reverberation time reduction should be obtained by shortening the travel paths of the sound reflections rather than by sound absorption measures which tend to reduce loudness. The separation of room volumes (e.g., seats on the upper circle, reverberation chambers) leads mostly to undesirable timbre changes, unless these volumes are carefully dimensioned.

Electro-acoustical systems have in multi-genre theaters mostly mutual hearing and playing-in functions. Concert presentations on the stage with natural sound sources require the additional installation of a concert enclosure.

Opera houses having a large classical theater hall must be capable of transmitting speech and music presentations of natural sources in excellent acoustic quality without taking recourse to sound reinforcement. Speech is mainly delivered by singing. In room-acoustical planning of modern opera houses the parameters are therefore chosen so as to be more in line with musical requirements (longer average reverberation time of up to 1.8s, greater spaciousness, spatial and acoustical integration of the orchestra pit in the auditorium). Electro-acoustical means are used for reproducing all kinds of effects signals and for playing-in functions (e.g., remote choir or remote orchestra). This implies that the sound reinforcement system is becoming more and more an artistic instrument of the production director.

Concert presentations on the stage with natural sound sources also require the additional installation of a concert enclosure which has to form a unity with the auditorium as regards proper sound mixing and irradiation.

Multipurpose halls cover the widest utilization scope ranging from sports events to concerts. This implies that variable natural acoustics are not efficient as a planning concept, since the expenditure for structural elements generally exceeds the achievable benefit. Parting from a room-acoustical compromise solution tuned to the main intended use, with a somewhat shorter reverberation time and consequently high definition and clarity, appropriately built-in structural elements (enclosures) have to provide for the proper sound mixing required for concerts with natural music instruments, while prolongation of reverberation time as well as enhancement of spatial impression and loudness can be achieved by means of electroacoustical systems of "electronic architecture", see section 9.4.

To a greater extent sound systems are used here to cover speech amplification and the needs of modern rock and pop concerts.

Classical concert halls serve first of all for music events ranging from soloist presentations to the great symphony concert with or without choir. They are mostly equipped with a pipe organ and their room-acoustical parameters must satisfy the highest quality demands. Electro-acoustical systems are used for vocal information and for mutual listening with special compositions, but are generally still ruled out for influencing the overall room-acoustical parameters.

General concert halls can be used for numerous music performances, among others for popular or pop concerts. Here it is the use of an electro-acoustical sound reinforcement system that overrides the room-acoustical parameters of the hall. In accordance with the variety of events to be covered by these halls, they should be tuned to a frequency-independent reverberation time of the order of 1.2s and feature high clarity.

Sports halls, Gymnasiums have to provide an acoustical support for the mutual emotional experience. This concerns first of all the supporting acoustic correspondence of the spectators between themselves and with the performers. Thus there are only few sound absorbing materials to be used in the spectators' areas and sound reflecting elements to be provided towards the playing field. The ceiling of the playing field should be more heavily damped so as to enable it to be used also for music events, in which case an electro-acoustical sound reinforcement system is to be used. The same applies to open and partially or fully covered stadiums where sound absorption above the playing field is a natural feature with the open ones.

Show theaters are generally used only on negligibly few occasions with natural acoustics, an exception being "Singspiel" theaters with orchestra pit. Predominantly, however, an electro-acoustical sound reinforcement system is used for the functions of play-in and mutual hearing as well as half or full playback. The room-acoustical parameters of the theater room have with this form of utilization to comply with the electro-acoustical requirements. The reverberation time should therefore not exceed a frequency-independent value of 1.4 s and the sound field should have a high diffusivity so that the electro-acoustically generated sound pattern does not get distorted by the acoustics of the room.

Rooms with variable acoustics controlled by mechanical means show some positive result only in a certain frequency range, if corresponding geometric modifications of the room become simultaneously visible. The room-acoustical parameters have always to coincide with the listening experience that means they must also be perceived in a room-size and room-shape-related manner. Experimental rooms and effect realization (e.g., in a virtual stage setting of a show theater) are, of course, excluded from this mode of consideration. In theater rooms and multi-purpose halls it is possible to vary the reverberation time by mechanical means within a range of about 0.5 s without detrimental effect on spatial impression and timbre. At any rate one should abstain from continuously variable acoustic parameters "house superintendent acoustics," since possible intermediate steps could lead to uncontrolled and undesirable acoustic settings.

Sacral rooms. Here we have to distinguish between classical church rooms and contemporary modern sacral buildings. With the classical rooms it is their size and importance that determine their room-acoustical parameters, e.g., a long reverberation time and an extreme spaciousness. Short reverberation times sound inadequate in such an environment. The resulting deficiency in definition, inconvenient, e.g., during the sermon, has to be compensated by providing additional initial reflections through architecturally configured reflectors or nowadays mostly through an electroacoustical sound system. With music presentations one has, in various frequency domains, to adapt the style of playing to the long decay time (cf. Baroque and Romanesque churches). Electroacoustical means can serve here only for providing loudness.

From the acoustical point of view, modern church buildings acquire to an increasing degree the character of multi-purpose halls. Thanks to appropriately adapted acoustics and the use of sound reinforcement systems they are not only adequate for holding religious services, but can also be used as venues for concerts and conferences in good quality.

9.3.2 Structuring the Room Acoustic Planning Work

9.3.2.1 General Structure

The aim of room-acoustical planning consists in safeguarding the acoustical functionality under the envisaged utilization concepts of the auditorium for the performers as well as for the audience. With new buildings such details should be considered already in the planning phase, whereas with already existing rooms an appropriate debugging should be an essential part of the refurbishment. Point of departure in this respect is a purposeful influencing control of the *primary structure* of the performance room. This concerns, among other things:

- The size of the room.
- The shape of the room.
- Functional-technological circumstances, for instance the platform or stage arrangement, the installation of balconies or galleries, lighting installations, and the arrangement of multi-media equipment.
- The topography regarding the arrangement of performers and listeners, like for instance the sloping of tiers or the proscenium area in front of the stage opening.

Basing on these premises, the *secondary structure* of the room will be acoustically determined. This structure concerns essentially:

- The arrangement and distribution of frequency-dependent sound-absorbing as well as sound-reflecting faces.

- The subdivision of the surface structure for directional and diffuse sound reflections.
- The frequency-dependent effect of uneven surfaces.
- The architectural-stylistic conformation of all boundary surfaces of the room.

9.3.2.2 Room Form and Sound Form

There exists a correspondence between the shape of a room (room form) in its primary structure and the resulting sound. The term "sound form" refers in this context to the reverberation timbre which is herewith divided into its low-frequency portion (warmth) and its high-frequency portion (brilliance).

The method used for assessing the acoustical quality of concert halls is based on a paper by Beranek[46] in which one finds a list of 76 concert halls arranged in six subjective categories according to their acoustical quality. Of all these there are three halls listed in the category A+ as outstanding "superior" and six halls in the category A as "excellent." Eight of these are shoebox-shaped, a fact that gives rise to the question as to whether a good room-acoustical quality is linked to a rectangular shape of the room.

The subjective assessment parameters used are, on the one hand, the "warmth" of the sound pattern and, on the other hand, the brilliance of the same. Warmth and brilliance refer in this context mainly to the influence of the sound energy density on the lower and the higher frequency ranges, respectively. Questions concerning initial reflections will be left out of consideration for the time being; only the timbre in the decay process will be considered.

The criterion "bass ratio" *BR* (Beranek)[9], provides indisputable evidence on the warmth of the sound, see section 9.2.1.2. The desirable optimum value range for music performances is between 1.0 to 1.3. According to Beranek for rooms having a reverberation time lower than 1.8s, it is permissible to have a bass ratio of up to 1.45.

For objective assessment of the timbre, Schmidt[23] has defined the timbre measure. By analogy with the *BR* and the timbre measures there was an equivalent measure deduced and introduced as *TR1* (Timbre Ratio). It is used only for evaluating comparative aspects of brilliance

$$TR1 = \frac{T_{2000\,\mathrm{Hz}} + T_{4000\,\mathrm{Hz}}}{T_{125\,\mathrm{Hz}} + T_{250\,\mathrm{Hz}}} \qquad (9\text{-}46)$$

This numerical relationship is used to evaluate timbre as the ratio of the reverberation time at high compared to low frequencies. Thus the value $TR1 > 1$ stands for a longer reverberation time at higher frequencies rather than at lower frequencies, and hence in this context higher brilliance in the sound pattern.

As regards the primary room structure of a concert hall there are four basic forms considered: rectangle (shoebox), polygon, circle, and various trapezoidal forms.

The concert halls selected by Beranek for the categories A+ and A allow the following pairs of values to be ascertained in an occupied hall, see Table. 9-6.

Table 9-7 shows that, in shoebox-shaped rooms, the brilliance is lower in comparison to rooms of polygonal primary shape. However, on the basis of the brilliance ratio *TR*1, no such significant difference can be shown between rooms of a quasi-circular ground plan (five halls) and those having diverse trapezoidal primary shapes (nine halls).

Table 9-6. $T_{30,mid}$, *BR*, and *TR1* for Outstanding and Excellent Concert Halls

Room (in Alphabetical Order)	$T_{30,mid}$ in s	BR	TR1	Primary Structure
Amsterdam Concertgebouw	2.0	1.09	0.77	Rectangle
Basel, Stadt-Casino	1.8	1.18	0.74	Rectangle
Berlin, Konzerthaus	2.05	1.08	0.79	Rectangle
Boston Symphony Hall	1.85	1.03	0.78	Rectangle
Cardiff, David's Hall	1.95	0.98	0.87	Polygon
New York Carnegie Hall	1.8	1.14	0.78	Rectangle
Tokyo Hamarikyu Asahi	1.7	0.93	1.04	Rectangle
Vienna Musikvereinssaal	2.0	1.11	0.77	Rectangle
Zürich Tonhallensaal	2.05	1.32	0.58	Rectangle

9.3.3 Primary Structure of Rooms

9.3.3.1 Volume of the Room

As a rule, the first room-acoustical criterion to be determined as soon as the intended purpose of the room has been clearly established, is the reverberation time, see section 9.2.1.1. From Eq. 9-6 and the correlation between reverberation time, room volume and equivalent sound absorption area, graphically depicted in Fig. 9-6, it becomes evident that the room volume must not fall short of a certain minimum if the desired reverberation time is to be achieved with the planned audience capacity.

Table 9-7. Brilliance Ratio for 36 Examined Concert Halls

Room shape	Number of Examined Halls	Average Value $TR1$	Confidence Range $TR1$	Limit Values $TR1$
Rectangle	12	0.75	±0.05	0.70–0.80
Polygon	10	0.91	±0.05	0.86–0.97
Circle	5	0.75	±0.16	0.59–0.91
Trapezoidal	9	0.75	±0.15	0.63–0.86

For enabling a tentative estimate of the acoustically effective room size required with regard to its specific use there serves the volume index *k*, which indicates the minimum room volume in m^3/listener seat, Table 9-8. In case an auditorium is used for concert events, the volume of the concert enclosure is added to the volume of the auditorium without increasing, however, the seating capacity of the auditorium by the number of the additional performers (orchestra, choir). For theater functions the volume of the stage house behind the portal is left out of account.

The minimum required acoustically effective room volume is calculated as follows:

$$V = k \cdot N \tag{9-47}$$

where,

V is the acoustically effective room volume in m^3 (ft^3)

k is the volume index according to Table 9-5 in m^3/seat (ft^3/seat),

N is the seating capacity in the audience area.

If a given room is to be evaluated regarding its suitability for acoustic performances, the volume index may be useful for providing a rough estimate and simultaneously for determining the scope of additional sound-absorptive measures.

If the volume index falls short of the established guide values, the desirable reverberation time cannot be achieved by natural acoustics. With very small rooms, especially orchestra rehearsal rooms, it is moreover possible that loudness results in excessive fortissimo (in a rehearsal room of 400m^3 (14,000ft^3) volume and with 25 musicians it may reach up to 120dB in the diffuse field). In rooms of less than 100m^3 (3500ft^3) the eigenfrequency density results are insufficient.[2] This leads to a very unbalanced frequency transmission function of the room giving rise to inadmissible timbre changes.

Table 9-8. Volume Index *k* versus Room Volume

No.	Main use	Volume Index *k* in m^3/seat (in ft^3/Seat)	Maximum Effective Room Volume with Natural Acoustics in m^3(ft^3)
1	Speech performances, e.g., spoken drama, congress hall and auditorium, lecture room, room for audiovisual performances	3–6 (110–210)	5000 (180,000)
2	Music and speech performances, e.g., musical theater, multi-purpose hall, town hall	5-8 (180–280)	15,000 (550,000)
3	Music performances, e.g., concert hall	7-12 (250–420)	25,000 (900,000)
4	Rooms for oratorios and organ music	10-14 (350–500)	30,000 (1,100,000)
5	Orchestra rehearsal rooms	25-30 (900–1100)	-

Excessive loudness values require additional sound-absorptive measures, which may bring about too heavy a loudness reduction for low-level sound sources.

On the other hand it is not possible to increase seating capacity and room volume just as you like, since owing to the increase of the equivalent sound absorption area and the unavoidable air absorption, see Fig. 9-4 the attainable sound energy density in the diffuse field decreases and so as well the performance loudness, see Eq. 9-8. Moreover the distances within the performance area and to the listener are dissatisfactory expanded this way. For these reasons it is possible to establish an upper volume limit for rooms without electroacoustic sound reinforcement equipment, i.e., with natural acoustics, that should not be exceeded, see Table 9-2. These values depend, of course, on the maximum possible power of the sound source. By choosing Eq. 9-8 in level representation and using the reverberation time formula by Sabine,[6] one obtains the correlation between sound power level of the sound source L_W in dB and sound pressure level L_{diff} in dB in the diffuse sound field, as a function of the room parameters volume V in m^3, and reverberation time RT in s.[6]

$$L_{diff} = L_W - 10\log\frac{V}{T}\,\text{dB} + 14\,\text{dB}^* \qquad (9\text{-}48)$$

* add 29.5 dB in U.S. system.

The graphical representation of this mathematical relation is shown in Fig. 9-19. For determining the attainable sound pressure level in the diffuse sound field one can proceed from the following sound power levels L_W[3,29,30]

Music (Mean Sound Power Level with "Forte")	
Tail piano, open	L_W= 77-102dB
String instruments	L_W= 77-90dB
Woodwind instruments	L_W= 84-93dB
Brass instruments	L_W= 94-102dB
Chamber orchestra of 8 violins	L_W= 98dB
Small orchestra with 31 string instruments, 8 woodwind instruments, and 4 brass instruments (without percussion)	L_W= 110dB
Big orchestra with 58 string instruments, 16 woodwind instruments, and 11 brass instruments (without percussion)	L_W= 114dB
Singer	L_W= 80-105dB
Choir	LW = 90dB
***Speech* (Mean Sound Pressure Level with Raised to Loud Articulation)**	
Whispering	L_W= 40-45dB
Speaking	L_W=68-75dB
Crying	L_W= 92-100dB

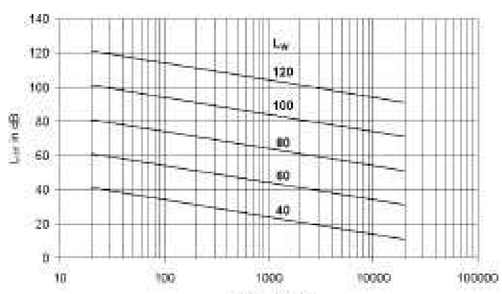


Figure 9-19. Correlation between the sound pressure level Ld_{iff} in the diffuse field and the sound power level L_W as a function of room volume V and reverberation time RT. L_{diff} in dB, V/RT in m^3/s.

With musical performances, for instance in forte and piano passages, perception of the dynamic range plays a decisive role for the listening experience, independently of the prevailing volume level. A sound passage or a spoken text submerging in the surrounding noise level is no longer acoustically registered and the performance considered being faulty. The mean dynamic range of solo instruments lies with slowly played tones according to reference 30 between 25dB and 30dB, with orchestra music it is about 65dB and with singers in a choir about 26dB. The dynamic range of a talker is about 40dB and that of soloist singers about 50dB. Without taking into account the timbre of a sound source, the signal to noise ratio (S/N) should generally be at least 10dB with pianissimo or whispering.

With large room volumes and high frequencies, the increase of energy attenuation loss caused by the medium air should not be neglected. This shall be illustrated by an example: in a concert hall with a room volume of 20,000m^3, the unavoidable air attenuation at 20°C and 40% relative humidity accounts for an additional equivalent sound absorption area which at 1000Hz corresponds to an additional 110 persons and at 10kHz even to 5000 additional persons.

9.3.3.2 Room Shape

The shape of a room allows a wide margin of variability, since from the acoustical point of view it is not possible to define an optimum. Depending on the intended purpose, the shape implies acoustical advantages and disadvantages, but even in the spherical room of a large planetarium it is by room-acoustical means (full absorbing surfaces) possible to achieve good speech intelligibility. Acoustically unfavorable, however, are room shapes that do not ensure an unhindered direct sound supply nor any omnidirectional incidence of energy-rich initial reflections in the reception area, as is the case, for instance, with coupled adjoining rooms and low-level audience areas under balconies or galleries of low room height.

When selecting different room shapes of equal acoustically effective volume and equal seating capacity there may result very distinct characteristics as regards the overall room-acoustical impression. A more or less pronounced inclination of the lateral boundary surfaces may produce different reverberation times, see reference 31. In combination with a sound reflecting and not much structured ceiling layout, prolongation of the reverberation time up to a factor of two results especially large, if long-delayed sound reflection groups are produced by side wall surfaces that are inclined outwards and upwards. But if these wall surfaces are inclined towards the sound-absorbing audience area, the shorter path lengths thus achieved may considerably reduce the reverberation time as compared to the usual calculating methods with vertical boundary surfaces.

Also with similar room shapes, different room-acoustical conditions are obtained by just varying the furnishing of the room (platform, audience areas).

All acoustically usable room shapes have in common that the unhindered direct sound and energy-rich initial reflections reach the listener. Deviations from this rule occur through direct-sound shading in the orchestra pit of an opera theater. Diffraction compensates this effect partially and the listening experience is adapted to a different sound impression which is similar to the case of unhindered sound irradiation. The initial reflections must arrive at the listener's seat within a path difference to direct sound of approximately 17 m (50ms) for speech and 27m (80ms) for music performances.

Decisive for an adequate spatial impression with musical performances are first of all the lateral sound reflections. The more the spaciousness is supported this way, the more the orchestra sound gains, according to Meyer,[30] in “volume” and “width.” The increase of sound intensity perceptible with forte-play is thus enhanced beyond the mere loudness effect so that the subjectively perceived dynamic range is expanded. By the same token, also spaciousness is subjectively enhanced by an increased loudness of the sound source.

From these general premises it is for different arrangement patterns between performers and listeners possible to derive universally valid guidelines for fundamental room-acoustical problems of certain typical room ground-plan layouts. In this regard one can distinguish between *purely geometrical layouts* with parallel boundary lines (rectangle, square, hexagon) on all sides, with at least two mutually slanted boundary lines (trapezoid) and generally curved boundary lines (circle, semicircle, ellipse) and *irregular layouts* with asymmetric or polygonal boundary lines.

9.3.3.2.1 Ground Plan

For obtaining lateral sound reflections, a room with a rectangular ground plan is very well suited if the performance zone is arranged at an end wall and the width of the room is in the range of 20m (66ft), see Fig. 9-20A. This is the typical example of the linear contact in a shoebox layout of a classic concert hall (Musik-vereinssaal Vienna, Symphony Hall Boston, Konzerthaus Berlin).

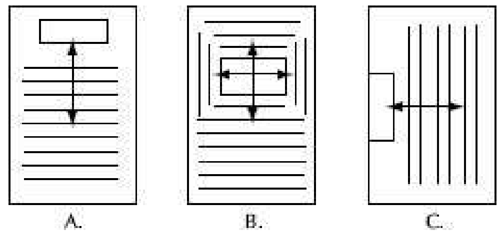


Figure 9-20. Examples of so-called arrangement patterns between performers and listeners in a room with a rectangular ground plan.

If the performance zone is shifted from the end wall towards the middle of the room, Fig. 9-20B, a circular contact may come into being as a borderline case, where audience or a choir may be arranged laterally or behind the platform. Owing to the relatively pronounced frequency-dependent directional characteristic of most sound sources (singers, high-pitched string instruments, etc.) there occur herewith, especially in the audience area arranged behind the platform, intense balance problems which may even lead to unintelligibility of the sung word and to disturbing timbre changes. On lateral seats at the side of the platform, the listening experience can be significant impaired if due to room reflections, visually disadvantaged instruments are perceived louder than instruments located at closer range. This effect is even enhanced by lateral platform boundary surfaces, whereas an additional rear sound reflection area supports sound mixing. Often these acoustical disadvantages are, however, subordinated to the more eventful visual experiences.

If the performance zone is arranged in front of a longitudinal wall, see Fig. 9-20C, short-time lateral initial reflections get missing especially with broad sound sources (orchestras), whereby the mutual hearing and consequently the intonation get impaired. Soloist concerts or small orchestras (up to about seven musicians) may still provide satisfactory listening conditions, if ceiling height and structure provide clarity-enhancing sound reflections. By means of a sound-reflecting rear wall combined with adjustable lateral wall elements which do not necessarily disturb the visual impression, it is possible to attain good room-acoustical conditions with not too long rooms (up to about 20m or 66ft). For spoken performances this way of utilization provides advantages on account of the short distance to the talker, but disadvantages due to timbre changes impairing intelligibility. As the talker chooses to speak towards the audience seats in the middle of the room, the lateral seating areas are bound to be disadvantaged on account of the frequency-dependent directional characteristic. According to Meyer[3], the sound pressure level reduction laterally to the talker when articulating the vowels "o," "a," and "e" is 0dB, 1dB, and 7dB, respectively.

A special case of the rectangular room ground plan is the **square** with approximately equal side lengths. Especially for multi-purpose use such halls allow the realization of diverse forms of confrontation with the audience for which good acoustical conditions are given in small rooms with about 500 seats,[29] assuming some basic principles are considered, see Fig. 9-21A to C. Room variant A represents the classical linear contact ensuring a good direct-sound supply to the listeners, especially with directional sound sources (talkers, singers, directional instrumental groups). Variant B offers an acoustically good solution for sound sources of little extension (talkers, singers, chamber music groups), since a good lateral radiation into the room is given. It is true, however, that in the primary structure there is a lack of lateral sound reflections for supporting mutual hearing and intonation in the performance zone. The amphitheatrical arrangement shown in variant C is suitable for only a few kinds of performance, since apart from visual specialities there are above all acoustic balance problems to be expected. With directional sound sources, e.g., talkers and singers, the decrease of the direct sound by at least 12 dB versus the straight-ahead viewing direction produces intelligibility problems behind the sound source.

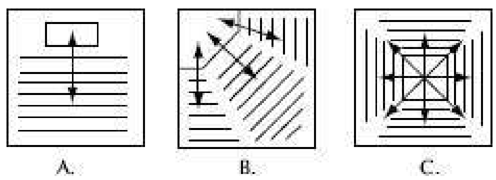


Figure 9-21. Diverse platform arrangements in a room with a square ground plan.

The **trapezoidal ground plan** enables or principle two forms of confrontation: the diverging trapezoid with the lateral wall surfaces diverging from the sound source and the *converging* trapezoid with the sound source located at the broad end side. The latter ground plan layout, however, is from the architectural point of view not used in its pure form, see Fig. 9-22.

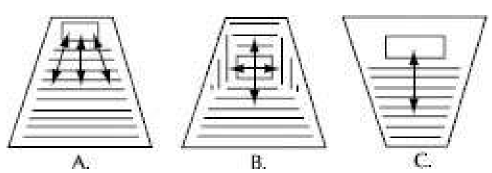


Figure 9-22. Examples of so-called arrangement patterns between performers and listeners in a room with a trapezoidal ground plan.

Variations of the first ground plan layout with a curved rear wall are designated as “fan-shaped” or “piece of pie.” The room-acoustical effect of the trapezoidal layout depends essentially on the diverging angle of the sidewall surfaces. The room shape shown in Fig. 9-22A produces room-acoustical conditions, which are similarly favorable as those in a rectangular room used for music, if the diverging angle is slight. With wider diverging angles, the energy-rich initial reflections especially from the side walls are lacking in the whole central seating area, which is a characteristic conditioned by this primary structure. A principal comparison of the fraction of lateral sound energy produced merely by the ground plan layout is shown in Fig. 9-23. As was to be expected, the comparable, relatively narrow ground plan of the rectangular shape shows a higher lateral sound fraction than the diverging trapezoid. For spoken performances this situation is relatively uninteresting, since in most cases the lacking early lateral reflections can be compensated by early reflections from the ceiling. If the performance zone is shifted in an amphitheater-like solution to the one-third point of the room ground plan, see Fig. 9-22B, this variant is suitable only for musical performances. Especially the listeners seated behind the performance zone receive a very spatial, lateral-sound accentuated sound impression.

The room acoustically most favorable with a trapezoidal layout is that of a converging trapezoid with the performance zone located at the broad end side, Fig. 9-22C. Already without additional measures on the side of the platform, the audience areas receiving low early lateral sound energy gets reduced to a very small area in front of the sound source; almost all the other part of the audience area receives a strong lateral energy fraction, see Fig. 9-23. Unfortunately, this room shape has only perspectives for architectural realization in combination with a diverging trapezoid as a platform area. The arrangement of so-called “vineyard terraces” constitutes herewith a very favorable compromise solution in which wall elements in the shape of converging trapezoids are additionally integrated in the seating area. The effective surfaces of these elements direct energy-rich initial reflections into the reception area according to reference 32, see Fig. 9-24. Examples of projects accomplished in this technique are the concert halls of the Gewandhaus in Leipzig and De Doelen in Rotterdam.[33]

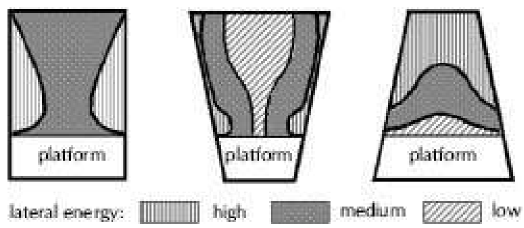


Figure 9-23. Principal portions of early lateral sound reflections in rectangular and trapezoidal rooms platform.

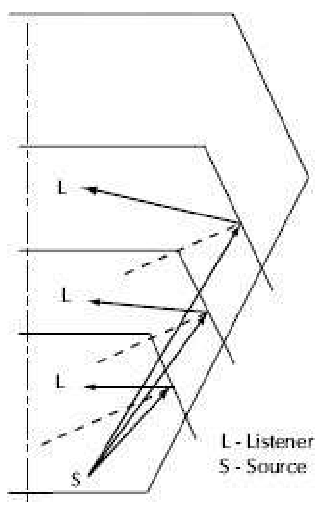


Figure 9-24. Lateral sound reflections produced by “vineyard terraces.”

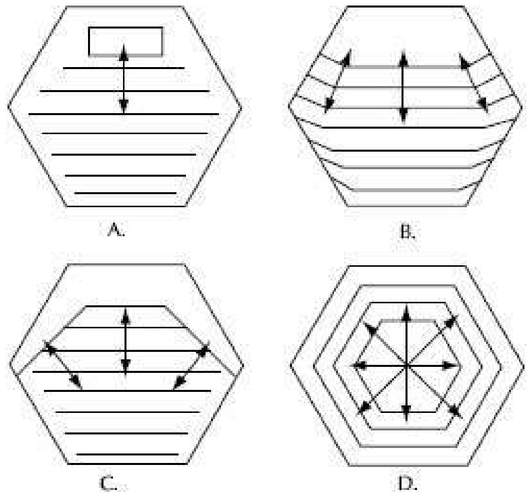


Figure 9-25. Diverse platform arrangements in a room of hexagonal shape.

This combination of a ground plan layout can also be realized in the shape of a hexagon which is a common application of a regular polygon ground plan. Elongated hexagons show room-acoustical properties similar to those of the combination of a diverging and a converging trapezoid or, provided the converging or diverging angle is slight, to those of a rectangular room. If the ground plan is that of a regular hexagon, see Fig. 9-25, the necessary lateral sound reflections are lacking especially with musical performances. Thanks to its varied uses and the short distances between the performance and reception zones it provides, this shape is rather advantageous for congress and multi-purpose halls also from the acoustical point of view. The amphitheater-like arrangement of stage and audience of Fig. 9-25D shows acoustical similarities to the rectangular variant of Fig. 9-20C. With sound sources having a pronounced directional characteristic there occur timbre and clarity problems for listeners seated at the sides and behind the platform area, which cannot be compensated by means of additional secondary structures along the walls.

Ground plans with monotonically c**urved boundary surfaces** (circle, semicircle, Figs. 9-26A to D) produce, due to their concave configuration towards the sound source and especially if the tiers are only slightly sloped or not at all, undesirable sound concentrations. On account of the curved surfaces the sound pressure level may, in the concentration point even surpass that of the original sound source by 10dB and thus become an additional disturbing sound source. According to reference 34 their result, depending on frequency, travel time, and circle diameter, are shown in Fig. 9-27. One recognizes instances of migrating punctual and flat-spread sound concentration (the so-called caustic), which even after long travel times never do lead to a uniform sound distribution. Without any structuring in the vertical plane and without broadband secondary structures, rooms having a circular ground plane are acoustically suited neither for speech nor for musical performances.

With asymmetrical ground plans, see Fig. 9-26E, there exists for musical performances the risk of a very poor correlation between the two ear signals, an effect that may give rise to an exaggerated spaciousness. Energy-rich initial reflections are to underline the visual asymmetry only as far as required for the architectural "comprehension" of the room, otherwise the room produces balance problems with musical performances, leaving questions regarding the arrangement of the orchestra instrument groups unsolved. Elliptical ground plans, see Fig. 9-26F, are without reflection-supporting measures acoustically suited only for locally fixed sound sources, a general utilization is not recommended owing to the focus formation in the performance zone as well as in the audience area. This refers especially to the atrium courtyards of unstructured glass walls and plane floor in large office buildings, which are a modern architectural trend. These functionally designed entrance foyers are often used for large musical events which, however, can in no way satisfy any room-acoustical requirements.

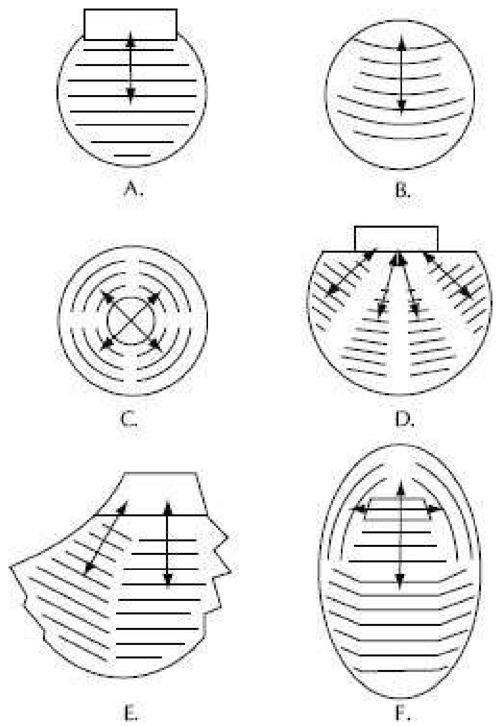


Figure 9-26. Examples of so-called arrangement patterns (confrontations) between performers and listeners in a room with curved boundary surfaces.

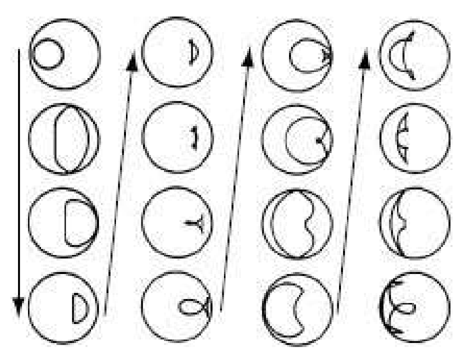


Figure 9-27. Propagation of the wave front in a circle (caustic).

9.3.3.2.2 Ceiling

In general, the ceiling configuration contributes little to spaciousness of the sound field, but all the more to achieving intelligibility with speech, clarity with music, volume, and guidance of reverberation-determining room reflections. For speech, the reverberation time should be dimensioned as short as possible. Therefore the ceiling should be configured in such a way that possibly each first sound reflection reaches the middle and rear audience areas, see Fig. 9-28. For musical performances the mean ceiling height has to comply with the volume-index requirements. For achieving an as-long-as possible reverberation time, the ceiling should have its maximum height where the length or width of the room are maximal. The repeated reflection of the sound energy by the involved boundary surfaces produces long travel times, while the required slightness of energy reduction by the reflections has to be ensured by a negligible sound absorption coefficient of these surfaces.[35] Thanks to an adequately chosen geometry and size of the surfaces involved in this reverberation-time generating mechanism it is possible to reduce the reverberation time in the low frequency range in a desirable fashion, while the sound impression is not deprived of its stimulating sound energy by sound absorption measures. In the concert hall of the Gewandhaus in Leipzig[33] the room has its widest extension in the rear audience area, so that here—as a result of simulation measurements in a physical model (see section 9-3)—the height of the ceiling was chosen to have its maximum, see Fig. 9-29A and B. In contrast to that, the maximum room width of the concert hall of the Philharmonie Berlin, see Fig. 9-29C and D is in the region of the platform, so that for realizing an optimum reverberation time, the maximum room height has to be above the platform.

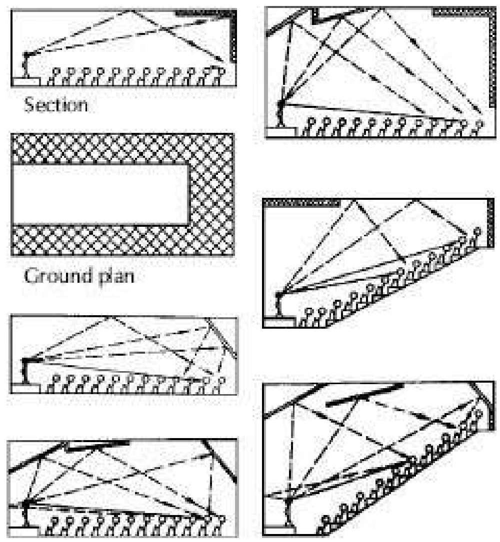


Figure 9-28. Examples of acoustically favorable ceiling designs.

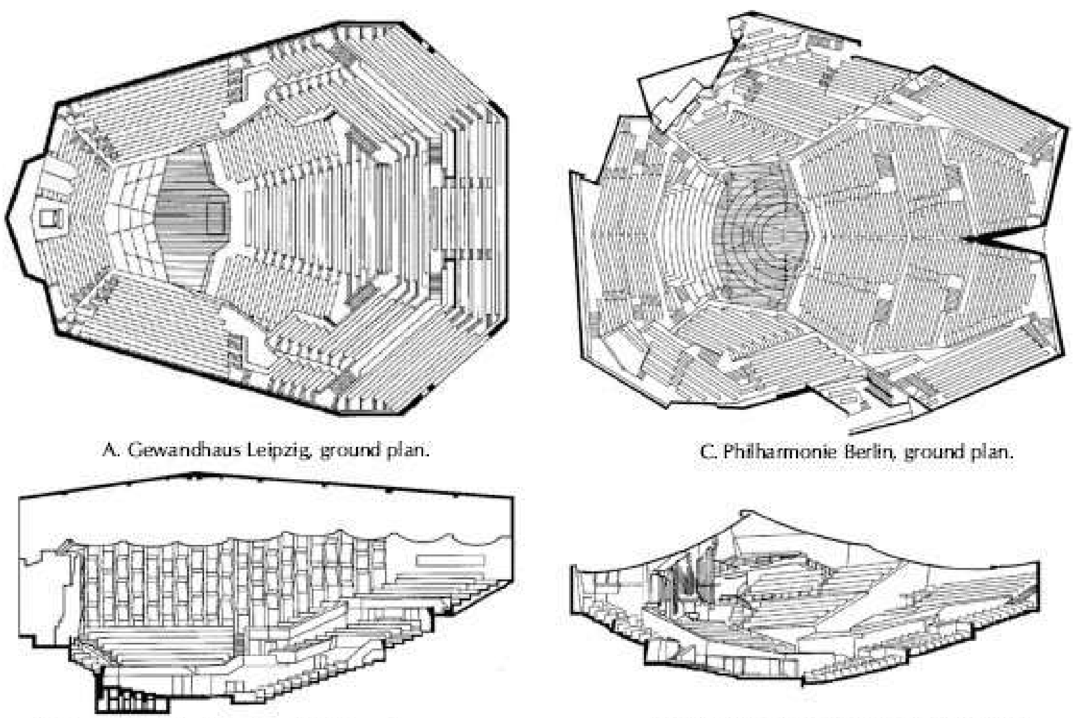


Figure 9-29. Various concert halls.

This explains also why it was necessary to arrange room-height reducing panels in this concert hall. With music, the ceiling above the performance zone must neither fall below nor exceed a certain height in order to support the mutual hearing of the musicians and to avoid simultaneously the generation of disturbing reflections. According to reference 3, the lower limit of the ceiling height in musical performance rooms is 5 to 6m (16 to 19ft), the upper limit about 13m (43ft).

In large rooms for concert performances, the ceiling configuration should provide clarity-enhancing sound reflections in the middle and rear audience areas and simultaneously avoid disturbing reflections via remote boundary surfaces. Owing to the geometrical reflection, a plain ceiling arrangement, Fig. 9-30A, supplies only a slight portion of sound energy to the rear reception area, but in the front area (strong direct sound), the sound reflected by the ceiling is not required. In the rear ceiling area, however, the sound energy is reflected towards the rear wall from where it is returned, according to the unfavorable room geometry, as a disturbing echo (so-called theater echo), to the talker or the first listener rows. Keeping this in mind, the ceiling surfaces above the performance zone and in front of the rear wall should point perpendicular towards the middle seating area, Figs. 9-30B to D.

Monotonically curved ceilings in the shape of barrel vaults or cupolas show focusing effects, which in the neighborhood of such focuses may produce considerable disturbances in the listener or performer areas. The center of curvature should therefore be above half the total height of the room or below twice the height, see Fig. 9-31.

$$r \leq \frac{h}{2} \text{ or } r \geq 2h \qquad (9\text{-}49)$$

According to Cremer[9] it is guaranteed in this case that there do not originate from the curved ceiling any stronger reflections towards the receiving area than from a plain ceiling at apex height.

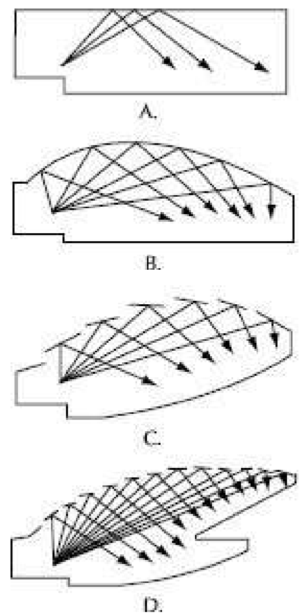


Figure 9-30. Ceiling configurations for obtaining energy-rich initial reflections in the middle and rear listener areas.

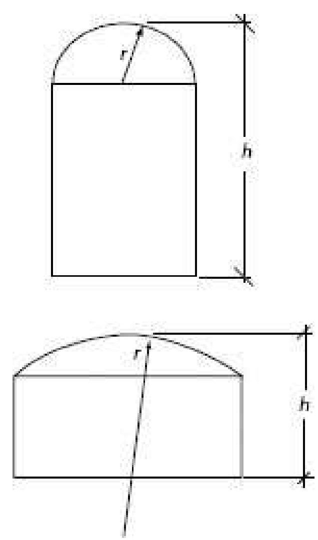
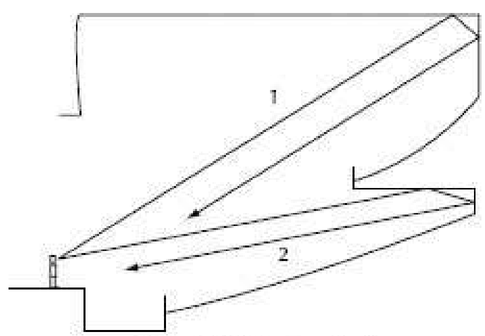


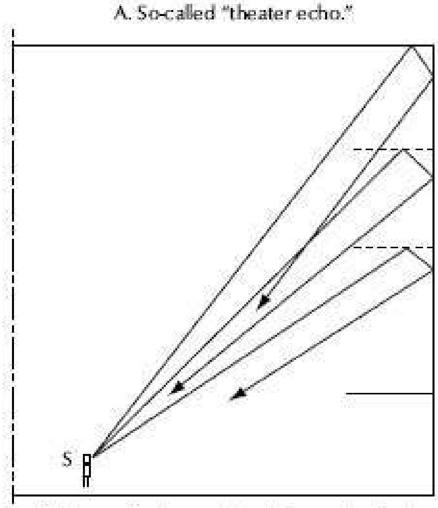
Figure 9-31. Focusing due to vaulted ceilings.

9.3.3.2.3 Balconies, Galleries, Circles

With proper arrangement and dimensioning, balconies and circles may have an acoustically favorable effect, since they contribute to a broadband diffuse sound dispersion and are also able to supply initial reflections for enhancing clarity and spatial impression. In this respect it is necessary, however, to decide if these reflections are desirable. Fig. 9-32A shows a graph of long-delayed sound reflections which give rise to very disturbing echo phenomena, the so-called theater echoes. Instrumental in the generation of these reflections is first of all the rear wall in combination with horizontal architectural elements (circle, gallery, balcony, ceiling). The disturbing effect of these sound reflections has to be avoided. Protruding balconies are, thanks to their horizontal depth, capable of shading these "corner reflectors" and to turn the reflections into useful sound, see Fig 9-32B.



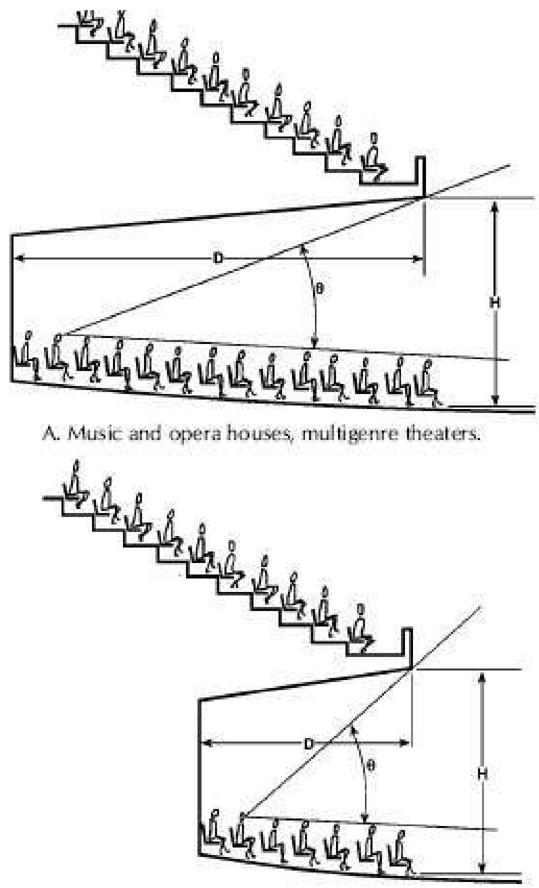
A. So-called "theater echo."



B. Edge reflections under circles and galleries.

Figure 9-32. Echo phenomena due to edge reflections.

The arrangement of far protruding circles is acoustically problematical with regard to the depth *D* (distance of the balustrade or of a room corner above it from the rear wall) and the clearance height *H* of the circle above the parquet or between two circles arranged one above the other. If the protrusion is very deep, the room area situated below it is shaded against reverberant sound and clarity-enhancing ceiling reflections. So his area may be cut off from the main room and have an acoustic pattern of its own with strongly reduced loudness, unless certain construction parameters are observed,[2,7,9], see Fig. 9-33.



B. Concert halls.

Room	Music and opera houses, multigenre theaters, Fig. 9-33A	Concert hall Fig. 9-33B
Circle depth D	$2H$	$<H$
Angle θ	25°	45°

Figure 9-33. Geometry of circle arrangement in A. Music and opera houses, multi-genre theaters and B. Concert halls.

9.3.3.3 Room Topography

9.3.3.3.1 Sloping of Tiers, Sight Lines

For all room-acoustical parameters describing time and registers clarity, the energy proportion of direct sound and initial reflections is of great importance. With the sound propagating in a grazing fashion over a plain audience area there occurs a strong, frequency-dependent attenuation (see section 9.3.4.4.4). Also visually such a situation implies considerable disadvantages by obstruction of the view towards the performance area. These disturbing effects are avoided by a sufficient and, if possible, constant super elevation of the visual line. According to Fig. 9-34, this is the super elevation of the visual line (virtual line between eye and reference point) of a tier $n + 1$ as against tier n.

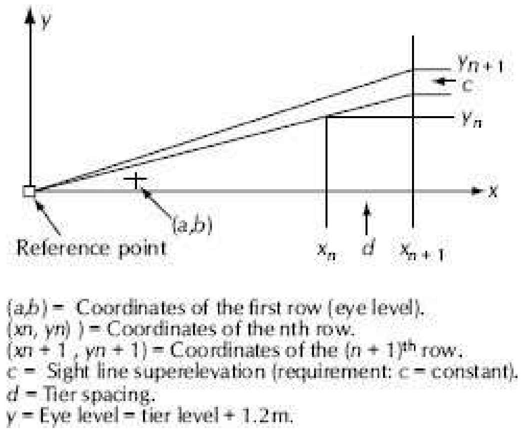


Figure 9-34. Sloping of tiers (schematic view).

With a tier arrangement having a constant step height, i.e., continuous sloping in longitudinal direction of the room, it is not possible to achieve a constant super elevation c. Mathematically it is the curve of a logarithmic spiral[34] in which the super elevation increases alongside with the distance from the reference point that realizes a constant super elevation of the visual line.

As this implies, however, steps of different height for the individual tiers, a compromise must be found by either adapting the step height or by combining several tiers in small areas of constant sloping. In concert halls, the areas arranged in the shape of “vineyard terraces” (see section 9.3.3.2.1) constitute in this respect an acoustically and optically satisfactory solution.

The eye level $y(x)$ is calculated with

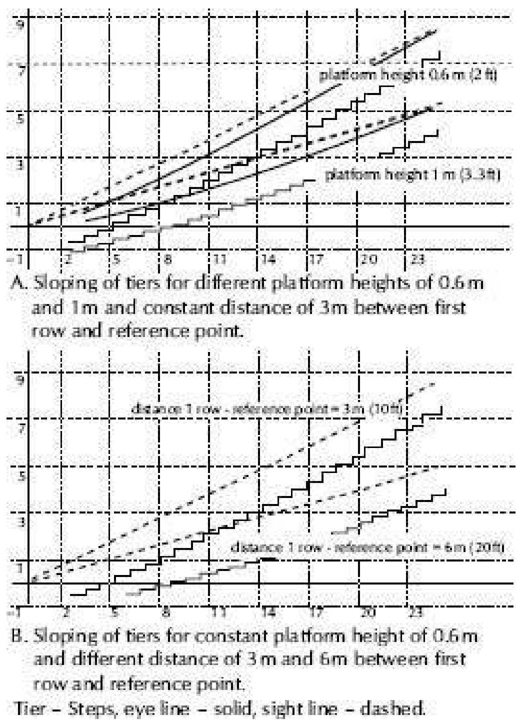
$$y = y_0 + \frac{c \cdot x}{d} \cdot \ln\frac{x}{a} + \frac{x}{a}(b - y_0) \tag{9-50}$$

where,

for $y_0 = 0$ there is: $y = \frac{cx}{d} \cdot \ln\frac{x}{a} + \frac{b \cdot x}{a}$

The superelevation of the visual line should amount to at least 6cm (2.5in).

For estimating the required basic sloping of tiers one should keep in mind that the platform must be completely observable from all seats during the performances. The reference point to be chosen to this effect should, if possible, be the front edge of the platform. Depending on a reasonable platform height of between 0.6 to 1m (2 to 3.3ft), the results of the sloping values are shown in Fig. 9-35A.



A. Sloping of tiers for different platform heights of 0.6 m and 1m and constant distance of 3m between first row and reference point.

B. Sloping of tiers for constant platform height of 0.6 m and different distance of 3 m and 6m between first row and reference point.

Tier – Steps, eye line – solid, sight line – dashed.

Figure 9-35. Effect of sloping tiers.

By increasing the distance between first tier and viewing point (observable area of the platform) it is, of course, possible to notably reduce the necessary sloping of tiers, see Fig. 9-35B.

With plain parquet arrangement which used to be the case in concert halls serving also for banquets or with classicistic architecture (Musikvereinssaal Vienna, Konzerthaus Berlin, Symphony Hall Boston, Herkulessaal Munich, etc.), a certain though normally somewhat dissatisfactory compensation is possible by means of an appropriate vertical staggering within the performance area (especially feasible for concert performances). For a basic platform height of 0.6 to 0.8m (2 to 2.6ft) it is possible to derive the theoretical required sloping heights from Eq. 9-50: with a length of a plain seating area of about 14m (46ft), a vertical staggering on the platform of about 3m (10ft) has to aimed at, and with 18m (59ft) this amounts to 4m (13ft). These values are generally not easy to realize, but show the necessity of an ample vertical staggering of the orchestra on the platform in rooms with plain parquet arrangement. However, if the optimum sloping of tiers according to Eq. 9-50 is realized on principle, it is possible for a sound source situated in the middle of the orchestra about 6m (20 ft) behind the front edge of the platform to achieve the required view field angle by an elevation of 0.25m (≈1ft), and for the entire depth of the orchestra arrangement this elevation amounts to only about 1m (≈3.3ft). In concert halls with a sufficient sloping of tiers in the audience area, vertical staggering of the orchestra plays no more than a subordinate role for the unhindered direct sound supply to the audience area.

9.3.3.3.2 Platform Configuration in Concert Halls

With concert performances the performance area for the orchestra (platform) must be an acoustical component of the auditorium, which means that both sections of the room must form a mutually attuned unity. This unity must not be disturbed by intermediate or other built-in elements. Any individual room-acoustical behavior of its own of a too small concert stage enclosure must be avoided. As used to be the case with many opera houses, this has a sound coloration deviating from that of the main auditorium and will be perceived as alienated. According to reference 36 the volume of a concert stage enclosure should be at least 1000m^3 (35,300ft^3). The sloping angles of the lateral boundary walls, referred to the longitudinal axis of the room, should be relatively flat. Takaku[36] defines an inclination index *K* according to Eq. 9-51

$$K = \frac{\sqrt{\frac{WH}{\pi}} - \sqrt{\frac{wh}{\pi}}}{D} \tag{9-51}$$

where,

K is the inclination index,

W is the proscenium width,

H is the proscenium height,

w is the width of the rear wall,

h is the height of the rear wall,

D is the enclosure depth.

According to reference 36, optimum conditions for the mutual hearing of the musicians are achieved for a concert stage enclosure in the shape of a truncated pyramid with $K \leq 0.3$, see Fig. 9-36.

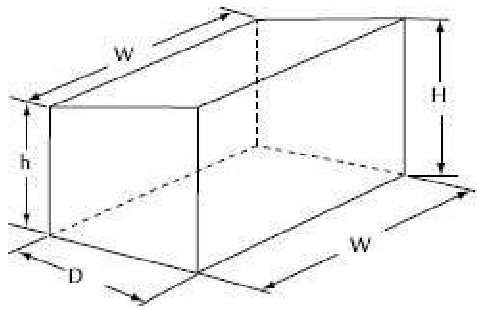


Figure 9-36. Geometrical parameters of a concert enclosure.

The more pronounced the diffuse subdivision of the inner surfaces of the concert stage enclosure, the smaller results the dependence of the room-acoustical parameters on the inclination index *K*.

If the platform boundaries are not formed by acoustically favorable solid wall and ceiling surfaces, additional elements have to be installed. The surface-related mass of the planking of these platform boundary surfaces should be chosen in such a way that the sound energy reduction by absorption results as little as possible. (The thinner the boundary walls the higher the low frequency absorption.) To this effect, area-related masses of about 20kg/m^2 (0.85 lbs/ft^2) are generally sufficient, in the neighborhood of bass instruments about 40kg/m^2 (1.7lbs/ft^2).

The vibration ability of the platform floor has only an insignificant influence on its sound radiation. With a relatively thin platform floor (12.5mm (0.5 in) plywood[37] there may well result a sound amplification of between 3dB to 5dB in the lower frequency range, but one should also not forget in this respect the positive psychological feedback of a vibrating floor on the players.[3] As a rule, the area-related mass of the platform floor should not fall below 40kg/m^2.

By comparison with a rigid floor a vibrating platform floor has for the sound radiation of the bass string instruments with pizzicato play (faster decay resulting in a dry sound), the disadvantage of a reduced airborne sound energy, which can, however, be technically compensated with bow strokes.[3] This is why the platform floor should be frequency-tuned as low as possible.

The platform boundary surfaces should be structured in such a way that the mutual hearing of the musicians is supported, disturbing echo phenomena (e.g., by parallel wall surfaces) are avoided and a well-mixed sound pattern gets radiated into the audience area. Obtaining a thorough mixing of the sound pattern requires a frequency-independent substructure of the boundary surfaces.

The space required per musician is about 1.4m^2 (15ft^2) for high-pitched string and brass instruments, 1.7m^2 (18ft^2) for low-pitched string instruments, 1.2m^2 (13ft^2) for woodwind instruments and 2.5m^2 (27ft^2) for the percussion. From this one can infer that, with due consideration of the participation of soloists (tail piano, etc.), the area of a concert platform (without choir) should generally not fall much below 200m^2 (2200ft^2), in which case the width should be about 18m (60ft) at the level of the high-pitched strings, and the maximum depth about 11m (36ft).

Depending on the sloping of tiers in the audience area (see section 9.3.3.3.1), a vertical staggering of the orchestra is necessary especially if the audience area in the parquet is level or only slightly sloping. In the Musikvereinssaal Vienna the level difference on the platform is 1.8m (6ft), in the Berliner Philharmonie which was destroyed during WWII, it was 2.8m (9.2ft). In such a case it is necessary to have one step in the string group approximately 250mm (10 in), the following steps to and between the two rows of woodwind instruments should each be 500mm (20in) high. For the brass instruments or the percussion a further step of about 150mm (6in) is sufficient.

A choir which in the staging of a grand concert is normally lined up behind the orchestra, can profit only from the lateral wall surfaces and the ceiling of the room with regard to clarity-enhancing sound reflections, the floor area is shaded. Since according to Meyer[3] the main radiation axis of the singers' strongest sound fractions is inclined about 20° downwards, the choir line-up should be relatively steeply staggered in order to ensure clarity and definition of articulation in the choir sound. With a flat line-up, however, only reverberance is increased. This is perceived as disturbing in rooms with a long reverberation time, whereas it may be rather desirable in reverberation-poor rooms. The optimum value of vertical staggering within a choir is about 45°, see reference 3, i.e., the steps should be equal in breadth and height in order to enable simultaneously an unhindered sound radiation to the lateral boundary surfaces of the room.[3]

9.3.3.3.3 Orchestra Pit

On principle, the arrangement of the orchestra with musical stage plays in the so-called pit at the border line between stage and auditorium is acoustically unfavorable by comparison with orchestra arrangements on the stage, e.g., stage music, but has developed historically from the performing practice in the 19th Century. In most baroque theaters the musicians were seated either at the same level as the first listeners' rows or only a few steps lower.[3] They were separated from the audience area only by an about 1m (3.3ft) high balustrade. With the introduction of an orchestra pit the visual contact between listener and stage was later on reduced, especially when the orchestras grew larger. Room-acoustical shortcomings lie herewith in the problem of balance between singing/speech on stage and the accompanying orchestra in the pit. Owing to the size and equipment of the stage area, the loudness of the singers gets altered with growing distance from the orchestra so that balance problems increase especially in case of low singing loudness and unfavorable pitch levels.

A further aspect concerns the register and time correspondence between stage and pit on which depend the intonation and ensemble playing.

The geometrical separation between the two performance areas (stage and orchestra pit) should in modern opera houses be as little as possible, not only in dependence on dramaturgical arguments, but also for visual and functional reasons. Consequently, the orchestra pit slides beneath the stage, so as to avoid that the distance of the first rows from the stage increases still further. The practicability of the thus formed covered area of the orchestra pit (proscenium area), required for dramaturgical reasons, implies that the covered area becomes bigger and bigger, while the open coupling space of the pit to the auditorium gets smaller and smaller. The orchestra pit thus becomes an independent room “tightly packed with musicians” and with low boundary surfaces, a low volume index, and a non-reflecting “subceiling” (opening) representing the outlet for irradiation of the more or less well mixed orchestra sound to the auditorium. Owing to the reduced distance of the musicians from the boundary surfaces, the sound pressure level in the pit increases by up to about 4dB, whereby the mutual hearing of the musicians is supported for low volume playing. With increased loudness the mutual hearing gets disadvantageously limited to loud instrumental groups in the low and medium frequency range.

Sound-absorbing wall or ceiling coverings or adjustable wall elements with preferential effect in the low-and medium-frequency range, arranged in the neighborhood of loud instruments, reduce the loudness desirably, but not the direct sound irradiation into the auditorium. This supports the clarity of the sound pattern.[3] If the orchestra pit level is very low, about 2.5m (8.2ft), direct sound fractions reach the parquet level only by diffraction, causing the sound pattern to be very bass-accentuated. Brilliance and temporal clarity become adequate only at those places where visual contact to the instrument groups is given (circles).

Acoustic improvement of this situation may be achieved on the one hand by a wider "opening" of the orchestra pit, so that energy-rich initial reflections are enabled via a corresponding structure of the adjacent proscenium area (proscenium ceiling and side-wall design). On the other hand the pit depth should not exceed certain limits. By means of subjective investigations with varying height of the pit floor, optimum solutions may easily be found here in combination with an adequate positioning of the instruments in the pit. With a balustrade height of about 0.8m (2.6ft), lowering the front seating area of the pit floor (high-pitch strings) to about 1.4m (4.6ft) produces generally good acoustical conditions. Towards the rear the staggering should go deeper.

Provided the orchestra plays with adapted loudness, an acceptable solution consists in an almost complete opening of the orchestra pit towards the proscenium side walls and an as little as possible covering towards the stage. If the open area amounts to at least 80% of the pit area, the orchestra pit becomes acoustically part of the auditorium and the unity of the sound source is ensured also with respect to coloration (example: Semperoper Dresden). Another solution consists in an almost completely covered orchestra pit with a small coupling area to the auditorium. This requires, however, a correspondingly large pit volume with a room height of at least 3m (10ft) (example: Festspielhaus Bayreuth). Common opera houses lie with their orchestra pit problems half-way between these two extremes. If there is a large orchestration accommodated in the pit, less “powerful” singers on the stage may easily become acoustically eclipsed. More favorable conditions can be obtained in this case by means of a pit covering, provided a sufficient volume is given, or by positioning the orchestra on a lower pit floor level.

Apart from the sound reflecting and sound absorbing boundary surfaces arranged in the pit for supporting the mutual hearing and the intonation, the inner faces of the pit balustrade should point perpendicularly towards the stage (slight inclinations on the side of the balustrade). In this way the stage is better supplied with initial reflections from the pit, whereas the pit receives a first reflection from sound sources on the stage. Convex curves (in the vertical domain) combined with a sound-absorbing effect in the low-frequency range are in this respect especially advantageous for making the supporting effect register-independent on the one hand and brilliance enhancing on the other hand. The edge of the stage above the pit vis-à-vis the conductor should be conformed geometrically in such a way that additional initial reflections are directed to the audience area. The lateral configuration of the pit opening, combined with an appropriate subconstruction of the proscenium side wall, should ensure a maximum of sound reflections towards the pit and the stage.

9.3.4 Secondary Structure of Rooms

9.3.4.1 Sound Reflections at Smooth Plane Surfaces

With the reflection of sound rays from boundary surfaces one can principally define three types of reflection which differ from one another by the relation between the linear dimensions and the wavelength and by the relation between the reflected and the incident sound ray, see Fig. 9-37.

- **Geometrical reflection**, Fig. 9-37A: $b < \lambda$, $\alpha = \beta$ (specular reflection according to the reflection law in one plane perpendicular to the carrier wall).
- **Directed (local) reflection**, Fig. 9-37B: $b > \lambda$, $\alpha = \beta$ (specular reflection according to the reflection law, referred to the effective structural surface).
- **Diffuse reflection**, Fig. 9-37C: $b \approx \lambda$, (no specular reflection, without a preferred direction).

A geometrical sound reflection occurs at a sufficiently large surface analogously to the reflection law of optics: the angle of incidence a is equal to the angle of reflection ß and lies in a plane perpendicular to the surface, see Fig. 9-38. This reflection takes place only down to a lower limit frequency f_{low}

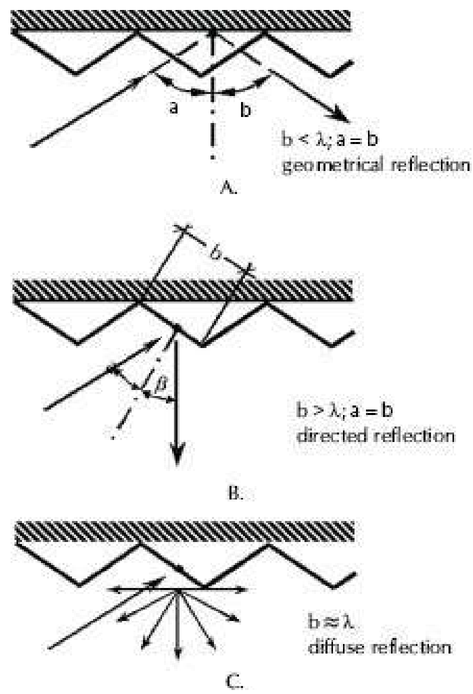


Figure 9-37. Basic sound reflections at smooth, plane surfaces.

$$f_{low} = \frac{2c}{(b \cdot \cos\alpha)^2} \cdot \frac{a_1 \cdot a_2}{a_1 + a_2} \tag{9-52}$$

where,

$a_1, a_2 > b$,

c is the velocity of sound in air.

Below f_{low} the sound pressure level decay amounts to 6dB/octave.[38]

Eq. 9-52 has been graphically processed by reference 3, see Fig. 9-39. With a reflector extension of 2m (6.6ft) at a distance of 10m (3ft) each from the sound source and to the listener, the lower limiting frequency is, e.g., about 800Hz with vertical sound incidence and about 1600Hz with an incidence angle of 45°. If this reflector is installed as a panel element in the front part of the platform, the frequency region of the sound reflections is about one octave lower with almost platform-parallel arrangement than in a

45° inclined position. The desired limiting frequency goes down to lower values under the following circumstances:

- The bigger the effective surface.
- The nearer to the sound source and to the listener the reflector is installed.
- The smaller the sound incidence angle.

Apart from the geometry of the reflectors, the area-related mass of the same also has to be consistent with certain limit values in order to obtain a reflection with as little a loss as possible. If the reflectors are employed for speech and singing in the medium and high frequency ranges, a mass of about 10kg/m^2 (1.7lbs/ft^2) is sufficient, e.g., a 12mm (½in) plywood plate. If the effective frequency range is expanded to bass instruments, a mass of about 40kg/m^2 (1.7lbs/ft^2) has to be aspired, e.g., 36mm [1.5in] chipboard. With reflectors additionally suspended above the performance zone, the statically admissible load often plays a restrictive role for the possible mass of the reflectors. For spoken performances an area-related mass of 5 to 7kg/m^2 (0.2 to 0.3lbs/ft^2) may still produce acceptable results, to which effect plastic mats of high surface density are suitable. The additional room-acoustical measure usually employed for enhancing the sound reflection of bass instruments with music performances consists in appropriate wall surfaces, so that the installation of heavy panels can be abandoned. In this case an area-related mass of 20kg/m^2 (0.8lbs/ft^2) is sufficient.

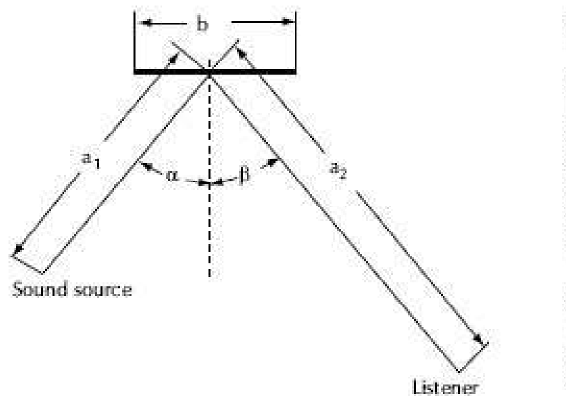


Figure 9-38. Geometrical sound reflections.

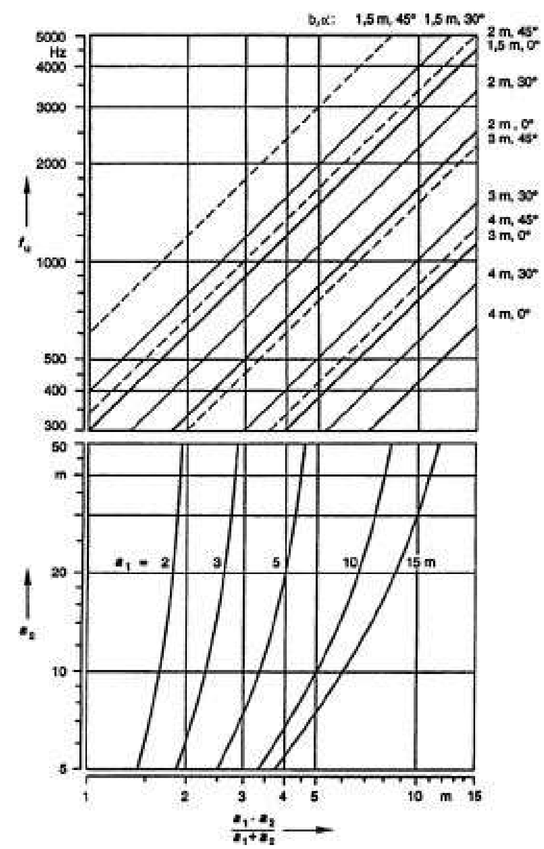


Figure 9-39. Minimum size of surfaces for geometrical sound reflections.

If a multiple reflection occurs close to edges of surfaces, there results, if the edge is at right angle to the surface, a sound reflection with a path parallel to the direction of the sound incidence, see Fig.

9-40. In corners this effect acquires a three-dimensional nature, so that the sound always gets reflected to its source, independently of the angle of incidence. With long travel paths it is possible that very disturbing sound reflections are caused at built-in doors, lighting stations, setoffs in wall paneling, which for the primary structure of a room are known as “theater echo,” see section 9.3.3.2.2.

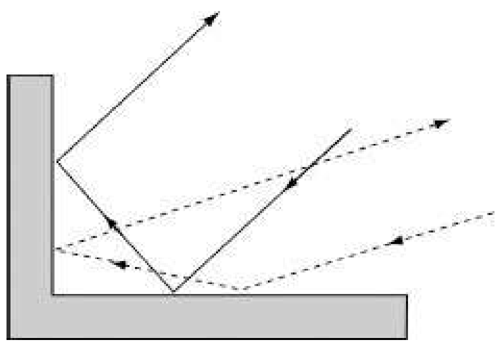


Figure 9-40. Multiple reflection in room edges.

9.3.4.2 Sound Reflection at Smooth Curved Surfaces

If the linear dimensions of smooth curved surfaces are much bigger than the wavelength of the effective sound components, the sound is reflected from these surfaces according to the laws of concentrating reflectors. Concavely curved two-dimensional (2D) or three-dimensional (3D) surface elements may under certain geometrical conditions lead to sound concentrations while convex curvatures always have a sound scattering effect.

For axis-near reflection areas (incident angle less than 45°) of a surface curved around the center of curvature M, it is possible to derive the following important reflection variants, see Fig. 9-41.

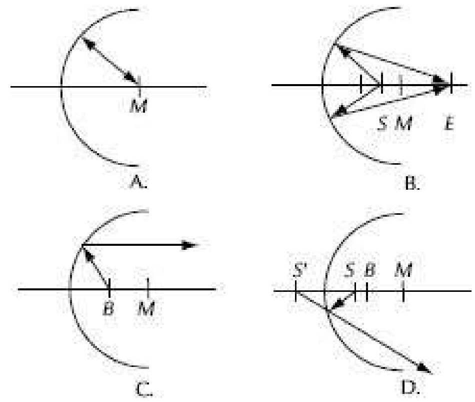


Figure 9-41. Sound reflection at smooth, curved surfaces.

Circular effect. The sound source is located in the center of curvature M of the reflecting surface, see Fig. 9-41A. All irradiated sound rays become concentrated in M after having covered the radius twice, so that a speaker may, for instance, be heavily disturbed by his own speech.

Elliptical effect. If the sound source is located between half the radius of curvature and the full radius of curvature in front of the reflecting surface, a second sound concentration point is formed outside the center of curvature, see Fig. 9-41B. If this second focus is located within the performance zone or the audience area, it is perceived as very disturbing, since distribution of the reflected sound results is very unbalanced. With extended sound sources like an orchestra, curved surfaces of this kind produce a heavily register-dependent sound balance.

Parabolic effect. If in a rather narrow arrangement the sound source is located at half the center of curvature, see Fig. 9-41C, the curved surface acts like a so-called parabolic reflector that generates an axis-parallel bundle of rays. This produces, on the one hand, a very uniform distribution of the reflected portion of the sound irradiated by the source, but on the other hand there occurs an unwanted concentration of noise from the audience area at the location of the sound source.

Hyperbolic effect. If the distance of the sound source from the curved surface is smaller than half the radius of curvature, see Fig. 9-41D, the reflecting sound rays leave the surface in a divergent fashion. But the divergence is less and thus the sound intensity at the listener's seat is higher than with reflections from a plain surface.[2] The acoustically favorable scattering effect thus produced is comparable to that of a convexly curved surface, but the diverging effect is independent of the distance from the curved reflecting surface.

9.3.4.3 Sound Reflections at Uneven Surfaces

Uneven surfaces serve as secondary structure of directional or diffuse sound reflections. This refers to structured surfaces with different geometrical intersections in the horizontal and vertical planes (rectangles, triangles, sawteeth, circle segments, polygons) as well as three-dimensional structures of geometrical layout (sphere segments, paraboloids, cones, etc.) and free forms (relievos, mouldings, coves, caps, ornaments, etc.). Also by means of a sequence of varying wall impedances (alternation of sound reflecting and sound absorbing surfaces), it is possible to achieve a secondary structure with scattering effect.

To characterize this sound dispersion of the secondary structure one makes a distinction between a *degree of diffusivity d* and a *scattering coefficient s*.

Typically for the homogeneity of the distribution of the sound reflections is the so-called frequency-dependent **degree of diffusivity *d***.[47]

$$d = \frac{\left(\sum_{i=1}^{n} 10^{\frac{L_i}{10}}\right)^2 - \sum_{i=1}^{n}\left(10^{\frac{L_i}{10}}\right)^2}{(n-1)\sum_{i=1}^{n}\left(10^{\frac{L_i}{10}}\right)^2} \qquad (9\text{-}53)$$

This way angle-dependent diffusion balloons may be generated. Depending from the number of n receiver positions hi-res level values are supplied to form the balloon.

High diffusion degrees close to one will be reached for half-cylinder or half-sphere structures. Nevertheless the **diffusion degree *d*** is more or less a **qualitative** measure to evaluate the homogeneity of scattering.

On the other side and as a **quantitative** measure to characterize the amount of scattered energy in contrast to the specular reflected or absorbed energy, the frequency-dependent **scattering coefficient *s*** is used.[50]

This scattering coefficient *s* is used in computer programs to simulate the scattered part of energy especially by using ray-tracing methods.

The coefficient *s* will be determined as the ratio of the non-specular, i.e., of the diffuse reflected, to the overall reflected energy.

$$s = \frac{diffuse - reflected - Energy}{overall - reflected - Energy} = \left(1 - \frac{geometric - reflected - Energy}{overall - reflected - Energy}\right) \qquad (9\text{-}54)$$

The measurement and calculation of the scattering coefficient under random sound impact takes place in the reverberation chamber.[39,48,49]

All these parameters don't say too much about the angular distribution of the reflected sound energy. But there exist many examples of rooms in which the secondary structure is intended to realize a directional reflection in which the angle of sound reflection does not correspond to the angle of sound incidence, as referred to the basic surface underlying the primary structure. In this case of **directional** sound reflection, one has to consider parameters determining, among other things, the diffusivity ratio D_{diff} and the maximum displacement d_α,[42] see Fig. 9.42.

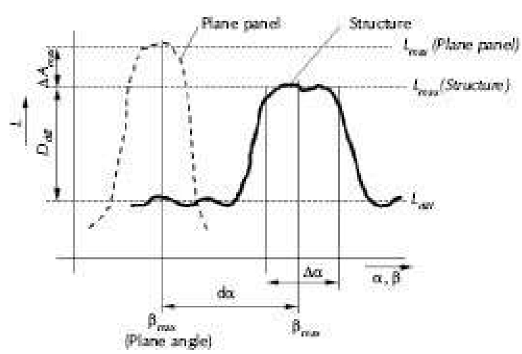


Figure 9-42. Parameters for characterizing the directivity of uneven surfaces.

- *Diffusivity ratio Ddiff*: sound pressure level difference between the directional and the diffuse sound components Lmax and Ldiff, respectively.

  → Characterizes the directional effect of a structure.
- *Attenuation of Maximum* Δa_{max}: sound pressure level difference between the directional reflection (local maximum, β*max*) of the structure, as compared with a plain surface.

→ Characterizes the sound pressure level of the reflection.

- *Displacement of Maximum d*α: angle between geometrical and directional reflections.

  → Characterizes the desired change of direction of the reflection.

- Angular range of uniform irradiation Δα: 3dB bandwidth of the reflection.

  → Characterizes the solid-angle range of uniform sound reflection.

Guide values D_{diff} for the octave mid-band frequency of 1000Hz are based on subjective sound-field investigations in a synthetic sound field, Table 9-9.

Table 9-9. Perception of Diffuse and Directed Sound Reflections as a Function of the Diffusivity Ratio D_{diff}

Perception	D_{diff} in dB
Ideal diffuse sound reflection	0
Diffuse sound reflection	< 3
Range of appropriate perception of diffuse and directed sound reflections	3–10
RT around 1.0s with energy-rich ceiling reflections	2–6
RT around 2.0s with energy-rich ceiling reflections	4–8
Spatial sound fields with low direct sound energy, but big part of lateral reflections	6–8
Sound fields with high direct sound energy, e.g., more distant listener groups	3–6
Low sound energy of ceiling reflections and big part of lateral sound	8–10
Directed sound reflection	>10
Ideal directed sound reflection	oo

An example for a sawtooth structure is shown in Fig. 9-43. This side-wall structure has at a sound incidence angle of 50° and a speech center frequency of about 1000Hz energy-rich directional sound reflections ($D_{diff} \geq 10$dB) with a displacement of maximum of $d\alpha = -20°$ (reflection angle 30°). Additionally directional and diffuse sound components being perceptible from about 3000 Hz (D_{diff} = 7 to 8dB). The attenuation of maximum Δa_{max} was 5dB at 1000Hz and 11dB at 5000Hz by comparison with a carrier panel of geometrical sound reflection.

Periodical structures of elements having a regular geometrical cut (rectangle, isosceles triangle, sawtooth, cylinder segment) may according to reference 3 and 40 show high degrees of scattering, if the following dimensions are complied with, see Fig. 9-44.

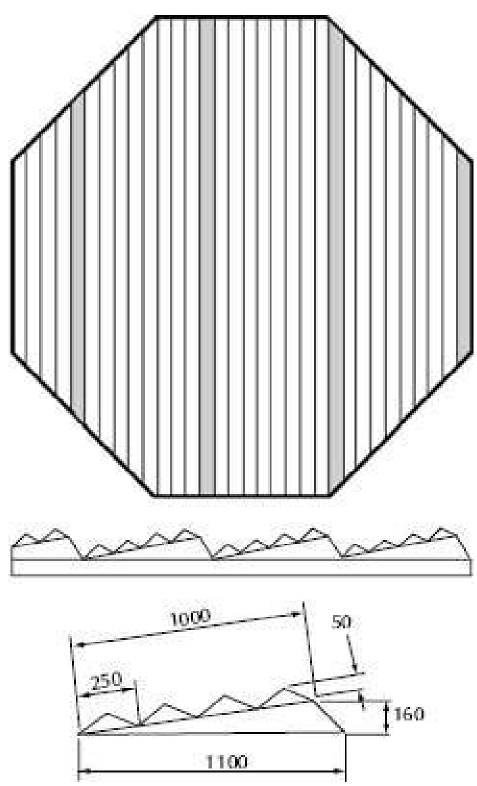
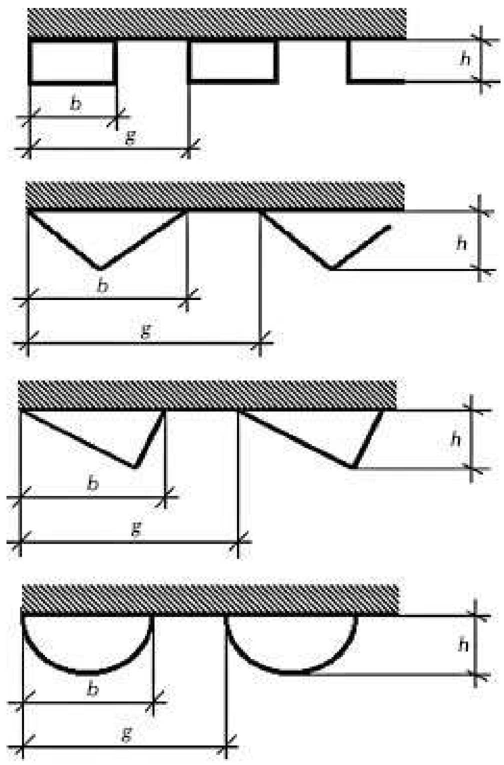


Figure 9-43. Example of an acoustically active sawtooth structure. Measures in mm.

For a diffuse scattering in the maximum of the speech-frequency range, the structure periods are thus about 0.6m (2ft), the structure widths between 0.1 to 0.4m (0.33 to 1.3ft), and the structure heights maximally about 0.3m (1ft). With rectangular structures the sound scattering effect is limited to the relatively small band of about 1 octave, with triangular structures to maximally 2 octaves. Cylinder segments or geometrical combinations can favorably be used for more broadband structures, see Fig. 9-43. In a wide frequency range between 500Hz and 2000Hz a cylinder segment structure results are sufficiently diffuse, if the structure width of about 1.2m (4ft) is equal to the structure period, and the structure height is between 0.15 and 0.20m (0.5 and 0.7ft). With a given structure height *h* and a given structure width *b* it is, according to Eq. 9-55, possible to calculate the required curvature radius *r* as

$$r = \frac{\left(\frac{b}{2}\right)^2 + h^2}{2h} \qquad (9\text{-}55)$$



Structure	Structure period g	Structure width b	Structure height h
Rectangle	$\sim(1-2)\lambda$	$\sim 0.2g$	$\sim 0.2g$
Isosceles triangle	$\sim(1-2)\lambda$	$\sim(0.5g-0.67)g$	$\sim(0.25-0.33)g$
Sawtooth	$\sim 2\lambda$		$\sim 0.33\lambda$
Cylinder segment	$\sim(1-2)\lambda$	$\sim(0.17-1.0)g$	$\sim(0.25-0.5)g$

Figure 9-44. Geometrical parameters at structures with rectangular, triangular, sawtooth-formed, and cylinder-segment-formed intersection.

A special form of a diffusely reflecting surface can be realized by "lining up" phase-grating structures of varying depths. Based on the effect of coupled $\lambda/2$ runtime units, these structures produce on the surface a local distribution of the reflection factor and hence of the sound particle velocity. Every component of this velocity distribution produces thereby a sound irradiation into another direction. If according to Schroeder[41] one distributes these reflection factors in accordance with the maximum sequences of the number theory (e.g., Barker code, primitive root diffusor PRD, square-law residual series QRD), and separates these trough structures from each other by thin wall surfaces, one obtains diffuse structures of a relatively broadband effect (up to 2 and more octaves), see Fig. 9-45. With perpendicular sound incidence, the minimum frequency limit f_{low} for the occurrence of additional reflection directions is approximately

$$f_{low} \approx \frac{c}{2d_{max}} \qquad (9\text{-}56)$$

where,

c is the velocity of sound in air in m/s (ft/s),

d_{max} is the maximum depth of structure in m (ft).

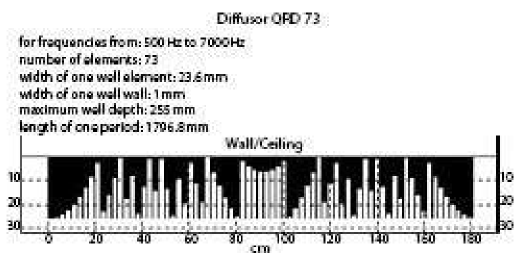


Figure 9-45. SCHOEDER diffusor with primitive root structure.

Nowadays calculation programs are available to calculate the scattering coefficients for angle-dependent sound impact by using Boundary Element Methods, see Fig. 9-46.

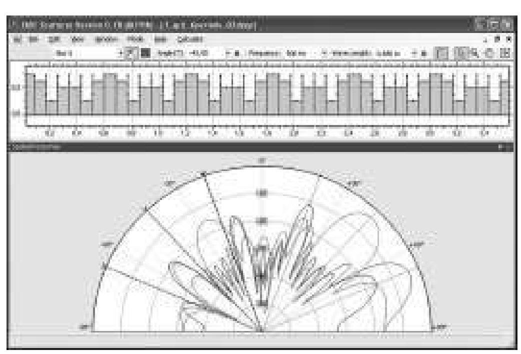


Figure 9-46. Boundary Element Methods (BEM) based software

tool for calculating scattering coefficients.

9.3.4.4 Sound Absorbers

Sound absorbers can occur in the shape of surfaces, built-in elements, pieces of furniture, or in the form of unavoidable environmental conditions, e.g., air, as well as arrangements conditioned by utilization of the room (e.g., spectators, decorations). According to their preferential effect in a determined frequency range one distinguishes on principle between:

- Absorbers in the low-frequency range between approximately 32Hz and 250Hz.
- Absorbers in the medium-frequency range between approximately 315Hz and 1000Hz.
- Absorbers in the high-frequency range between approximately 1250Hz and 12kHz.
- Broadband absorbers.

For acoustical characterization of a sound absorber there serves its frequency-dependent sound absorption coefficient a or the equivalent sound absorption area *A*. For an area of size *S* one determines the equivalent sound absorption area *A* as

$$A = \alpha S \tag{9-57}$$

The sound power W_i being incident on an area of size *S* of a sound-absorbing material or a construction is designated as sound intensity I_i, part of which is reflected as sound intensity I_r, and the rest is absorbed as sound intensity I_{abs}. The absorbed sound intensity consists of the sound absorption by *dissipation* (transformation of the sound intensity I_δ in heat, internal losses by friction at the microstructure or in coupled resounding hollow spaces), and of the sound absorption by *transmission* (transmission of the sound intensity I_τ into the coupled room behind the sound absorber or into adjacent structural elements).

$$\begin{aligned} I_i &= I_r + I_{abs} \\ &= I_r + I_\delta + I_\tau \end{aligned} \tag{9-58}$$

With the *sound reflection coefficient* ρ defined as

$$\rho = \frac{I_r}{I_i} \tag{9-59}$$

and the *sound absorption coefficient* α as

$$\begin{aligned} \alpha &= \frac{I_\delta + I_\tau}{I_i} \\ &= \frac{I_{abs}}{I_i} \end{aligned} \tag{9-60}$$

as a sum of the *dissipation coefficient δ*

$$\delta = \frac{I_\delta}{I_i} \tag{9-61}$$

and the *transmission coefficient* τ

$$\tau = \frac{I_\tau}{I_i} \tag{9-62}$$

Eq. 9-57 becomes

$$
\begin{aligned}
1 &= \rho + \delta + \tau \\
&= (\rho + \alpha)
\end{aligned} \tag{9-63}
$$

The transmission coefficient τ plays a role when considering the sound insulation of structural components. For non-movable, monocoque, acoustically hard material surfaces, e.g., walls, windows, it is according to Cremer[48] possible to consider the frequency dependence of the transmission coefficient as a low-pass behavior which surpasses the given value up to a limit frequency f_τ. With a negligible dissipation coefficient it is furthermore possible to equate the transmission coefficient τ numerically to the sound absorption coefficient α.

9.3.4.4.1 Sound Absorption Through Porous Materials

The effect of sound absorption is based essentially on the transformation of sound energy in thermal energy by air particles moving in open, narrow, and deep pores. Closed pores like those existing in foamed materials used for thermal insulation are unsuited for sound insulation. For characterizing the materials the so-called *Porosity* σ is used. This represents the ratio between open air volume V_{air} existing in the pores and the overall volume V_{tot} of the material

$$
\sigma = \frac{V_{air}}{V_{tot}} \tag{9-64}
$$

With a porosity of σ = 0.125, it is possible for high frequencies to obtain a maximum sound absorption coefficient of only α = 0.4, and with σ = 0.25 of α = 0.65. Materials with a porosity of σ ≥ 0.5 enable a maximum sound absorption coefficient of at least 0.9.

Usual mineral, organic, and naturally growing fibrous insulating materials feature porosities of between 0.9 and 1.0 and are thus very well suited for sound absorption purposes in the medium and high frequency ranges.[29]

Apart from porosity it is also the structure coefficient s and the flow resistance Ξ which influence the sound absorbing capacity of materials. The structure coefficient s can be calculated from the ratio between the total air volume V_{air} contained in the pores and the effective porous volume V_w

$$s = \frac{V_{air}}{V_w} \qquad (9\text{-}65)$$

The insulating materials most frequently used in practice have structure factors of between 1 and 2, i.e., either the total porous volume is involved in sound transmission or the "dead" volume equals the effective volume. Materials with a structure factor of the order of 10 show a sound absorption coefficient of maximally 0.8 for high frequencies.[8]

The flow resistance exerts an essentially higher influence on sound absorption by porous materials than the structure factor and the porosity. With equal porosity, for instance, narrow partial volumes offer a higher resistance to particle movement than wide ones. This is why the specific flow resistance R_s is defined as the ratio of the pressure difference before and behind the material with regard to the speed of the air flowing through the material v_{air}

$$R_s = \frac{\Delta p}{v_{air}} \qquad (9\text{-}66)$$

where,

R_s is the specific flow resistance in Pa s/m (lb s/ft^3),
Δp is the pressure difference in Pa (lb/ft^2),
v_{air} is the velocity of the passing air in m/s (ft/s).

With increasing material thickness the specific flow resistance in the direction of flow increases as well.

9.3.4.4.2 *Sound Absorption by Panel Resonances*

Thin panels or foils (vibrating mass) can be arranged at a defined distance in front of a rigid wall so that the withdrawal of energy from the sound field in the region of the resonance frequency of this spring-mass vibrating system makes the system act as a sound absorber. The spring action is produced herewith by the rigidity of the air cushion and the flexural rigidity of the vibrating panel. The attenuation depends essentially on the loss factor of the panel material, but also on friction losses at the points of fixation.[43] The schematic diagram is shown in Fig. 9-47, where d_L is the thickness of the air cushion and m' the area-related mass of the vibrating panel.

The resonance frequency of the vibrating panel mounted in front of a rigid wall with attenuated air space and lateral coffering is calculated approximately as

$$f_R \approx \frac{60^{*}}{\sqrt{m' d_L}} \qquad (9\text{-}67)$$

* 73 in U.S. units

where,
f_R is in Hz,

m' is in kg/m^2 (lb/ft^2),
d_L is in m (ft).

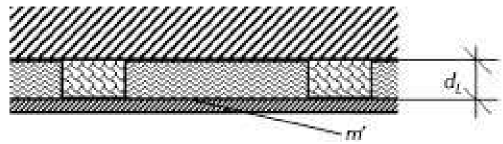


Figure 9-47. General structure of a panel resonator.

In practical design one should take into account the following:

- According to reference 43 the loss factor of the vibrating panel should be as high as possible.
- The clear spacing of the coffering should in every direction be in general smaller than 0.5 times the wavelength in case of resonance, but not fall short of 0.5m (1.7ft).
- The minimum size of the vibrating panel must not fall short of 0.4m^2 (4.3ft^2).
- The air-space damping material should be attached to the solid wall so that the panel vibration is not impaired in any way.
- The sound absorption coefficient depends on the Q factor of the resonance circuit and amounts at the resonance frequency to between 0.4 to 0.7 with air-cushion damping and to between 0.3 to 0.4 without air-cushion damping. At an interval of one octave from the resonance frequency one must reckon that the sound absorption coefficient is halved.

An effective method for increasing the acoustically effective resonance frequencies of panel resonators consists in reducing the vibrating mass of heavy panels by means of holes arranged in defined patterns. In this case the correlations are governed by analogous regularities, if the area-related mass m' of the panels is replaced by the effective hole mass m'_L. For circular holes of radius R and a hole-surface ratio e, see Fig. 9-48, the hole mass is calculated as

$$m'_L = 1.2^{**}\frac{l^*}{\varepsilon} \qquad (9\text{-}68)$$

** 19.2 for U.S. system where

where,

m'_L is the area-related air mass of circular openings in kg/m^2 (lb/ft^2),

l^* is the effective panel thickness with due consideration of the mouth correction of circular openings of radius R in m (ft)

$$l^* \approx 1 + \frac{\pi}{2}R\,, \qquad (9\text{-}69)$$

ε is the hole-area ratio according to Fig. 9-48 for circular openings

$$\varepsilon = \frac{\pi \cdot R^2}{a \cdot b}\,. \qquad (9\text{-}70)$$

Provided the hole diameters are sufficiently narrow, the damping material layer arranged between the perforated panel and solid wall can be replaced by the friction losses produced in the openings. By using transparent materials, e.g., glass, it is possible to fabricate optically transparent, so-called micro-perforated absorbers. The diameters of the holes are in the region of 0.5mm (0.02in) with a panel thickness of 4 to 6mm (0.16 to 0.24in) and a hole-area ratio of 6%. For obtaining broadband sound absorbers, it is possible to resort to variable perforation parameters, e.g., scattered perforation, varying thickness of the air cushion and composite absorbers combined of various perforated panels.

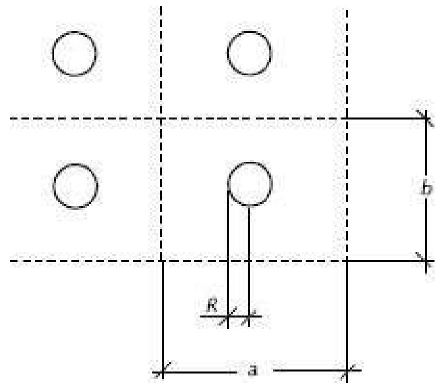


Figure 9-48. Hole-area ratio of perforated panels with round holes.

A very recent development are micro-perforated foils of less than 1mm (0.04in) thickness which also produce remarkable absorption when placed in front of solid surfaces. The transparent absorber foil can be advantageously arranged in front of windows either fixed or also as roll-type blind in single or double layer.[44]

9.3.4.4.3 Helmholtz Resonators

Helmholtz resonators are mainly used for sound absorption in the low-frequency range. Their advantage, as compared with panel absorbers, see section 9.3.4.4.2, lies in their posterior variability regarding resonance frequency and sound absorption coefficient as well as in the utilization of existing structural cavities which must not necessarily be clearly visible. According to Fig. 9-49, a Helmholtz resonator is a resonance-capable spring-mass system which consists of a resonator volume V acting as an acoustical spring and of the mass of the resonator throat characterized by the opening cross section S and the throat depth I. In resonance condition and if the characteristic impedance of the resonator matches that of the air, the ambient sound field is deprived of a large amount of energy. To this effect a damping material of a defined specific sound impedance is placed in the resonator throat or in the cavity volume.

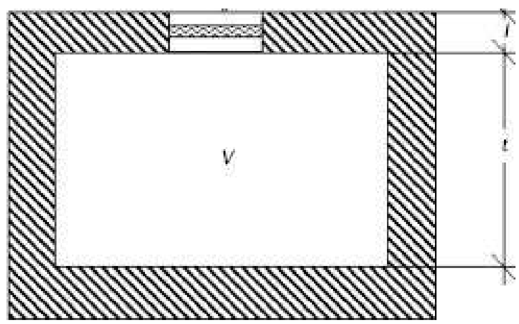


Figure 9-49. General structure of a Helmholtz resonator.

The resonance frequency of a Helmholtz resonator is generally calculated as

$$f_R = \frac{c}{2\pi}\sqrt{\frac{S}{V(l+2\Delta l)}} \qquad (9\text{-}71)$$

where,

c is the speed of sound in air, approximately 343m/s (1130ft/s),

S is the cross-sectional area of the resonator in m^2 (ft^2),

V is the resonator volume in m^3 (ft^3),

l is the length of the resonator throat in m (ft),

$2\Delta l$ is the mouth correction.

In case of a square opening

$$2\Delta l \approx 0.9a\,, \qquad (9\text{-}72)$$

where,

a is the edge length of the square opening.

For circular openings the resonance frequency f_R in Hz is calculated approximately from Eq. 9-71:

$$f_R \approx \frac{100^* \cdot R}{\sqrt{V(l + 1.6R)}} \qquad (9\text{-}73)$$

* 30.5 for U.S. units

where,

R is the radius of the circular opening in m (ft),

V is the resonator volume in m^3 (ft^3),

l is the length of the resonator throat in m (ft).

9.3.4.4.4 Sound Absorption by the Audience

The efficiency of sound absorption by the audience depends on many factors, like for instance the occupation density, the spacing of seats and rows, the clothing, the type and property of the seats, the sloping of tiers and the distribution of the persons in the room. In a diffuse sound field the location of the sound source towards the audience area is of minor importance in this regard. Fig. 9-50 shows a survey of the values of the equivalent sound absorption area per person for a variety of occupation densities and seating patterns in a diffuse sound field. Since in many types of rooms the reverberation time for medium and high frequencies is determined almost exclusively by the sound absorption of the audience, one has to reckon with a rather high error rate, if the range of dispersion of the factors influencing the sound absorption capacity of the audience is to be taken into account when determining the reverberation time, see Fig. 9-50. A still wider range of dispersion of the sound absorption area occurs with the musicians and their instruments, see Fig. 9-51. The unilateral arrangement of the listener's or musician's areas prevailing in most rooms tends to disturb the diffusivity of the sound field heavily so that the above-mentioned measured values may be faulty, Figs. 9-50 and 9-51.

Especially with an almost plain arrangement of the audience and performance areas there occurs for the direct sound and the initial reflections a frequency-dependent additional attenuation through the grazing sound incidence on the audience area. This is intensified by the fact that the sound receivers, i.e., the ears, are located in this indifferent acoustical boundary region, so that the influence of this additional attenuation becomes particularly relevant for the auditory impression.

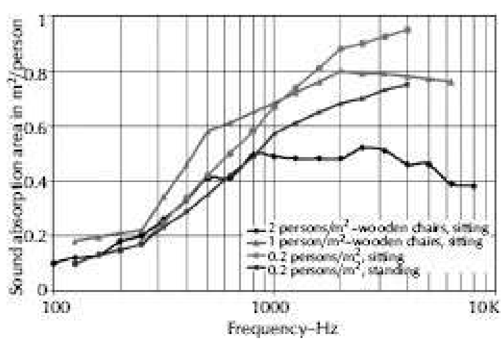


Figure 9-50. Equivalent sound absorption area in m^2/person of audience.

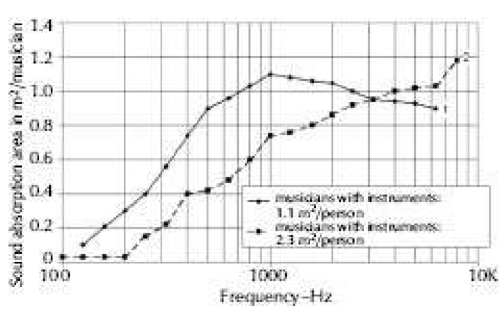


Figure 9-51. Equivalent sound absorption area in m^2/person of musicians.

According to Mommertz[45] this effect of additional attenuation can be attributed to three causes:

1. The periodical structure of seat arrangement compels a guided wave propagation for low frequencies. In the frequency range between 150Hz and 250Hz, this additional attenuation causes a frequency-selective level dip which is designated as *seat dip effect*. An example is given in Fig. 9-52 for a frequency of about 200Hz from reference 45.
2. The scattering of sound at the heads produces an additional attenuation especially in the frequency range between 1.5kHz and 4kHz, which is designated as "head dip effect," Fig. 9-52. The magnitude of the effect depends largely on the seat arrangement and the orientation of the head with regard to the sound source.
3. In combination with the incident direct sound, the scattering at the shoulders produces a very broadband additional attenuation through interference. According to reference 45 it is possible to define a simple correlation between the so-called elevation angle γ, see Fig. 9-53, and the sound level reduction ΔL in the medium-frequency range at ear level of a sitting person.[45]

$$\Delta L = -20\log(0.2 + 0.1\gamma) \qquad (9\text{-}74)$$

where,

ΔL is in dB,

γ is in degrees, $\gamma < 8°$.

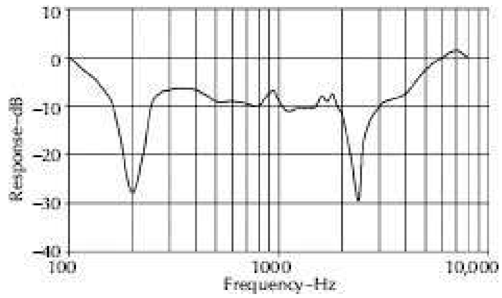


Figure 9-52. Measured quantity spectrum of the transfer function above a plain audience arrangement.[45]

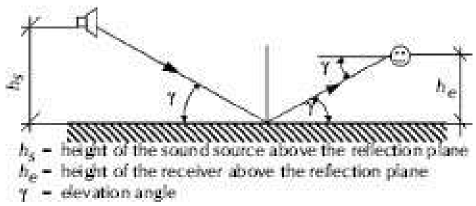


Figure 9-53. Geometric data for determining the elevation angle above a sound reflecting plane.

Fig. 9-54 shows a graphical representation of the correlation resulting from Eq. 9-72. One sees that with a plain arrangement of source and receiver the resulting level reduction may be up to about 14dB, whereas an elevation angle of 7° suffices already to cut the level reduction to a negligible amount of less than 1dB. The reflection plane defined in Fig. 9-53 lies herewith h_L = 0.15m (0.5ft) below ear level, i.e., at shoulder level of a sitting person (approximately 1.05m [3.5ft] above the upper edge of floor). According to reference 45, the additional attenuation depends herewith only on the elevation angle, no matter if tier sloping is effected in the audience area or in the performance area.

Fig. 9-55 shows the influence of the height of the sound source above the ear level of a person sitting in a row at a distance source-receiver of 15m (50ft), on the frequency-dependent additional attenuation caused according to reference 45 by a grazing sound incidence over an audience. The receiver level is herewith 1.2m (4ft) above upper edge of floor, the height of the source above floor is represented in this example as being 1.4m and 2.0m (4.6ft and 6.6ft). With a level difference of only 0.2m (0.66ft) between source and receiver one can clearly recognize the additional timbre change of the direct sound component and of the initial reflections by attenuation in the low and medium frequency ranges, whereas with a level difference of 0.8m (2.6 ft) the sound level attenuations get reduced to below 3dB.

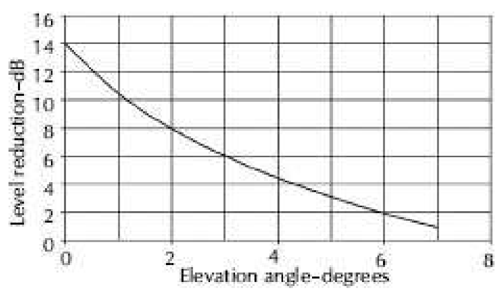


Figure 9-54. Sound pressure level reduction by sound dispersion at shoulder level of sitting persons as a function of the elevation angle.

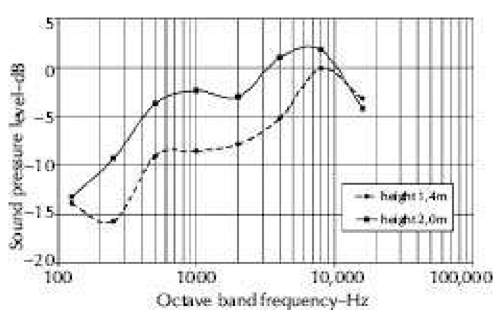


Figure 9-55. Influence of the vertical distance ear levelsound source level on the sound pressure level at a listener seat referred to free field radiation.

9.4 Variation of Room Acoustics by Construction

or Electro-acoustic Methods

9.4.1 Variable Acoustics

The manipulation of room acoustic properties is universally known by the term "vario-acoustics" as well as "variable acoustics." What variable manipulations are possible in variable acoustics? The primary structures (volume, dimensions, shape) and the secondary structures (reflective, diffuse, absorptive) of a room have a huge influence on its acoustic properties.

Acoustical parameters describing the acoustical overall impression of rooms are determined by the utilization function, see section 9.3.1. If this function is unambiguously defined, as is the case, e.g., with auditoriums and concert halls for symphonic music, the result corresponds, provided an appropriate planning was carried out, to the utilization-relevant roomacoustical requirements. Things look quite different, however, for rooms having a wide utilization range, i.e., so-called multi-purpose halls. For speech and music performances which use exclusively sound reinforcement, a short reverberation time with little rise in the low frequency range as well as a reduced spaciousness of the natural sound field are desirable. For vocal and music performances with mainly natural sound sources, however, a longer reverberation time with enhanced spaciousness is aspired. The timbre of the room should herewith show more warmth in the lower frequency range. As regards their roomacoustical planning, most multi-purpose halls feature a compromise solution which is harmonized with their main utilization variant and does not allow any variability. The acoustically better solution lies in the realization of "variable acoustics" within certain limits. This aim can be achieved by architectural or electro-acoustical means.

Another range of application for purposefully variable room-acoustical parameters is by influencing the reverberation time, clarity and spaciousness of rooms which owing to their form and size meet with narrow physical boundaries in this respect. This concerns mainly rooms with too small a volume index, see section 9.3.2.1, or such containing a large portion of sound-absorbing materials. Architectural measures for achieving desirable modifications of room-acoustical parameters are applicable here only to a limited extent, since they are bound to implement methods allowing a deliberate modification of the temporal and frequency-dependent sound energy behavior of the sound field. The effectiveness of these methods is herewith determined by the correspondingly relevant sound energy component of the room-acoustical parameter. Achieving a desired reverberation-time and spaciousness enhancement requires, for instance, a prolongation of the travel path of sound reflections and a reduction of the sound absorption of late sound reflections (enhancement of reverberant energy). In this respect, more favorable results can be obtained by electroacoustical means, particularly because in such rooms the sound-field structure does not contribute essentially to the manipulated parameters. From a practical point of view section 9.4 is mainly dedicated to the presentation of electronic procedures for reverberation-time prolongation. Equivalent architectural measures will be explained only on fundamental lines.

9.4.1.1 Fundamentals of Variable Acoustics

In the planning and realization of measures enabling variation of roomacoustical parameters, it is necessary to comply with essential aspects, so that the listener's subjective listening expectation in the room is not spoiled by an excessive range of variability:

1. The measures viable for realizing variable acoustics by architectural as well as electro-acoustical means can be derived from the definitions of the room-acoustical parameters to be varied, see section 9.2.2. Additional sound reflections arriving exclusively from the direction of the sound source surely enhance clarity, but boost spaciousness as little as an additional lateral sound energy prolongs reverberation time. Spaciousness-enhancing and reverberation-time prolonging sound reflections must essentially impact on the listener from all directions of the room. By means of appropriately dimensioned additional architectural measures it is possible to achieve good results in this respect. Realization of the same, however, often implies high technical expenditure and costs. For influencing the reverberation time, for instance, these include the coupling or uncoupling of additional room volumes or the prolongation of the travel path of sound reflections with simultaneous reduction of the sound absorption of late sound reflections. A desired reduction of reverberation time and spatial impression can be achieved by means of variable sound-absorbing materials (curtains, slewable wall elements of different acoustic wall impedance) which, however, have to be effective over the whole frequency range required by the performance concerned.
2. The coupling of acoustically effective room volumes has to be done in such a way that these get acoustically to form a unity with the original room. Otherwise there occur disturbing effects like timbre changes and double-slope reverberation-time curves. Incorrect dimensioning often results, owing to an

acoustical orientation towards the additional room volume, in a heavily frequency-dependent spaciousness of the decay process in the sound field. The frequency-dependent reverberation time of the additional room volume must be a bit longer than or at least as long as that of the original room.

In the opposite case of reducing the reverberation time by uncoupling the additional room volume it is for the remaining room volume necessary to provide the sound-field structure required for the desired variation. For instance, there is more sound energy to be allocated to the initial reflections and in the decay process—which is now to be supplied with less sound energy—there must not occur any irregularities.

3. The variation depth achievable by means of variable acoustics must be acoustically perceptible to a significant degree. The distinctive threshold of, e.g., subjectively perceived reverberation time changes is not independent of the absolute value of the reverberation time. Variations of 0.1s to 0.2s are at medium frequencies and a reverberation time of up to 1.4s to 1.5s is subjectively less clearly perceived than above this limit value. Thus a reverberation-time prolongation from 1.0s to 1.2s attained with much technical effort is almost not audible, whereas one from 1.6s to 1.8s is already significantly audible.
4. The listening experience has to tally with the overall visual impression of the room—too heavy deviations are perceived as disturbing and unnatural. This aspect has to be taken into account especially with small room volumes, if an excessively long reverberation time is produced by an electronic enhancement system (except for acoustic disassociation effects).
5. The sound-field structure of the original room has to remain unchanged also if measures of variable acoustics are

implemented. Additionally modified sound reflections have to correspond to the frequency and time structure of the room. This aspect holds true for architectural as well as electro-acoustical measures, e.g., for the reverberation enhancement. Coupled additional room volumes must thus not involve any distinctive timbre changes compared with the main room. Electro-acoustical room sound simulators with synthetically produced sound fields are allowed to change the transmission function only in compliance with the original room, except if alienation effects are required for special play-ins.

6. An enhancement of reverberation time and spaciousness is possible only within permissible boundaries in which the overall acoustic impression is not noticeably disturbed. This boundary is all the lower the more the manipulation makes the sound field structure deviate from that of the original room.

Aspects to be considered with the realization of variable acoustics. In keeping with the envisaged target, the following main realization objectives can be formulated for variable acoustics:

1. Large room volume (large volume index) or reverberant rooms.

   - **Task of variable acoustics**: Enhancement of clarity and definition. Reduction of reverberation time and spaciousness.
   - **Architectural solution**: Apart from an appropriate tiering arrangement of the sound sources, variable ceiling panels and movable wall elements have to be placed at defined distances for enhancing the clarity of music and the definition of speech. Modified inclinations of walls, built-in concert shells near stage areas in theatres, etc., create new primary reflections that are in harmony with the variants of purpose.

Broadband sound absorbers in the shape of variable mechanisms for low frequencies, combined with curtain elements or slewable boundary elements of differing acoustic wall impedance reduce reverberation time and diminish spaciousness. When arranging variable sound absorbers it is necessary to pay attention to the frequency dependence of these elements. Slots between the installed slewable wall elements may, depending on the position of the elements, function as unwanted additional bass absorbers. In case of exclusive use of curtain arrangements the low-frequency fraction is at a disadvantage giving rise to a brilliance-poor sound pattern. An effective broadband reduction of the reverberation time can be achieved by deliberately influencing the late sound reflection mechanism. This may be realized by means of mobile room dividing wall parts which shorten the travel distances of sound reflections at the points of maximum length or width of the room and direct the reflections toward the sound-absorbing listener area. To this effect it is also possible to perform a room reduction, e.g., by detaching the room volume above or below a balcony or a gallery.

- **Electronic solution**: An additional “electronic architecture” system serves for enhancing definition and clarity. Reducing reverberation time or diminishing spaciousness however, are not possible by electronic means.

2. Little room volume (small volume index) or highly sound-absorbent rooms.

- **Task of variable acoustics**: Enhancement of reverberation time and spaciousness.

- **Architectural solution**: One solution for enhancing reverberation time consists in the coupling of additional room volumes in acoustical unity with the main room. By means of a purposive sound reflection guidance at the point of maximum room length or width it is possible, moreover, to realize long travel paths by letting the sound repeatedly be reflected between the room boundary faces and the room ceiling and thus having it belatedly absorbed by the listener area (cf. large concert hall in the Neues Gewandhaus Leipzig). This way it is first of all the early decay time, which is mainly responsible for the subjectively perceived reverberation duration, which is prolonged.
- **Electronic solution**: Influencing reverberation time and spaciousness is quite possible by electronic means, if the physical and auditory-psychological limitations are observed. Viable solutions are described in detail in section 9.4.2.

In general variable-acoustics is steadily losing ground because of its high costs and low effect in comparison with the use of correctly designed sound-systems.

9.4.2 Electronic Architecture

Establishing good audibility in rooms, indoors as well as in the open air, has been and remains the object of room acoustics. This is the reason why room acoustics is called "architectural acoustics" in some countries.

The architectural measures have far-reaching limitations. These shortcomings are:

- The sound source in question has only a limited sound power rate.

- Changes of room acoustics may make huge changes in the architectural design and thus cannot always be optimally applied.
- The measures regarding room acoustics may make a considerable amount of constructional changes and these can only be done optimally for one intended purpose of the room.
- The constructional change, despite its high costs, results in only a very limited effect.

Because of these reasons sound systems are increasingly being used to influence specific room acoustic properties, thus improving audibility. This holds true regarding an improvement of intelligibility as well as of spaciousness. So one can speak of a good acoustic design if one cannot distinguish, when listening to an event, whether the sound quality is caused only by the original source or by using an electroacoustic installation.

Another task of sound installation techniques consists of electronically modified signals that remain uninfluenced by specific properties of the listener room. It is necessary to suppress, as far as possible, the acoustic properties of the respective listener room by using directed loudspeaker systems. It is also possible to create a "dry" acoustic atmosphere by using suitable supplementary room-acoustic measures.

Reverberation (the reverberation time of a room) cannot be reduced by means of sound systems. At the typical listener seat the level of the direct sound is of great significance. Also "short-time reflections", enhancing intelligibility of speech and clarity of music, can be provided by means of sound reinforcement.

The following sound-field components can be manipulated or generated:

- Direct sound.
- Initial reflections with direct-sound effect.
- Reverberant initial reflections.
- Reverberation.

For this reason electronic techniques were developed that opened up the possibility of increasing the direct or reverberation time and energy in the hall, i.e., directly influencing the acoustic room properties.

Such a method for changing the room-acoustic properties of rooms is now called the application of **"electronic architecture."**

9.4.2.1 Use of Sound Delay Systems for Enhancing Spaciousness

These procedures act in particular on the sound energy of the initial reflections affecting the reverberant sound.

9.4.2.1.1 Ambiophony

This procedure which is already obsolete by now makes use of delaying devices reproducing not only the discrete "initial reflections," but also the "reverberation tail." The "reflection sequences" have herewith to be chosen in such a way that there will be no comb-filter effects, as for instance flutter echoes, as are produced with "impulsive" music motifs. The functioning of a simple ambiophonic system can be described as follows: To the direct sound emanating directly from the original source and directly irradiated into the room, there are admixed delayed signals produced by an adequate sound delaying system (in the initial stages this was just a magnetic sound recording system) which are then irradiated like reflections arriving with corresponding delay from the walls or the ceiling. This requires additional loudspeakers appropriately distributed in the room for irradiating the delayed sound as diffusely as possible. For further delaying the sound it is possible to arrange an additional feedback from the last output of the delay chain to the input. A system of this kind was first suggested by KLEIS[51] and was installed in several large halls.[52,53]

9.4.2.1.2 ERES (Electronic Reflected Energy System)

This procedure was suggested by Jaffe and is based on a simulation of early reflections used for producing so-called reverberant-sound-efficient initial reflections,[54] see Fig. 9-56.

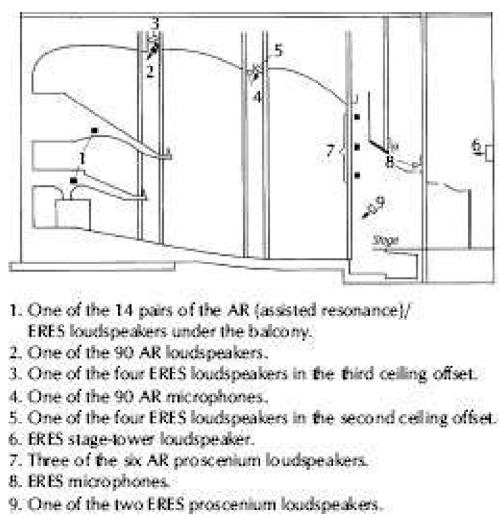


Figure 9-56. ERES/AR system in the Sivia Hall in the Eugene Performing Arts Center, Eugene, Oregon, USA.

Thanks to the arrangement of the loudspeakers in the walls of the stage-near hall area and to the variability range available for delay, filtering and level regulation of the signals supplied to them, adapted lateral reflections can be irradiated. The spatial impression can thus be amply influenced by simulating an acoustically wider portal by means of a longer reverberation time or a narrower portal by using a shorter reverberation time. Owing to the thus enabled capability of:

- Adaptation to acoustical requirements.
- Simulation of different hall sizes.
- Optimization of definition and clarity.

Jaffe and collaborators speak of “electronic architecture.” It is certainly true that this selective play-in of reflections does simulate roomacoustical properties the room in question is devoid of, so as to compensate shortcomings in its room-acoustical structure. After installing the first system of this kind in the “Eugene Performing Arts Center” in Oregon,[55] Jaffe-Acoustics have installed further ones in a large number of halls in the USA, Canada and other countries.

The electronic delay procedure in sound reinforcement systems has meanwhile become a general practice all over the world and is now the standard technique used for the play-in of delayed signals (e.g., for simulating late reflections). In this sense one may well say that “electronic architecture” is used in all instances where such reflections are used on purpose or unintentionally.

9.4.2.2 Travel-Time-Based Reverberation-Time Enhancing Systems

This procedure is mainly used for enhancing the late reverberant sound energy combined with an enhancement of the reverberation time.

9.4.2.2.1 Assisted Resonance

For optimizing the reverberation time in the "Royal Festival Hall" built in London in 1951, Parkin and Morgan[56,57] suggested a procedure permitting an enhancement of the reverberation time especially for low frequencies, see Fig. 9-57.

Parkin and Morgan proceeded on the assumption that in any room there exist a multitude of Eigenfrequencies which give rise to standing waves with nodes and antinodes decaying by an e-function according to the absorption characteristics of the surface involved. This decay process is characteristic for the reverberation time of the room at the corresponding frequency. Any standing wave has its specific orientation in space and a microphone is installed at the point where a sound pressure maximum (vibration antinode) occurs for a given frequency. The energy picked up from the microphone is supplied via an amplifier to a loudspeaker installed at a distant antinode of the same standing wave, so that the energy lost by absorption is compensated. The energy at that frequency can thus be sustained for a longer period (assisted resonance). By enhancing the amplification it is possible to considerably prolong the reverberation time for this frequency (until feedback sets in). Thanks to the spatial distribution of the irradiating loudspeakers this applies accordingly to the spatial impression.

These considerations hold true for all Eigenfrequencies of the room. The arrangement of the microphones and loudspeakers at the locations determined by the antinodes of the individual eigenfrequencies may, however, be difficult. The microphones and loudspeakers are therefore installed at less critical points and driven via phase shifters. In the transmission path there are additionally inserted filters (Helmholtz resonators, bandwidth approximately 3Hz) which allow the transmission channel to respond only at the corresponding eigenfrequency. Care should be taken that the irradiating loudspeakers are not arranged at a greater distance from the performance area than their corresponding microphones, since otherwise the first arrival of the reverberant signal may produce mislocalization of the source.

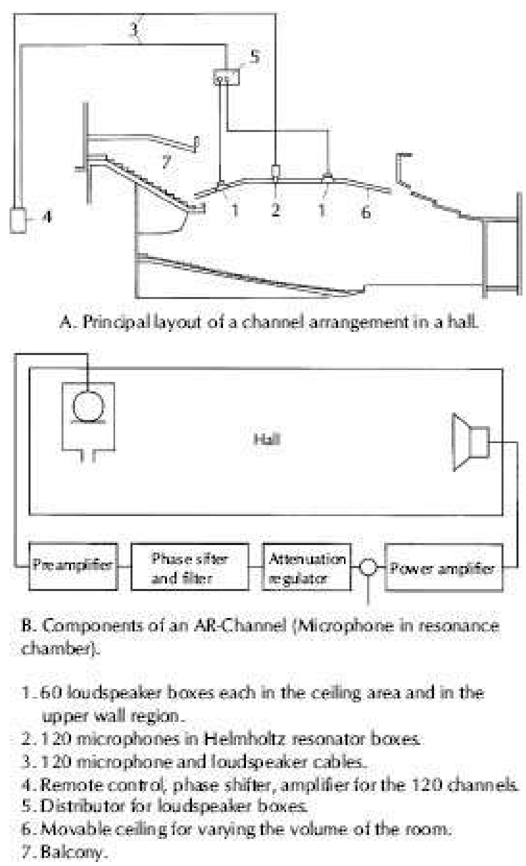


Figure 9-57. Assisted resonance system.

This procedure which has meanwhile become obsolete was installed in a large number of halls. In spite of its high technical expenditure and the fact that the system required can be used only for the “assisted resonance”, it was for a long period one of the most reliable solutions for enhancing the reverberation time without affecting the sound, particularly at low frequencies.

9.4.2.2.2 Multi-Channel-Reverberation, MCR

Using a large number of broadband transmission channels whose amplification per channel is so low that no timbre change due to commencing feedback can occur, was suggested first by Franssen.[58] While the individual channel remaining below the positive feedback threshold provides only little amplification, the multitude of channels are able to produce an energy density capable of notably enhancing the spatial impression and the reverberation time.

The enhancement of the reverberation time is determined by

$$\frac{T_m}{T_o} = 1 + \frac{n}{50} \qquad (9\text{-}75)$$

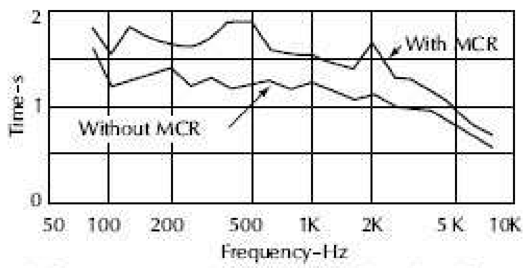
If the reverberation time is for instance to be doubled (which means doubling the energy density), there are n = 50 individual amplification chains required. Oshmann[59] has investigated in an extensive paper the functional principle of these loudspeaker systems and has shown that the prognosticated results regarding enhancement of the reverberation time cannot be achieved in practice. He quotes also the fact that Franssen "did not sufficiently consider the cross couplings between the channels" as a possible reason for the deviation from theory.[1]

A system technologically based on this procedure is offered by Philips under the name of "Multi-Channel Amplification of Reverberation System" (MCR). It serves for enhancing reverberation and spaciousness.[60] According to manufacturer's specifications a prolongation of the average reverberation time from approximately 1.2 to 1.7s is achieved for 90 channels. Even longer reverberation enhancements are said to be possible. There exist numerous implementations in medium-sized and large halls (the first was in the POC Theater in Eindhoven, Fig. 9-58).

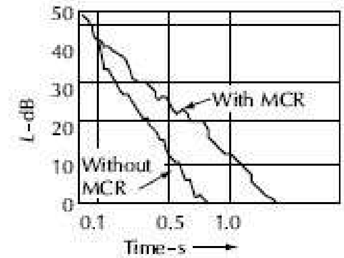
9.4.2.3 Modern Procedures for Enhancing Reverberation and Spaciousness

9.4.2.3.1 Acoustic Control System (ACS)

This procedure was developed by Berkhout and de Vries at the University of Delft.[61] Basing on a wave-field synthesis approach (WFS) the authors speak of a "holographic" attempt for enhancing the reverberation in rooms. In essence it is really more than the result of a (mathematical-physical) convolution of signals captured by means of microphones in an in-line arrangement (as is the case with WFS) with room characteristics predetermined by a processor, which produces in the end a new room characteristic with a new reverberation time behavior, see Fig. 9-59.



A. Frequency response of the reverberation time with and without MCR.



B. Reverberation behavior at 400 Hz.

Technical data of the system: hall 3100 m^3, stage 900 m^3.
90 channels (preamplifier, filter, power amplifier).
90 microphones at the ceiling.
110 loudspeakers in the side walls, in the ceiling and under the balcony.
Remote control of the reverberation in 10 steps.

Figure 9-58. MCR system in the POC Theater in Eindhoven.

The block diagram shows the principle of the ACS circuit for a loudspeaker-microphone pair. One sees that the acoustician formulates the characteristics of a desired room, e.g., in a computer model, transfers these characteristics by means of suitable parameters to a reflection simulator and convolutes these reflection patterns with the real acoustical characteristics of a hall. Fig. 9-60 shows the complete block diagram of an ACS system.

Unlike other systems, the ACS does not use any feedback loops thus timbre changes owing to self-excitation phenomena should not be expected. The system is functioning already in a series of halls in the Netherlands, Great Britain and the USA.

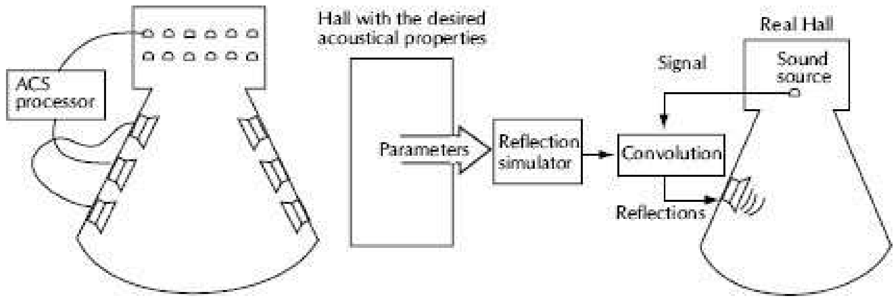


Figure 9-59. Basic block diagram of the Acoustic Control System (ACS) illustrated for a loudspeaker–microphone pair.

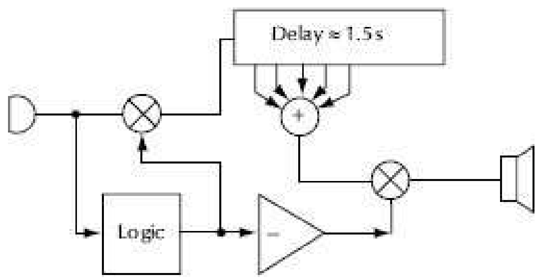


Figure 9-60. Principle of the Reverberation on Demand System (RODS).

9.4.2.3.2 Reverberation on Demand System, RODS

With this system a microphone signal is picked up near the source and passed through a logical switching gate before reaching a delay line with branched members. This output is equipped with a similar gate. A logical control circuit opens the input gate and closes the output gate when the microphone signal is constant or rising. Vice versa, it closes the input gate and opens the output gate when the microphone signal is falling, see Fig 9-60.[62]

An acoustical feedback is thus avoided, but this system fails to enhance the lateral energy with continuous music, which makes it unsuitable for music performances. It is no longer used.

9.4.2.3.3 LARES

The LARES System by the Lexicon Company uses modules of the standardized Room Processor 480L which, when fed by special software, allows to simulate desired decay curves, Fig. 9-61. Also here a large number of loudspeakers are required in the wall and ceiling areas. The input signals are picked up by just a few microphones in an area near the source.[63,64] On account of the "time-variant" signal processing (large quantity of independent time-variant reverberation devices) the adjustment of reverberation times is not exactly repeatable. Common computer-controlled measuring software based, e.g., on MLS, is thus unable to measure decay curves. Apart from the ASC system, LARES installations are very widespread in Europe and the USA. Well known are the systems installed in the Staatsoper Berlin, the Staatsschauspiel Dresden, and the Seebühne (floating stage) in Mörbisch/Austria.

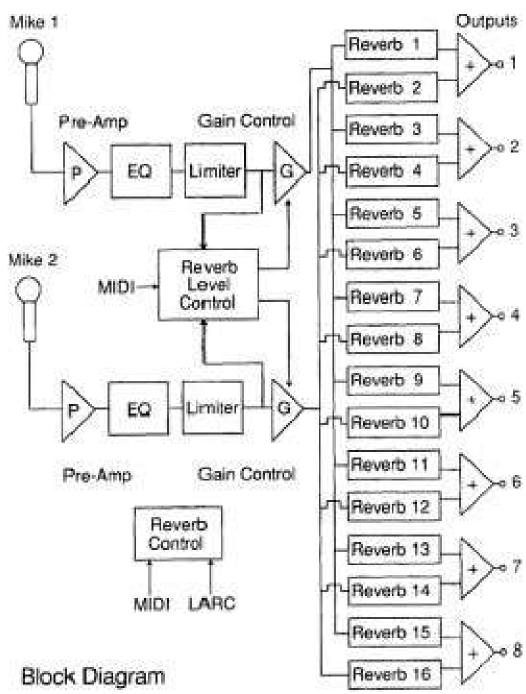


Figure 9-61. LARES block diagram.

9.4.2.3.4 System for Improved Acoustic Performance (SIAP)

The basic principle of SIAP consists in picking up the sound produced by the source by means of a relatively small number of microphones, processing it appropriately (by means of processors which convolute, that is overlay electronically the room-acoustical parameters of a room with target parameters) and then feeding it back into the hall by an adequate number of loudspeakers, see Fig. 9-62. The aim is to produce desired natural acoustical properties by electronic means. For obtaining spatial diffusivity a large number of different output channels are required. Moreover, the maximally attainable acoustic amplification is dependent on the number of uncorrelated paths. Compared with a simple feedback channel, a system with 4 inputs and 25 outputs is able to produce a 20dB higher amplification before feedback sets in. This holds true, of course, only under the assumption that each and every input and output path is sufficiently decoupled from the other input/output paths. Each listener seat receives sound from several loudspeakers, each of which irradiates a signal somewhat differently processed than any of the others (!).[65]

9.4.2.3.5 Active Field Control, AFC

The AFC System by Yamaha[66] makes active use of acoustic feedback for enhancing the sound energy density and thereby also the reverberation time. When using the acoustic feedback it is, however, important to avoid timbre changes and to ensure the stability of the system. To this effect one uses a specific switching circuit, the so-called Time Varying Control (TVC) which consists of two components:

- Electronic Microphone Rotator (EMR).
- Fluctuating FIR (fluc-FIR).

The EMR unit scans the boundary microphones in cycles while the FIR filters impede feedback.

For enhancing the reverberation the microphones are arranged in the diffuse sound field and still in the close-range source area (gray dots in Fig. 9-63 on the right). The loudspeakers are located in the wall and ceiling areas of the room. For enhancing the early reflections there are 4 to 8 microphones arranged in the ceiling area near the sources. The signals picked up by these are passed through FIR filters and reproduced as lateral reflections by loudspeakers located in the wall and ceiling areas of the room. The loudspeakers are arranged in such a way that they cannot be located, since their signals are to be perceived as natural reflections.

Furthermore the AFC system allows signals to be picked up, e.g., in the central region of the audience area and the reproduction of them via ceiling loudspeakers in the area below the balcony for the sake of enhancing spaciousness.

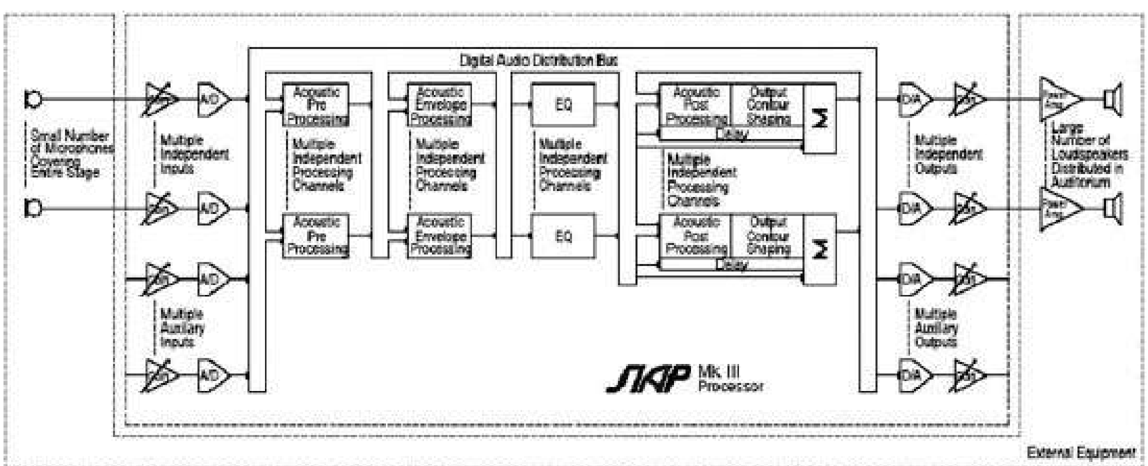


Figure 9-62. Schematic circuit diagram of SIAP.

9.4.2.3.6 Virtual Room Acoustic System CONSTELLATION™

CONSTELLATION™ by Meyer Sound Inc. is a multi-channel regenerative system for reverberation enhancement. Its development is based on ideas already considered in the sixties of last century by Franssen[58] when developing the above-mentioned MCR procedure.[67] The principle is already rather old and was already described by users.[68] The biggest difference is that today CONSTELLATION™ uses instead of the second reverberation room an electronic reverberation processor which is of course more easily adaptable.

But modern electronic elements and DSPs have made it possible to design circuits which widely exclude timbre changes. This is achieved by coupling a primary room A (the theater or concert hall) with a secondary room B (the reverberant room processor). Simultaneously the number of reproduction channels is reduced along with the timbre change of sound events. An enhancement of the early reflections is obtained as well, see Fig. 9-64.

Contrary to other systems, CONSTELLATION™ uses a comparable number of microphones and loudspeakers in a room. To this effect the microphones are located in the reverberant or diffuse field of all sound sources within the room and connected via preamplifiers to a digital processor. Then the outputs of the processor are connected to power amplifiers and loudspeakers for reproduction of the signal.

With the CONSTELLATION™ system there is a multitude of small loudspeakers L_1 to L_N (40 to 50) distributed in the room, which, of course, may also be used for panorama and effect purposes. Ten to 15 strategically located and visually inconspicuous microphones m_1 to m_N pick up the sound and transmit it to the effect processor $X(\omega)$ in which the desired and adjustable reverberation takes place. The output signals thus obtained are fed back into the room. The advantage of this solution lies in the precise tuning of the reverberation processor enabling well reproducible and thus also measurable results.

9.4.2.3.7 CARMEN

The underlying principle is that of an active wall whose reflection properties can be electronically modified.[69] The system was called CARMEN which is the French abbreviation of Active Reverberation Regulation through the Natural Effect of Virtual Walls. On the wall there are arranged so-called active cells forming a new virtual wall. The cells consist of a microphone, an electronic filter device, and a loudspeaker by which the picked-up signal is irradiated, see Fig. 9-65. The microphones are typically located at 1m distance from the loudspeaker of the respective cell, i.e., at approximately 1/5 of the diffuse-field distance in typical halls. Therefore one speaks also of a "locally active system."

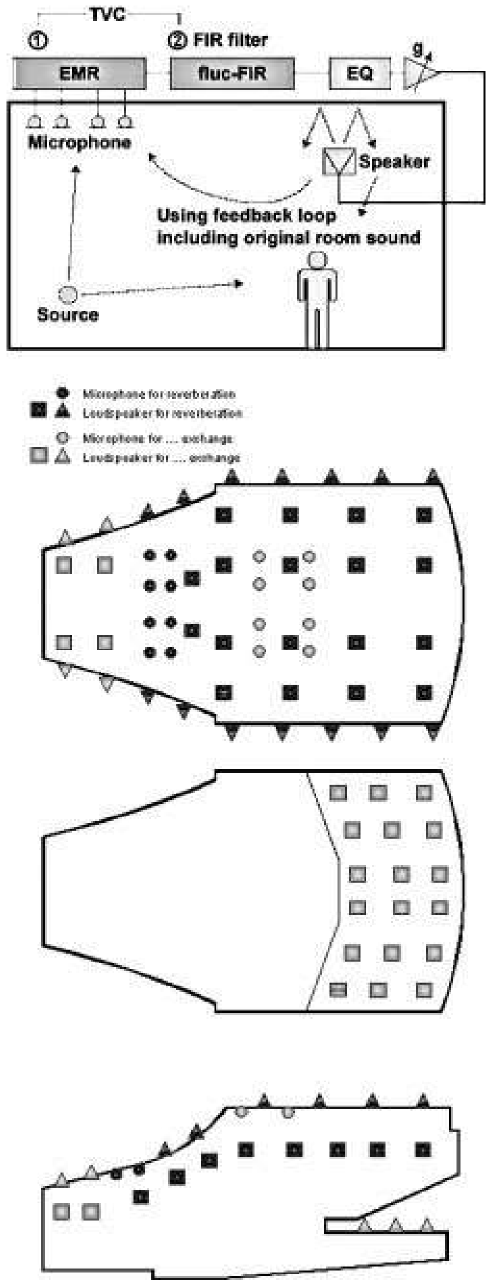


Figure 9-63. Active Field Control System (AFC) by Yamaha, Japan.

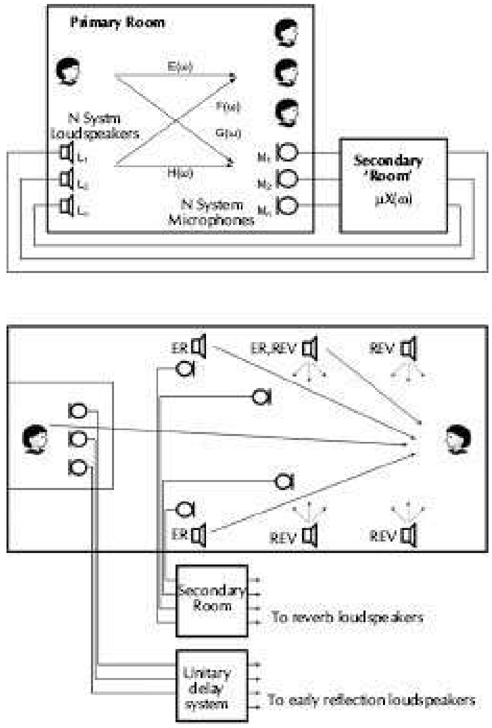


Figure 9-64. Principle of the Virtual Room Acoustic System, CONSTELLATION™.

Every cell produces a desired decay of the artificial reflections, provided one does not choose an excessive cell gain liable to provoke feedback. To avoid feedback an adequate microphone directivity characteristic as well as internal echo canceling algorithms are used. In addition it is possible to delay the microphone signal electronically, a feature which allows the cell to be virtually shifted and the room volume to be apparently enlarged.

Since 1998 CARMEN has been installed and tested in more than 10 halls used by important orchestras. It has proven to be particularly effective in theaters which are also used for orchestra performances. In these it improves most positively the acoustics in the distant areas under the balconies. In the MOGADOR Theater in Paris acoustics were significantly improved by installing CARMEN cells in the side walls and in the ceiling of the balcony.

By means of 24 cells in a room with a reverberation time of 1.2s at 500Hz, it was possible to enhance this reverberation time to 2.1s. Additionally there resulted various spatial effects like a "broadening of the sound source" or an improved envelopment with lateral reflections, features often required for big orchestras, but also soloists.

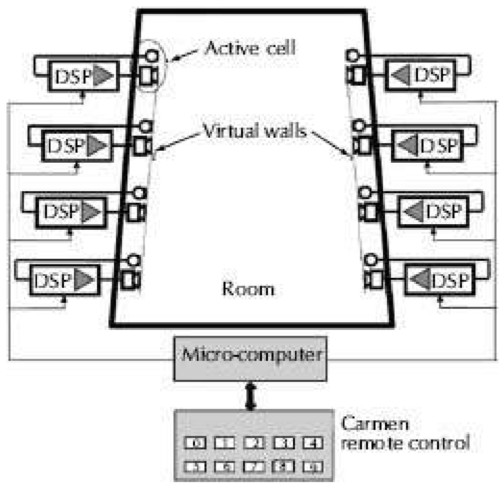


Figure 9-65. Principle of the Active Reverberation Regulation System, CARMEN.

9.4.2.3.8 VIVACE

VIVACE is a powerful system for electroacoustic enhancement of room acoustics. It creates virtual acoustics in indoor and even outdoor performance spaces and is used typically to improve the acoustic features of environments that are too small or insufficiently reverberant. VIVACE creates the acoustic impression of an ideal concert hall or any other room—either permanently or just for a specific event.

VIVACE ensures a high degree of detail veracity, perfect transient response, and exceptional feedback stability. The result is a homogeneous and entirely realistic three-dimensional sound which meets individual on-site acoustic requirements with the utmost flexibility and accuracy.

VIVACE also enables sources and effects to be moved around, virtually, in the acoustic environment.

A VIVACE system consists of a few microphones picking up the performance on the stage, the room-enhancement mainframe, an audio I/O matrix system, multichannel digital amplifiers, monitored remotely, and loudspeakers. VIVACE digitizes, analyses, and processes incoming stage-microphone signals in real-time and subsequently plays them back over precisely positioned loudspeakers, compare Fig. 9-66. Using an intelligent convolution algorithm VIVACE can recreate almost any environment in a low-reverberant space. Last installations are realized in the Dalian International Conference Center in China and the New Folklore Theater in Moscow/Russia (both in 2012).

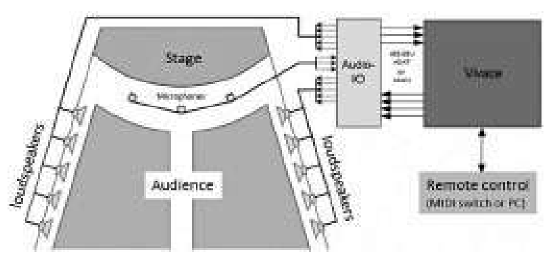


Figure 9-65. Block diagram of the VIVACE enhancement system.

9.4.3 Conclusions and Outlook

The above presented comparison shows that there exist a large number of procedures for enhancing reverberation and spaciousness, but only a part of it continue to be used nowadays. Thanks to ever increasing quality of the electronic transmission devices the prejudices still existing especially among musicians against the “electronic architecture” will diminish more and more, so that it will be increasingly possible to adapt concert halls even to the acoustical conditions characteristic of different creative and historic periods. Utilization in so-called multipurpose halls will, of course, prevail. The aim will have been achieved when musicians and audience perceive acoustical conditions established by means of "electronic architecture" as normal and natural. Simplicity of varying settings or security against acoustical feedback and unrelated timbre change will then be decisive factors in the choice of an enhancement system. Modern computer simulation will assist in banning the potential feedback risk.

Costly architectural measures for realizing the "variable acoustic" will be more and more discarded, particularly in view of their limited effectiveness.

References

1. W. Ahnert, W. Reichardt, *Grundlagen der Beschallungstechnik* (*Fundamentals of Sound Reinforcement Engineering*,) S. Hirzel Verlag Stuttgart, 1981.

2. Taschenbuch der Akustik (Handbook of Acoustics), Editors: W. Fasold, W. Kraak, and W. Schirmer. Verlag Technik, Berlin, 1984.

3. J. Meyer, *Akustik und musikalische Aufführungspraxis (Acoustics and Music Presentation Practice)*, Verlag Erwin Bochinsky, Frankfurt Am Main, 1995.

4. W. C. Sabine, *Collected Papers on Acoustics*, Cambridge: Harvard Univ. Press, 1923.

5. H. Kuttruff, *Room Acoustics*, 3rd Edition, Applied Science Publishers Ltd, London, 1991.

6. W. Ahnert, F. Steffen, *Sound Reinforcement Engineering, Fundamentals and Practice*, E&FN Spon, London, 2000.

7. L. L. Beranek, *Music, Acoustics and Architecture*, New York: Wiley, 1962.

8. W. Fasold, E. Veres, *Schallschutz + Raumakustik in der Praxis (Noise Control and Room Acoustics in Practice)*, Verlag fuer Bauwesen Berlin, 1998.

9. L. L. Beranek, *Concert and Opera Halls: How They Sound*, Woodbury: Acoustical Society of America, 1996.

10. ISO 3382, Messung der Nachhallzeit von Räumen mit Hinweisen auf andere akustische Parameter (Measuring the Reverberation Time of Rooms, with Hints for Other Parameters) (2000–03).

11. Y. Ando, *Architectural Acoustics*, Springer-Verlag edition New York, Inc., 1998.

12. J. Hoffmeier, Untersuchungen zum Einfluß von Raumklangfärbungen auf die Deutlichkeit von Sprache (Investigations on the Impact of Surround Sound Coloration on Speech Definition), Diploma thesis, Technical University, Diplomarbeit An Der Tu, Dresden, 1996.

13. L. Dietsch, "Objektive raumakustische Kriterien zur Erfassung von Echostörungen und Lautstärken bei Sprach und Musikdarbietungen" (Objective Room-Acoustical Criteria for Capturing Echo Disturbances and Loudness Variations with Speech and Musical Performing), Dissertation Technische Universität Dresden, 1983.

14. R. Hoehne, G. Schroth, "Zur Wahrnehmbarkeit von Deutlichkeits- und Durchsichtigkeitsunterschieden in Zuhörersälen" (Perceptibility of Definition and Clarity Differences in Audience Halls). *Acustica 81* (1995), pp. 309–319.

15. T. Houtgast, H.J.M. Steeneken, "A review of the MTF Concept in Room Acoustics and its use for Estimating Speech Intelligibility in Auditoria" *J. Acoust. Soc. Amer.* 77 (1985), pp. 1060–1077.

16. "RASTI-Sprachübertragungsmesser Typ 3361" (Speech Transmission Meter Type 3361), Datenblatt Fa. Brüel & Kjær.

17. V.M.A. Peutz, "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," *J. Audio Eng. Soc. 19* (1971), H 11, pp. 915–919.

18. W. Klein, "Articulation Loss of Consonants as a Basis for the Design and Judgement of Sound Reinforcement Systems," *J. Audio Eng. Soc.* 19 (1971).

19. Technical Report ISO Tr 4870, Acoustics: The Construction and Calibration of Speech Intelligibility Tests, 1991.

20. V.O. Knudsen, *Architectural Acoustics*. New York: John Wiley and Sons, Inc., 1932.

21. W. Reichardt, O. Abdel Alim, W. Schmidt, "Definitionen und Meßgrundlage eines objektiven Masses zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbie- tungen" (Definitions and Measuring Base of an Objective Measure for Ascertaining the Boundary Between Useful and Useless Transients at Musical Performances, *Acustica* 32 (1975) 3, p. 126.

22. W. Reichardt, O. Abdel Alim, W. Schmidt, "Zusammenhang zwischen Klarheitsmaß C und anderen objektiven raumakustischen Kriterien" (Correlation Between the Clarity Measure C and Other Objective Room-Acoustical Criteria). (1975), H 2, pp. 144–155.

23. W. Schmidt, "Raumakustische Gütekriterien und ihre objektive Bestimmung durch analoge oder digitale Auswertung von Impulsschalltestmessungen," Dissertation B Technische Universität

Dresden, 1979.

24. W. Kuhl, “Räumlichkeit als Komponente des Raumeindrucks” (Spaciousness as a Component of Spatial Impression). *Acustica* 40 (1978), p. 167 ff.

25. U. Lehmann, “Untersuchungen zur Bestimmung des Raumeindrucks bei Musikdarbietungen und Grundlagen der Optimierung” (Investigations on the Determination of Spatial Impression with Music Presentations and Bases for Optimization, Technische Universität Dresden, 1974.

26. W. Schmidt, “Zusammenhang zwischen Hallabstand und Nachhallzeit für den Raumeindruck (Correlation Between Direct to Reverberent Ratio and Reverberation Time with Regard to Spatial Impression). *Zeitschrift fuer Hochfrequenztechnik Und Elektroakustik* 77 1968, p. 37 ff.

27. M. Kleiner, “A New Way of Measuring Lateral Energy Fractions.” *App. Acoust.*, Vol. 27, p. 321, (1989).

28. W. Reichardt, A. Kussev, “Ein-und Ausschwingvorgang von Musikinstrumenten und integrierte Hüllkurven ganzer Musikinstrumentengruppen Eines Orchesters,” *Zeitschrift Für Elektrische Informations-und Energietechnik* 3 (1973) 2, pp. 73–88.

29. W. Fasold, E. Sonntag, H. Winkler, *Bauphysikalische Entwurfslehre: Bau-und Raumakustik (Theory of Structural-Physical Design, Architectural and Room Acoustics)*, Verlag für Bauwesen Berlin, 1987.

30. J. Meyer, “Zur Dynamik und Schalleistung von Orchesterinstrumenten,” (On the Dynamic Range and Acoustic Power of Orchestra Instruments). *Acustica* 71 (1990), pp. 277–278.

31. H. Winkler, U. Stevenson, “Einfluss geneigter Seitenwände auf die Nachhallzeit” (The Influence of Inclined Side Walls on Reverberation Time), Fortschritte der Akustik - DAGA, (1990), pp. 867–874.

32. L. Cremer, “Early Lateral Reflections in Some Modern Concert Halls,” *J. Acoust. Soc. Am. 85*, (March 1989).

33. H. P. Tennhardt, “Modellmessverfahren für Balanceuntersuchungen bei Musikdarbietungen am Beispiel der Projektierung des grossen Saales im Neuen Gewandhaus Leipzig” (Model Measuring Procedures for Balance Investigations of Musical Performances Demonstrated on the Example of the great Hall of the Neues Gewandhaus Leipzig), *Acustica* 56 (1984) Nr. 2 (Oktober 1984), pp. 126–135.

34. L. Cremer, H.A. Mueller, *Die Wissenschaftlichen Grundlagen der Raumakustik, Band I (The Scientific Foundations of Room Acoustics*, Vol. 1, S. Hirzel Verlag Stuttgart, 1978.

35. H. Winkler, “Das Nachhallreservoir, Bedeutung und Beeinflussung” (The Reverberation Time Reservoir, Importance and Control), *Fortschritte der Akustik—DAGA* 95, pp. 315–318.

36. K. Takaku, S. Nakamura, “A Report on the Relationship between Orchestra Shell Design and Musicians’ Acoustical Impression,” *15th Intern. Congress on Acoustics*, Trondheim, June 1995, pp. 525–528.

37. A. Askenfeld, Stage Floors and Risers—Supporting Resonant Bodies or Sound Traps? In: S. Ternström, *Acoustics for Choirs and Orchestra*. Stockholm, 1986.

38. J.H. Rindel, Acoustic Design of Reflectors in Auditoria. Proc. Inst. of Ac. 14, Part 2, p. 119.

39. M. Vorlaender, E. Mommertz, “Definition and Measurement of Random-Incidence Scattering Coefficients,” *Appl. Acoustics 60*, (2000), (2), 187–200.

40. E. Meyer, L. Bohn, “Schallreflexion an Flächen mit periodischer Struktur” (Sound Reflections at Surfaces with Periodical Structure), *Acustica* 2 (1952), H. 4, pp. 195–207.

41. M.R. Schroeder, "Binaural Dissimilarity and Optimum Ceilings for Concert Halls: More Lateral Sound Diffusion," *JASA*, Vol. 65 (1992), No.4, p. 958 ff.

42. M.R. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *Acustica*, Vol. 49 (1981).

43. N. Kiesewetter, "Schallabsorption durch Platten-Resonatoren" (Sound Absorption by Panel Resonators), *Gesundheits-Ingenieur Gi 101*, (1980), H.3, pp. 57–62.

44. The New MICRO**SORBER** from KAEFER, Product Information of KAEFER Isoliertechnik, Bremen, Germany, www.kaefer.com.

45. E. Mommertz, *Untersuchungen akustischer Wandeigenschaften und Modellierung der Schallrückwürfe in der binauralen Raumsimulation, (Investigations of Acoustic Wall Properties and Modelling of Sound Reflections in Binaural Room Simulations)*, Aachen: Shaker, 1996.

46. L. L. Beranek, *Concert halls and Opera Houses, Music, Acoustics, and Architecture*, New York: Springer, 2nd edition 2004.

47. T. Cox, P. D'Antonio, *Acoustic Diffusers and Absorbers: Theory, Design and Application*, London: T&F Group, SPON, 2004.

48. E. Mommertz, M. Vorlaender, "Measurement of Scattering Coefficients of Surfaces in the Reverberation Chamber and in the Free Field," *Proc. 15th ICA*, II, (1995), 577–580.

49. M.H. Gomes, M. Vorlaender, S.N. Gerges, "Anforderungen an die Probeflächengeometrie bei der Messung des Streugrades im Diffusfeld," *Fortschritte der Akustik—DAGA* '02 – Bochum, S. 584–585.

50. ISO 17497, 2004 (E), Acoustics—Sound-Scattering Properties of Surfaces.

51. D. Kleis, “Moderne Beschallungstechnik” (Modern Sound Reinforcement). Philips tech. Rdsch. 20 (1958/59) 9, S. 272ff. und 21 (1959/60) 3, S. 78ff.

52. E. Meyer, H. Kuttruff, “Zur Raumakustik einer großen Festhalle” (Regarding the Room Acoustic of a Large Festival Hall), *Acustica 14* (1964) 3, S. 138–147.

53. W. Kaczorowski, “Urzadzenia elektroakustyczne i naglosnia w Wojewodzkiej Hali Widowiskowo-Sportowej w Katowicach” (Electroacoustic Equipment in the Wojewod Sport Hall in Katowice). Technika Radia i Telewizji 4 (1973) S. 1–16.

54. C. McGregor, “Electronic Architecture: The Musical Realities of Its Application,” 6th Int. Conf. Sound Reinforcement, Nashville, 1988.

55. “Sound System Design for the Eugene Performing Arts Center, Oregon,” Recording-Eng./Producer, (1982) S.86–93.

56. P.H. Parkin, K. Morgan, “‘Assisted Resonance’ in the Royal Festival Hall, London,” J. Sound and Vibration 2 (1965) 1, S.74–85.

57. “A Special Report on the Experimental (“Assisted Resonance”) System in the Royal Festival Hall, London,” J. Sound and Vibration. 1 (1964) 3, S.335–342.

58. N.V. Franssen,. “Sur l`amplification des champs acoustiques,” Acustica 20 (1968), S. 315 ff.

59. M. Ohsmann, “Analyse von Mehrkanalanlagen” (Analyzing multichannel systems), Acustica 70 (1990) 4, S. 233–246.

60. S.H. de Koning, “Konzertsäle mit variablem Nachhall,” Funkschau 57 (1985) 14, S. 33–38.

61. A. J. Berkhout, “A Holographic Approach to Acoustic Control,” J. Audio Eng. Soc. 36 (1988) 12, S. 977–995.

62. P. W. Barnett, “A Review of Reverberation Enhancement Systems,” AMS Acoustics Ltd., London 1988 (U.S. Patent 4.649.564, March 10, 1987).

63. D. Griesinger, “Improving Room Acoustics Through Time-Variant Synthetic Reverberation,” AES Convention Paris, Preprint 3014 (1990).

64. D. Griesinger, (U.S. Patent 5.109.419, April 28, 1992).

65. W. Prinssen, B. Kok, “Technical Innovations in the Field of Electronic Modification of Acoustic Spaces,” Proc. of the IOA, (1994), Vol. 16, Part 4.

66. Hideo Miyazaki, Takayuki Watanabe, Shinji Kishinaga, Fukushi Kawakami, Yamaha Corporation, Advanced System Development Center, “Active Field Control (AFC), Reverberation Enhancement System Using Acoustical Feedback Control,” 115th AES Convention, (October 2003), N.Y.

67. M.A. Poletti, “On Controlling the Apparent Absorption and Volume in Assisted Reverberation Systems,” Acustica 78, (1993).

68. W. Ahnert, “Einsatz elektroakustischer Hilfsmittel zur Räumlichkeitssteigerung, Schallverstärkung und Vermeidung der akustischen Rückkopplung,” Diss. Techn. Univ., Dresden, 1975.

69. J.P. Vian, X. Meynal, “How Reverberation Enhancement Systems Based on the Virtual Wall Principle can Improve the Acoustic Qualities of Concert Halls from the Point of View of theMusicians, Experiences with the Carmen System,” Symposium on Authentic Music Performance and Variable Acoustics, University of Ghent, May 15, 1998.

Further Reading

J. Blauert, *Raeumliches Hoeren* (*Spatial hearing*). The MIT Press, Cambridge, Massachusetts and London/England 1983.

F. A. Everest, *The Master Handbook of Acoustics*, 3rd ed. TAB Books, Division of McGraw-Hill, Inc. 1994.

J. Cowan, *Architectural Acoustics Design Guide*. McGraw-Hill Companies, Inc., 2000.

L. Marshall, *Architectural Acoustics*, Elsevier Academic Press, 2006.

T. D. Rossing, *Handbook of Acoustics*. Springer 2007.

Chapter 10

Worship Styles in the Christian Church

by Doug Jones

10.1 Introduction

At first glance it seems odd indeed that a book such as this *Handbook for Sound Engineers* should have a chapter on the worship styles of the Christian Church. We have included this chapter for the simple reason that in the United States, the Christian Church is a very big business. According to the Federal Reserve bank of St. Louis Economic Research, in 2013 between 3 and 4 billion was spent on religious building construction, a lot of money to be sure, but down considerably from the pre-recession spending of just under 9 billion dollars in 2004! In 2009 considered by the feds to be a recession year, 7.5 billion was spent on "Construction Spending: Religious"[1] According to the American Religious Identification Survey (ARIS 2008) some 76% of Americans self-identified as Christian compared to 3.9% who self-identified as other.[2] Another group found similar numbers. According to the Pew Research; Religion & Public Life Project 2008, Christians made up 78% of the US population, all other religions combined made up 4.7%, with 16.1% stating that they were unaffiliated, and 0.8% declining to answer.[3]

It stands to reason that the majority of the money spent on "religious construction" is being spent on churches. The past two decades have seen the rise of the so called "Mega churches" especially in the Bible Belt where media budgets in the many millions of dollars are not uncommon. Some may be offended that other religions were not specifically included in this discussion. The author certainly intends no offence. If the reader is part of the 4.7%

non-Christian religious community, you are encouraged to examine this taxonomy and see if there are parallels to your particular religion. If you are part of the 16.1% non-religious community, please see this as a resource to get to know this market better. Finally to the 78% Christian community, this author has observed on numerous occasions that Christians themselves often do not understand the differences that exist between them. See this as a way to begin to understand the church that you did not attend.

Like any niche business, it is important to understand the nuances of what makes the business tick. If one is to be successful in selling goods and services to the Christian Church, one must understand the differences that exist among the various factions in the church. These differences play important roles in the design of the acoustic space for the church and in the use of media.

10.2 Worship Styles

There certainly are many ways to analyze and categorize the church. One could group churches based on a common theology, a shared history, or a shared governance philosophy. And of course any attempt at doing so would generate a considerable list of "outliers" or churches which defy categorization. Categorization is also difficult because it is not at all clear, even to people who describe themselves as Christians, which groups are a part of the Church and which do not make the cut.

If one zooms out far enough, one can begin to see groupings especially with respect to how church people interact with church buildings. If one further refines the study, and considers expectations for the way the rooms sound, i.e., the acoustics of the spaces and the way media, especially sound systems are used, one can begin to see distinct patterns in the chaos.

As it turns out, it is the *style* of worship which is the most accurate predictor of the kind of acoustic space and sound system required or preferred for any given church. Of course the style of worship is rooted in ecclesiology and to some degree in theology as well. There are four worship styles that are utilized by the vast majority of 21^{st} century Christian Churches in America.[4]

The first could be known by many names, but perhaps the best is the *Celebratory style*. These are the churches that follow a liturgy rich with tradition, which culminates in the celebration of what some call the Mass, others the Divine Liturgy. The next one might best be called the *Evangelical style* as it focuses on the proclamation of the Word and personal salvation. Next is the *Experiential style* because of the emphasis on the experience of the power of God to change lives. Finally there is the *Community style* which views Church as an expression of unity and commitment among believers.

As soon as one creates a taxonomy of churches someone will discover a church which does not seem to fit. This is likely true of this taxonomy as well. There are probably no churches which fit these descriptions exactly, and many churches which may fit into more than one category. This being said, the majority of American Christian Churches will fit *better* in one of these categories than in the others. The architect, acoustician and systems provider will better understand the Church as a client if these four worship styles are understood. It is also important to note that one cannot always infer a particular worship style from the name of a particular church.

10.3 The Celebratory Style of Worship

Perhaps the best way to determine if a given church employs the celebratory style of worship is to ask a member of the clergy "why do you do Church?" or maybe "what is the point of your church service?" Those who practice celebratory worship will answer immediately "to celebrate the Mass."

Those Christians who belong to the Orthodox, Catholic, or Episcopal churches all practice similar forms of Celebratory worship.

The *Catholics* refer to their celebration as the Mass. There is some discussion as to the ultimate source of this word but most Catholics trace it back to a Latin word *missio* which is the root of the English word mission. The word means to be sent out or dismissed. In early times there was a point during the church service when the uninitiated were dismissed. Gradually popular speech used the name of one of the parts of the ritual, the dismissal or the missio to indicate the whole.

The *Orthodox* do not generally use the term Mass. Although to the outside observer there seems to be much in common with the Roman Catholic Mass, it is inappropriate to use the Latin name for the Eastern worship service. In the Orthodox church, the celebration is referred to as the Divine Liturgy. The word liturgy is from a Greek word meaning common action or service, and its use signifies that this is an act of common worship that unites the faithful in service to God, Fig. 10-1.

The *Anglican* uses the term Mass to describe the worship service, but it should be very clear that although again to the outsider there is very little discernible difference between an Anglican Mass and a

Roman Catholic Mass, to the faithful in these two churches there are very important differences which must be respected.

10.3.1 History

The celebratory form of worship is by far the most ancient of all the forms of worship. The modern Mass (Devine Liturgy) can trace its roots back to the second century AD.

Before 1054 AD there essentially was one Christian Church. The word Catholic in fact means universal. In 1054 AD there was the first real serious schism in the Church. It split in two with the western part headquartered in Rome becoming known as the Roman Catholic church and the eastern part, which was centered in Constantinople (now Istanbul), becoming known as the Orthodox or Eastern Orthodox church. In 1534, the Church of England (Anglican church) was formed when it broke away from the Church in Rome, Fig. 10-2.

Figure 10-1. Typical Roman Catholic church.

Figure 10-2. St. Joseph Orthodox Church, Wheaton, IL.

In the Roman Catholic church there was a significant reform which took place in the early 1960s. It became known as Vatican II and it addressed virtually all of Roman Catholic life and worship. Pre-Vatican II Mass was spoken in Latin by priests at the front of the church with their backs to the congregation. Vatican II allowed for the Mass to be spoken in the vernacular and in a much more inclusive fashion with the Priest facing the congregation.

Churches built before Vatican II tend to be reverberant spaces, with very high aspect ratios with traditional architecture which can be traced back to medieval times. For all the changes that Vatican II did initiate, it is almost silent on the subject of the architecture of the Roman Catholic church. The only direct mention of the church structure states "*...when churches are built, let great care be taken that they be suitable for the celebration of liturgical services and for the active participation of the faithful.*"[5] This phrase "active participation of the faithful" has been interpreted to mean all sorts of things, and triggered a wave of renovations to traditional churches and new churches being built. Some of these more modern structures are very successful and well received by laity and clergy alike. Others, not so much! There has been considerable reaction to some of the modern architecture, Fig. 10-3. Writers like Steven Schloeder refer to modern Catholic architecture as banal and insists that churches should not be built simply as places for people to gather, but for the worship God.[6] Others like Michael Rose were even more out spoken in their criticism of modern church architecture. The title of his book published in 2001 says it all; *Ugly as Sin: Why They Changed Our Churches from Sacred Places to Meeting Spaces and How We Can Change Them Back Again.*[7]

Figure 10-3. St. Dominic Catholic Church, Miami, FL.

10.3.2 Celebratory Worship

There are two attributes which differentiate the celebratory style from the other forms of Christian worship. The first is that the church service or Mass is a ritual consisting of a number of elements, but aside from minor changes mandated by the church calendar, the Mass is the same every time it is spoken. There is little room for variability or personal preference. The only element that will change week to week is the homily or sermon. The homily however is not considered essential to the Mass. The other unique attribute is their Sacramental perspective. At the risk of oversimplification, a Sacramental understanding of Gods relationship to humankind is one that believes that "*God delights to use tangible, concrete earthy means—matter itself—to communicate His grace, redemption and presence to us….*"[8]

When a believer partakes of or participates in a Sacrament, something real, if mystical, happens. The Christian is changed, grace is imparted. When one undergoes the sacrament of baptism for example, it is much more than a symbolic exercise. The baptized is welcomed by God in to the community of the faithful. It is this

Sacramental world view, if you will, which determines virtually everything about the Celebratory church from the architecture of the buildings to the way the services are conducted.

10.3.3 Sacred Space

In all of the Sacramental or Celebratory churches, the Mass or Divine Liturgy must be conducted in consecrated or sacred space. As we shall see this position stands in sharp contrast to the other worship styles. In this tradition, a space is made sacred by an act performed by someone whose office entitles them to sanctify a space for holy use.

In the celebratory style of worship, symbolism is very important. Virtually all aspects of the architecture and furnishings have meaning and are there to support the Mass. It is the only one of the worship styles in which the visual is an important part of the act of worship. Indeed in some forms of the celebratory worship, all five senses are involved. All the symbolism of the altar, vestments, furnishings, indeed of the architecture itself appeal to the eye. The sounds of the bells, chants, prayers, spoken and musical responses are a central part of all celebratory worship. In the Eucharist the worshippers partake of the sacred bread and wine. In the passing of the peace, worshippers embrace or shake hands with each other.

10.3.4 Acoustical Considerations for the Celebratory Style

We have seen that the Celebratory Style of worship may only takes place in a space which is considered sacred. Sometimes, in the larger cities particularly, Celebratory churches are large cathedrals, Like New York's Cathedral of St. John the Devine which seats around 5000 people. Sometimes the church housing a local parish can barely accommodate 200. In this tradition, the size is not important. The Mass can be conducted for one, or for thousands. If in the Sacramental church the architecture must reflect the Sacred, then the acoustics which is a direct result of the architecture must also reflect the same. This of course is a much more difficult question. In architecture, and in most of the visual art forms there is at least a language rich with metaphor that can be used to describe how the visual might evoke or elicit the Sacred or the Holy. This language is woefully lacking in the field of acoustics, indeed it is lacking in things aural in general. In English which has more words by far than any other language, there are very few words which describe aural experience uniquely. The few that do like *silent, quiet, cacophonous, shrill* and *sibilant* either describe the *absence* of sound or undesirable properties of sound. Enter into a grand cathedral and what words come to mind to describe the acoustics of the space? We are forced to borrow terms from other senses or resort to metaphor or onomatopoeia. We might be able to speak of sacred architecture but can we speak of sacred acoustics? Can we speak of sacred sound? It is beyond the scope of this chapter to try to unravel the reasons for this lack of vocabulary much less to propose a language. Yet if there is such a thing as sacred space then there must exist also an acoustic space which if not sacred at least lends itself to sacred activity.

What can be done in the realm of acoustics to create or preserve an environment that is conducive to worship in the celebratory style? The first concrete step we must take is to insure that 1) we do no harm and 2) to realize that in the Celebratory church the architecture, because of the symbolism involved, trumps everything! It is as important to avoid doing harm to the acoustic space as it is to take active steps to produce an acoustically appropriate space. One form of “harm” of course is noise. Of all the forms of worship, the Celebratory is the most likely to be disturbed by extraneous noise. The design team should strive for noise levels of no higher than NC25. This can be a challenge in an urban setting, especially if there are many large windows incorporated in the architecture.

Another form of “harm” is to not respect the reverberation as something desirable. Acousticians sometimes tend to view reverberation as the enemy. Certainly there are situations where the reverberation time is excessive and needs to be creatively controlled. However, the perception of scale which often is associated with sacred space is experienced through two senses. The eye sees the large structure and is impressed. The auditory system at the same time is detecting the initial time gap which corroborates and reinforces the sense of scale, contributing to the impression of awe.

Many Celebratory churches still utilize pipe organs in their sanctuaries. Of all the types of churches in the USA, one is most likely to find a functional pipe organ in a Celebratory church. If a pipe organ is being considered for a new facility, it will likely dominate the discussion of the acoustics of the room. On some level it is reasonable to allow a million dollar plus investment to determine the acoustics of the room, but one should never lose sight of the fact that the purpose of the space is not just the organ. Ultimately this is not a performance space, this is Holy space, which exists primarily to serve the Mass. Other uses of the space such as organ recitals are secondary uses.

To summarize, in pre Vatican II churches the predominant acoustic issue is likely to be reverberation. The reverberation may be well suited to the pipe organ and chant but ill suited for everything else. Suggest treatment very judiciously. Not only will there be push-back due to the way the treatment will look, but it is likely that congregations worshiping in pre Vatican II churches have come to expect the church to sound a certain way. To them it most certainly is not broke, so don't fix it! Perhaps the ultimate in celebratory worship would be a church with variable acoustics that could be adjusted for a midband reverberation time of around three second for pipe organ music and chant but be turned down to around a second or less for "contemporary" or "alternative" forms of the Mass.

10.3.5 Sound System Considerations

Perhaps the most offensive things one can say about a Celebratory worship service is to call it a performance. It is *not* a performance! It may look like one to an outside observer, but it is a sacred act of worship. It contains pageantry even some choreography and music, but it is not a performance. For most churches that worship in the celebratory style, anything that hints of performance or commercialism is viewed very suspiciously. Churches may install lighting systems to light the Altar and enhance the beauty of the sanctuary, but balk at having the controls of the lighting system be anywhere in the congregation much less have someone "running lights" during a Mass. It also explains why for the most part it is very difficult if not impossible to get a Celebratory church to install a mixing console in the congregational seating area and have someone "mix" the service. Celebratory services are often held in the most difficult of acoustic spaces and require sophisticated sound systems to get intelligible speech to the congregation. Furthermore most parishes offer Masses at different times during the week where the numbers of those in attendance can range from very few to a full house. The different Masses are often officiated by different ministers or priests. All these conditions should be taken into consideration by those running the sound system, yet the systems in Celebratory churches rarely benefit from human control. The systems are most often automatic and represent significant compromise. Although it may seem odd, the way a sound system *looks* is very important in this style of worship. A large central cluster or centrally positioned line array may be desirable from an intelligibility point of view, but if it detracts from the center of visual focus which is the Altar, it will be vetoed.

Deal with the intelligibility problems with very high Q or at least very directional loudspeaker systems which can easily be integrated into the architecture. Steerable array speaker systems can be very useful in the older Celebratory churches, Fig. 10-4. The architecture will trump everything. If the sound system can be guaranteed to provide everyone with perfectly intelligible sound, but is deemed an eyesore, it will not be implemented. These churches need technology to be as inconspicuous as possible and in general will not tolerate anything which appears to suggest that the Mass is a performance.

One of the serious challenges facing the Celebratory church in the 21st century is a challenge that many leaders of the Church have not yet articulated but grapple with on a weekly basis. There is a very strong sense of history and often a desire to preserve the sort of church architecture prevalent before Vatican II. At the same time many parishes are trying to reach new audiences by trying to incorporate forms of music which often are at odds with the acoustics of the space. It is very difficult to use drums in a room with 6 seconds or more of reverberation!

Another situation that sometimes poses a challenge for sound systems in the traditional cathedral style church, is the location of the choir and organ. Often they are located in the rear of the church in a balcony or loft. Here the overriding rule, is “if it ain’t broke, don’t fix it!” Often the acoustics of the room are such that the choir is heard well enough without amplification. It can get tricky however if there is no choir and the organist is the cantor (singer) as well as the organist, or if there is a choir with a soloist who needs amplification. In those situations putting him or her into the system is desired, but remember that they will probably need a low level un-delayed monitor near them. If the main sound system loudspeaker(s) are towards the front of the church or if there is a distributed system with delayed loudspeakers, it is likely that the delay, either acoustic or electronic, will make it impossible for the organist to sing. A low power monitor loudspeaker situated very close to the organist will help but be careful of the levels in the monitor so that this un-delayed signal does not get back into the main part of the church.

Figure 10-4. St. Michael's Church, Chicago, IL.

10.4 The Evangelical Worship Style

The Evangelical style church has as its central focus the proclamation of the message, i.e., preaching. This is in contrast to the celebratory style where the celebration of the Mass is central. If you ask someone form this style why they do church or attend the service, they will reply "to *hear the word of God proclaimed*". The vast majority of the churches which worship in the evangelical style are part of the greater Protestant church. It is unwise to infer from the name of a church what their worship style is, however as a rule, Baptists, Congregationalists, Presbyterians, and Methodists to name a few are often utilize this style of worship.[*]

10.4.1 *History*

To the Roman Catholics it was a rebellion. To Protestants, it was their birth and a reformation in the truest sense of the word. To society it was a revolution. To the Orthodox it was a Roman Catholic problem and had little relevance. No matter what camp one is in, the Protestant Reformation certainly changed the Christian church. Although there were a number of "reformers" before Martin Luther[9] he is generally viewed as the one who really got the ball rolling as it were. Luther (1483–1546) was a Roman Catholic Priest from present day Germany who was dissatisfied with the sorry state of the Roman Catholic church, Fig. 10-5. Even Roman Catholic historians agree that this was a dark time for the Church. According to McBrien's *Lives of the Popes*, there were four popes who ruled the Church between 1492 and 1521; Innocent VIII, Alexander VI, Julius II and Leo X. All four made McBrien's list of the "worst of the worst", with Alexander VI (1431–1503, pope from 1492–1503) accorded the dubious honor of being called the "most notorious pope in history.[10] It was in this context that Luther wrote his famous *95 Theses* which he nailed to the door of the Castle Church in Wittenberg, Germany in 1517. The document was a list of what Luther felt to be the Churches' worst offenses, and an invitation to a public debate. This was extremely inflammatory in the sixteenth century, and it sparked what we now call the protestant reformation.

At the heart of Luther's message was the deep conviction that a Christian should be instructed by the Holy Scriptures only. When Scripture and Church disagreed, Scripture should win out. This message came at just the right time. Literacy was on the rise, and critical thinking was in vogue. In less than 20 years, Luther's reformation had swept through Europe. Only Spain and their infamous "Spanish Inquisition" managed to give any significant opposition.

Figure 10-5. Martin Luther.

10.4.2 Evangelical Worship

Worship in the evangelical style involves a number of elements, such as music, and prayer, but the main point of the service will always be the sermon. Unlike the celebratory style, no part of the service is prescribed or considered mandatory by a governing body, but an evangelical service without a sermon would be considered extremely unusual, and if it were to happen, the congregants would wonder if they had really been to church.

10.4.3 Architecture

Luther's insistence on the scripture as the sole authority can still be seen in the architecture of Protestant churches today. Go into just about any Baptist church and the first thing you will see, the center of visual focus if you will, is the pulpit, the place from which the

Word is proclaimed, Fig. 10-6. Contrast this with the center of visual focus in a Catholic church which will almost always be the sacred Altar, Fig. 10-7. Another important contrast to the celebratory style is the lack of sacred or Holy space. In the evangelical style of worship there is no need for sacred space, so the architecture has virtually no symbolic elements at all. The Evangelical church is a functional space, as long as the Word is lifted up (figuratively and literally) all is well. These churches can range in size from being so small, Fig. 10-8, that no sound system equipment is needed at all to seating many thousands and requiring a large sound system. A few of the Mega churches employ this style of worship.

Figure 10-6. First Baptist Church, Branford, CT.

Figure 10-7. St. George Catholic Church, Guilford, CT.

Figure 10-8. A small Evangelical church.

10.4.4 Acoustical Considerations

If you wanted to list the events of a church service in order of importance, in the celebratory style, the homily or preaching would be quite a few places down in the list. In the evangelical style it is right at the top. The congregation needs to be able to hear and understand the preaching. In the celebratory style, the ability to hear and understand is almost a luxury (remember that before 1963 the Mass was spoken in Latin) but not absolutely necessary for the Mass to be effective. In the evangelical style, it is critical. But running a close second is the need for the congregation to hear themselves sing. Music is present in all four worship styles, but in the evangelical style it is arguably the most developed and participatory. There is also some limited use of music as performance. So the acoustics for the evangelical style church needs to be optimized for speech originating at the pulpit and heard in the congregation, music originating on the stage and heard in the congregation, and music originating in the congregation and heard by itself. This is a very difficult juggling act indeed! In this style of worship, there are two surfaces that are critical to the success of the space. One is the back wall. It is very important that the back wall not reflect energy back to the stage and most of all not back to the pulpit. Reflections from the back wall can limit the potential gain before feedback and can make it very hard to speak from the pulpit. The other critical surface is the ceiling above the congregation. If this is too high, there will be no useful reflections back to the congregation and they will not hear themselves sing. If it is too low, and reflective, it will cause a harshness in the vocal range that also makes it hard to sing. Even worse is the use of acoustical tile on a low ceiling. Acoustical tile is generally not a very linear absorber and can make for a very unpleasant singing environment. The ceiling should provide reflections in the 16 to 24ms range, Fig. 10-9.[11]

Figure 10-9. High ceilings enhance the singing environment.

10.4.5 Sound System Considerations

In the evangelical style of worship, intelligibility is of the utmost importance. The reader is referred to Chapter 40 *Designing for Speech Intelligibility* for more information about the technology that is currently available for the quantification of intelligibility. Architects need to understand that sound systems especially loudspeakers need to be installed in specific places for important reasons. There are numerous types of loudspeaker systems and approaches to implementing sound reinforcement systems. As a general rule in order to insure that there is intelligibility in every seat in the congregation, everyone seated in the congregation must have an unobstructed view of one and only one loudspeaker. Often the optimal location for mounting loudspeaker systems will be directly above the pulpit. Especially in a larger Evangelical church this area needs to be left clear of visual symbols or elements which would get in the way of a loudspeaker cluster. In addition, structural support should be provided especially in larger rooms to accommodate the loudspeaker system. Ideally the planning of the facility will involve not only the architect/builder, but an acoustical consultant and a media systems contractor as well. The audio and video systems should be integrated into the architecture and not simply added as an afterthought, Fig. 10-10.

Figure 10-10. Parkview Community Church, Glen Ellyn, IL.

If the noise is kept to a minimum, the reverberation is controlled and kept appropriately low, surfaces which are creating harmful reflections are appropriately treated, focusing surfaces are avoided and there is system friendly architecture, then the sound system contractor can install a sound system which can be guaranteed to provide intelligible sound to every seat in the room.

10.5 The Experiential Style

In a sense, all Christian worship is experiential. In the Celebratory church, the believers experience the grace of God via the Sacrament of Eucharist. In the Evangelical church believers hear the word of God and experience salvation or find a sense of direction for living. However for some churches it is the *experience* which defines the worship. Even though one might be hard pressed to find a church that claims to worship in the experiential style, it is clear that there are a significant number of churches that conduct worship in a manner that emphasizes the experience. In these churches the worship service can only be described as a performance, and they don't seem to be troubled by this at all. This stands in stark contrast to the celebratory style, especially the Roman Catholic version where referring to the Mass as a performance would be seen as a grave insult. When there is an emphasis on experience, worship takes on a unique quality which has significant ramifications for acoustics and media systems. The Experiential church is likely to have sophisticated media systems which are a significant part of the experience.

10.5.1 History

Those who practice this form of worship with the primary emphasis being on experience trace their roots to two very different movements. One branch is the seeker sensitive or seeker friendly church. The branch traces its history to the Pentecostal movement. The Seeker Friendly church is a relatively modern form of church. Many credit Dr. Robert Schuller of the Crystal Cathedral of Orange County, California as being the founder of the seeker sensitive or seeker friendly movement. Schuller had a vision to build a church for people who would never consider darkening the door of a conventional church. The idea is to create a very friendly and exciting environment that is welcoming to all and does not feel like "Church." In this way it is argued, un-churched people will come and possibly discover Christianity. Some "Mega churches" like Saddleback in Southern California and Willow Creek in Illinois are churches that are seeker friendly but are also trying to deliver the traditional Gospel message as well.

The other branch of the Experiential church has its origins in the Pentecostal movement. The Pentecostals trace their roots back to the Holiness movement of the late 19th century. An important part of Pentecostal theology and praxis is the experience of the baptism in the Holy Spirit. This term, first coined by John Fletcher, a colleague of John Wesley of Methodist fame in the mid-18th century, refers to “an experience which brought spiritual power to the recipient as well as inner cleansing.”[12] This baptism was eventually connected with the practice of speaking in tongues or glossolalia.

10.5.2 The Charismatic Movement

Some Christians did not buy into the traditional theology of the Pentecostal church, nor would they attend Pentecostal churches. They however sought after similar experiences of being *Spirit Filled*, and many of them spoke in tongues sometimes to the dismay of the more traditional, non-Pentecostal churches to which they belonged. This movement gained in prominence throughout the 60s. These Neo-Pentecostals referred to themselves as the *Charismatic Movement* to distinguish themselves from the main stream Pentecostals. The Charismatic Movement was quite possibly the most important Christian movement of the late 20th century. By the end of the 1960s there was such momentum that Charismatics were popping up in virtually every Protestant denomination and even in some Roman Catholic and Orthodox churches.

10.5.3 Experiential Worship

Christians with Pentecostal origins are second only to Roman Catholics in number.[13] The worship can appear to the outsider as controlled chaos! It is very participatory and involves loud music, and energetic preaching interrupted with ecstatic out bursts from the congregation. The worship will often involve performed music and, especially in African-American churches, gospel choir music. The underlying element is the *experience* of the out pouring of the Holy Spirit and the power of God to change lives. This experience is sought by those who attend and everything, the singing, the performing, the preaching, the healing, the speaking in tongues, all combine to provide the experience.

The seeker friendly worship experience is much more controlled and generally scripted to give a very polished feel. It is a performance and in the larger churches with bigger budgets it can be a spectacular event. It is not uncommon to find that there are well paid professional staff in charge of the sound and lights. The service may have all the same elements that you would find in an evangelical style service, but the vibe and the emphasis is quite different. The music is generally not very participatory. The "praise and worship" band is usually quite good and the music is led by singers on mics. You aren't prevented from singing along but it is not really necessary or expected.

Both of these types of Experiential churches utilize music in much the same way. According to Calvin Johansson in his monograph *Music in the Pentecostal Church*,[14] music in the Pentecostal church evolved first as a reaction to the establishment. Whatever the traditional church did, the Pentecostals did **not** do. If mainstream music was formal and scripted, the Pentecostals adopted spontaneity. Secondly, Pentecostal music was functional and pragmatic. If it achieved the desired results, it was used. Again according to Johansson, *"If a Pentecostal musician were asked*, '*What is the function of music in worship?*'" the likely answer would be, "*music is a vehicle through which we praise God.*" But further answers would be sure to follow: "It *can cover up noise, fill silence, create atmosphere, be a means of service, accompany singing and manipulate and control people. Church music, functional art that it is has the ability to do many things*."[15] It is not unusual for a Pentecostal pastor to encourage the musicians to play on, as it creates a mood. Johansson goes on to trace the evolution of Pentecostal music into the 20th century and notes that the manipulative character of the music was refined and honed into a high art. Over time the spontaneity gave way to a much more scripted experience. "Standing people were admonished to praise and clap their hands along with rhythmic body swaying and whatever singing might take place prompted by an ever present video screen. The song service's function was to make people *feel*, which music is readily able to do. Objective theological biblical connections (as found in hymns for example) were abandoned for the emotional euphoria engendered by endless repetition of CCM (Contemporary Christian Music) songs. Part of the technique for accomplishing this had to do with the raw psychological manipulative power that extremely loud rhythmic music has over the nervous system. As the century progressed, trap set and microphone teamed up to make Pentecostal music an overt exercise in congregational control."[16]

This perspective is shared by the Seeker Sensitive church as well, but perhaps not as bluntly stated. In his book *The Purpose Driven Church*, Rick Warren writes about the power of music.

> *"I'm often asked what I would do differently if I could start Saddleback over. My answer is this: From the first day of the new church I'd put more energy and money into a first-class music ministry that matched our target. In the first years of*

Saddleback, I made the mistake of underestimating the power of music so I minimized the use of music in our services. I regret that now."[17]

10.5.4 Architecture

There is a great deal of diversity in the architecture of the Experiential church. At one end of the spectrum there are very small country churches, seating perhaps 100 with no sound system at all. At the other end there are the Mega churches seating many thousands with media budgets in the many millions of dollars. Of course the larger buildings get our attention because that is where the money is. However keep in mind that a small rural evangelical style church wanting to buy their first sound system, and a small experiential church of the same size wanting to buy their first sound system, will have dramatically different expectations about what that system should deliver, Fig. 10-11.

Figure 10-11. Architecture determines the type of sound system required.

10.5.5 Acoustic and Sound System Considerations

Given the role of music in this type of worship, it is important that the building be designed around music. It is strongly recommended that the design of the Experiential church be a cooperative venture between an architect and an acoustical engineer, Fig. 10-12. These rooms need to be aural spaces rather than visual ones. Especially in the Seeker Sensitive church, it is highly unlikely that there will be any form of visual symbolism. This is not to suggest that these rooms be designed as concert halls. The Experiential church requires a performance space which has more in common with a movie theatre than a concert hall. The music will all be amplified and be high impact. The system needs to be capable of delivering levels in excess of 110dB. That is not to say that the levels in any church setting should be that loud but the system needs to be able to deliver that kind of level if there is to be enough headroom. Acoustically, this demands that the room be as dead as possible, that is the midband RT should be very low. A slight bass multiplier is tolerable as long as it is under one second. A bigger problem is building an acoustical shell that will contain and not absorb the bass. If gypsum board is to be used especially if there are large surface area to be covered, there needs to be at least two layers of 5/8 inch gypsum board on 16 inch center (o.c.) studs. If only one layer is used or wider stud spacing is used, there will be excessive low end absorption and it will be very difficult for the sound system to produce the impact that is desired. CMU (Common Masonry Unit, i.e., cement block) is the preferred material for the sanctuary.

Figure 10-12. Typical Mega church.

Focusing surfaces must be avoided. The rear walls need to be absorbent as possible so that the sound from the sound system does not reflect either back to the stage or to the congregation. There needs to be absorption on stage as well in order to help control the sound from the stage monitors. ITE (in the ear) monitors should be encouraged to keep the stage levels to a minimum. The one surface where there may be a difference between the Seeker Sensitive Experiential church and the Pentecostal derivative church is the ceiling. In the Seeker Sensitive church, congregational or participatory singing is not a big deal. It is assumed that the "seeker" will probably not know the songs and would not be inclined to sing anyway about an experience he may not have had. The ceiling can be very effectively used for absorptive panels or treatment. The Pentecostal church on the other hand, values "audience participation." Even though the Pentecostal church should have a very low reverb time, and the actual deck should be absorptive, ceiling reflectors 15 feet or so above the congregation will allow the worshipers to feel like their participation matters.

For many experiential style churches, a choir is an important part of the worship experience. Choirs are a challenge for a number of reasons. First it is a challenge to mic a choir. If the microphones are placed close to individual singers, the sound systems does not have to “work as hard,” but you lose the sense of the whole. If you back the microphones away from the singers you can capture more of the whole, but the sound pressure drops off and the system has to “work harder” to get the sound of the choir out to the audience. Some have tried to use large numbers of wireless microphones and attempt to mic a significant percentage of the choir. This requires not only deep pockets and very large consoles but sound systems with the necessary potential gain to allow large numbers of microphones to be active at one time. What many sound system operators do not seem to understand is that for any sound system/acoustic space combination there is a maximum amount of potential gain that the sound system can provide. That total gain can be dedicated to one microphone, or apportioned to 50 microphones. The total potential gain does not change.

The second challenge that choirs present, is finding a way for them to hear themselves. This is sometimes attempted with elaborate monitor systems, but this also can cut into the total gain before feedback.

A better solution is to include a choir shell into the design of the worship space. A choir shell can be a very effective way to let the choir be heard and for them to hear themselves as well.

The Experiential church can be good candidate for reverberation enhancement systems. The room can be built acoustically dead, and the enhancement system used to create the sense of a large and live acoustic space when needed.

10.6 The Community Worship Style

The community worship style is by far the least common of the four worship styles. It shares many of the traits of the others, especially the evangelical style, but differs in emphasis in very important ways, especially in terms of the acoustics of the spaces and the use of media in the service. We are using the word *community* to mean that the focus and centrality of the worship experience is about doing it as a group or *community* of believers. The presence or absence of the word *community* in the name of the church does not necessarily indicate a community worship style.

In these types of churches the building is most often referred to as a meeting house rather than a church. The community style of worship is really at the opposite end of the spectrum from the Celebratory style. There is not sacred space at all. The only thing that is sacred is the community itself. The congregants will feel like they have "done church" if they have felt like they were a part of the community of believers and were able to minister and be ministered to. In these congregations, the idea of the members *being Jesus to each other* is well developed and in no way seems sacrilegious. The Church of the Brethren one of the groups who typically worship the community style, have as their motto *Continuing the Work of Jesus. Peacefully. Simply. Together.* Those who are most likely to adopt this form of worship are those who are descendant from Anabaptists including the Mennonites, the Brethren, the Amish and others. There are some churches who practice a community style of worship but are not historically related to the Anabaptists, most notably the Moravians.[18]

10.6.1 History

Most of the churches who use the community style of worship trace their roots back to the Anabaptist movement in the Netherlands in the 1500s. The name Anabaptist means the “re-baptizers” indicating the belief that a someone baptized as an infant could be re-baptized as an adult. This may not seem like a big deal or something a “movement” could be started over. However in the 16^{th} century this was a very big deal. The Roman Catholic church took great offence at the notion that someone might be re-baptized because it implied that the first baptism was not good enough. In fact there were two offenses that the Catholic church of that era viewed as capital offenses; one was denying the existence of the Holy Trinity, the other was being re-baptized! One of the central figures of the early Anabaptist movement was Menno Simons (1496–1561). The modern Mennonite church takes its name from this radical reformer. The Anabaptists had the dubious distinction of being the most persecuted of all the groups which came out of the reformation. Their belief in re-baptizing was enough to set the power of the Roman church against them. They also practiced pacifism refusing to serve in the military, citing Jesus’ teaching to love ones enemies. This of course set them up for persecution by the state. Many thousands of Anabaptists were tortured and killed in the 16^{th} century. They were simple folk, farmers mostly, who believed that their faith in God should be manifest in their relationships to their fellow man. They met each week in private homes often in secret, hence the use of the term meeting house rather than church, Fig. 10-13.

Figure 10-13. A Quaker Meeting House, Salem, MA.

10.6.2 Community Worship

Because of the emphasis on relationships and community, it is difficult to describe a typical Community worship service. On the surface it may look very similar to the evangelical style. In the community style there may be the usual elements of prayer, music, the Lord's Supper or communion, and teaching, but the emphasis will be markedly different from the other styles.

In this style of worship, prayer has the same meaning as it does for the evangelical; it is talking with God. However in these churches the prayers are more likely to be supplications on behalf of members of the church community in crisis, or on behalf of the poor or disenfranchised in the wider community. It is not unusual to hear prayers of supplication asking God to bless our enemies, meaning individuals with whom we may have unresolved issues as well as the enemies of the state. When there are acts of violence, Anabaptists will pray as much for the perpetrators of the violence as for the victims.

In the Community church tradition, music has a checkered past. Historically there were those of this tradition who prohibited musical instruments of any sort in the meeting house. Those who held this view, often had very well developed congregational singing with four or more part harmony. Others allowed instruments in the meeting, but the role was always as accompaniment. The notion of performance music would be out of place as the emphasis in the Community church is never on the individual but rather on the collective. Four part a capella singing is, for many in this form of worship, an exemplar of the community of believers forming the church.

Many contemporary Anabaptists have adopted more complex forms of music into the worship service. Even though there may be instruments and contemporary forms, the emphasis will still be on community rather than performance. Anabaptist communities find it difficult to use contemporary music from other traditions as it is often very individualistic and emotional. These values are in tension with Anabaptist core values which place the community above the individual.

In this style of worship there will be a wide variety of kinds of preaching. Many churches who practice this style will refer to this activity as teaching rather than preaching. It will generally consist of exegesis of some scripture passage and usually some very practical application to the life of the community. Some will adopt a more evangelical approach and preach a sermon that may fit very well in an Evangelical church. The difference is again in emphasis. In this style of worship the preaching/teaching is not the highpoint of the service, although it is an important part.

10.6.3 Acoustic and Sound System Considerations

In the Celebratory church we have sacred space where every element is there to provide a context for the celebration of the Eucharist. In the Evangelical church, the Word is preeminent and the acoustics must be optimized for intelligibility. In the Experiential church it is the total experience that informs the acoustic and sound system choices. Finally, in the community style of worship it is most important that the congregation be able to hear themselves and feel as though they are a part of the whole. This particular form of worship oddly enough finds itself in a very widely divergent collection of architecture. There are some Mennonite churches for example which most assuredly are committed to the values of community based worship, but who worship in almost Gothic style churches, which seem at odds with the style of worship practiced there. Others meet in purpose built or repurposed buildings and meet in the round. Interestingly enough when doing research for my book on this topic.* I found a church which clearly was of the community worship style, but meeting in a building that seemed totally wrong for this style. It was a Moravian church, not part of the greater Anabaptist movement, but with a strong commitment to community and wanting their worship to reflect those values. Their core values were conspicuously displayed on a poster in an ante room attached to the main worship space. The very first value on the poster was *community*. However, the architecture of the church building says nothing about community. Looking at the exterior, the steeple tower dominates the rectangular building which houses the sanctuary. It is a simple structure fitting in with the rural surroundings, but it is reminiscent of a Baptist church or a church with a strong hierarchy like Presbyterian or perhaps even Episcopal. Inside, there are pews in rows, all parallel to the raised stage. There is a pulpit conspicuously in the center of the stage with a rail in front, as if to separate the congregation from the clergy, Fig. 10-14. So I asked the pastor if they were comfortable in this building. I was wondering if the architecture of the church building fit the core values of the church community. Her answer was immediate and forceful, "*absolutely not*!" She recounted that the building was built some years ago by an architect who was a Presbyterian and who included those elements that *he* felt were important for worship. She longed for a structure that would reflect the values of the congregation. She would have liked seating in the round or at least in a semi-circle which would remove the "*paternalistic division*" (her words) between clergy and laity.

Figure 10-14. A typical community style of worship church.

In this worship style, there is no sacred space, there are no sacred elements, there is little need for media. Once again the governing principle should be "do no harm". The architecture should promote and support community. When Reba Place Church in Evanston Il. was repurposing a car paint shop into a meeting house, they chose to build the seating risers in a half circle, intending to complete the circle if growth ever demanded it. When asked why the semicircle (or circle) the reply was "*It is hard to carry unresolved resentments if you have to sit face to face with your neighbor in the meeting* *house.*"[18]

The governing acoustic aesthetic, should be intimacy. Fortunately churches that worship in the community style, are rarely large. It can be challenging to create a sense of intimacy in very large rooms. Generally intimacy is the result of placing reflecting surfaces fairly near the congregations. Low ceilings, relatively small rooms will all result in a short initial time gap and contribute to the impression of an intimate space. Intelligibility is also a concern because there is always a teaching and a significant amount of various sharing's in this style of worship. However, intelligibility in small, intimate non-reverberant spaces is rather easy to accomplish. Unless the ceiling is too low, a single point system is ideal as long as coverage is adequate.

10.7 Non-Christian Worship Spaces

This chapter obviously deals with the Christian church for reasons stated in the introduction. However, many of the principles in this chapter can apply to other religious spaces as well. If the religious service is constructed around very well defined traditions and rituals, it is likely that it will be similar to the celebratory style of worship, in the sense that the space will likely be considered sacred, and technology might be viewed as intrusive. If on the other hand the service is didactic in nature with an emphasis on teaching or dialog, then perhaps the evangelical style would serve as a model of sorts. If there is an emphasis on meditation and sharing, there may be similarities to the community style of worship. Maybe the most important point that can be made is that in worship spaces, the architecture matters. The acoustics matter. And not everyone will have the same need for media technology. Be sensitive and leave your presuppositions at home.

References

1. http://research.stlouisfed.org/fred2/series/TLRELCONS.
2. http://www.washingtonpost.com/wp-srv/metro/documents/aris030609.pdf.
3. Douglas R. Jones, *Sound Of Worship*, Focal Press 2011. p. 2.
4. *Sacrosanctum Concilium 1963*, chapter IV Sacred Music, articles 120, 121.
5. Steven J. Schloeder, *Architecture in Communion: Implementing the Second Vatican Council Through Liturgy and Architecture*, Ignatius, San Francisco 1998.
6. Michael S. Rose, *Ugly as Sin: Why They Changed our Churches from Sacred Places to Meeting Spaces and How We Can Change Them Back Again*, Sophia Institute Press, Manchester NH, 2001.
7. Christopher A. Hall, Worshipping with the Church Fathers, InterVarsity Press, Downers Grove IL 2009. p. 21
8. Douglas R. Jones, *Sound Of Worship*, Focal Press 2011, p. 97ff.
9. Richard McBrien, Lives of the Popes; Harper San Francisco 1997 P 267.
10. Douglas R. Jones, *Sound Of Worship*, Focal Press 2011, p. 130.
11. Ibid, p. 147.
12. Ibid, p. 149.

13. Eric Patterson, Edmund J. Rybarczyk, *The Future of Pentecostalism in the United States*, Lexington Books 2007, chapter 4. Music in the Pentecostal Church by Calvin Johansson, pp. 50–63.

14. Ibid, p. 56.

15. Ibid, p. 57.

16. Rick Warren, *The Purpose Driven Church*, Zondervan, Grand Rapids Michigan 1995, p. 279.

17. Douglas R. Jones, *Sound Of Worship*, Focal Press 2011. p. 174.

18. Private correspondence, C. Lehman, Senior Pastor, Reba Place Church, 2010.

* The term “evangelical” is used in this section to describe a particular style of worship. It should not be confused with evangelical as a theological or doctrinal position.

* Sound of Worship, Focal Press. 2011.

Chapter 11

Stadiums and Outdoor Venues

by Ron Baker and Jack Wrightson

11.1 Stadium Types and Geometries

There is no "typical" stadium across all sports and seating capacities. While there are distinct similarities, say among stadiums built for specific sport of similar seating capacity, their architectural design, the types of sports accommodated and wide variances in seating capacity lead to very different structures.

Common to all, is the desire for intelligible announcements in the seating bowl. How this is accomplished is highly dependent on the geometry of the seating bowl.

11.1.1 Rectangular, Oval and Baseball Stadiums

In most of the world, stadiums designed for sports that play on a rectangular field are the most common, Fig. 11-1. This includes American football, football/soccer, lacrosse, rugby, tennis, etc. Oval fields are utilized for cricket and Australian football, while the oval pitch is much larger than a standard American football field; the seating bowl designs are quite similar in their geometry, Fig. 11-2. The most unique, large stadium type is for baseball, to accommodate the diamond shaped infield and half circle shaped outfield, Fig. 11-3.

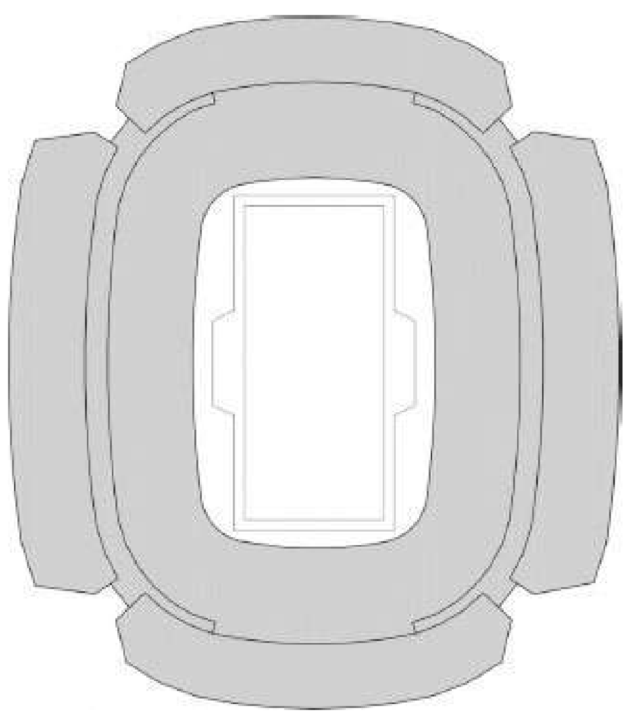


Figure 11-1. Plan view of a football or soccer stadium showing an example of a typical seating arrangement.

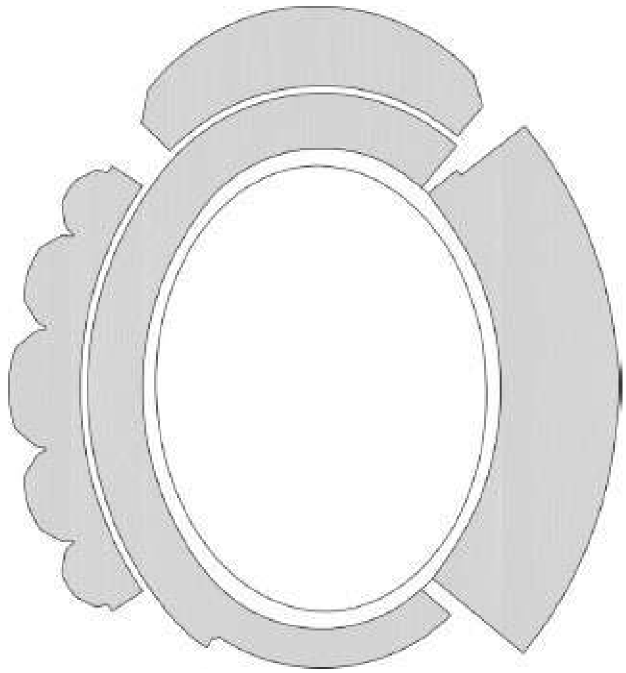


Figure 11-2. Plan view of a cricket or Australian football stadium showing an example of a typical seating arrangement.

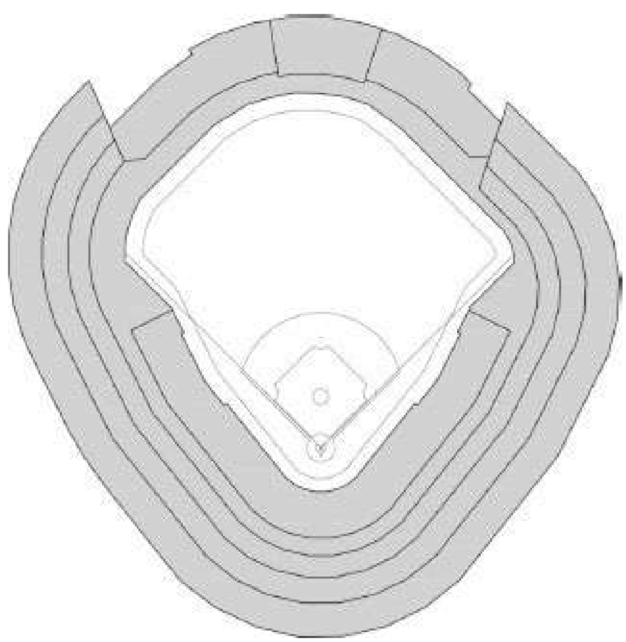


Figure 11-3. Plan view of a typical baseball stadium showing an

example of a typical seating arrangement.

Many smaller stadia have seats only along the long axis/sidelines of the playing field. In the case of baseball stadiums, seating may only be located directly behind home plate. While an interesting problem on their own, very small stadiums are not the focus of this chapter.

11.1.2 Seating Bowl

Most of this chapter will be concerned with seating bowl sound systems, as they are the most important and costly element of a stadium audio system. Other typical systems are provided in the overview below.

The seating bowl is defined by the fixed seating areas of the building. In most cases, this is easily identified by the permanently installed seating; however it can include grass berms, temporary seating elements, retractable seats and standing areas. It is important to identify all potential ticketed viewing areas to be sure that the sound system design is covering all locations where a paying customer may be located.

It is also important to recognize sightlines for seated and standing fans to not only the field area, but also important game presentation elements such as scoreboards, video displays and flags. The design of the seating bowl sound system must not only achieve the desired acoustical performance, but also be coordinated with other technical and architectural elements of the buildings while maintaining good viewing.

11.1.3 Control Room

The audio control room typically will serve as the central content origination and distribution point for audio through public spaces in any stadium. The control room should be located to provide good sight lines to the crowd and field, so as to be able to respond to crowd and game activity, but also to have a good listening environment to hear the seating bowl sound system as well as the crowd, Fig. 11-4. This requirement recommends that the control position have operable glazing and be located where it is "in the bowl" from a sound system coverage standpoint. This is not always possible, due to architectural and seating capacity requirements. In these cases, a remote mixing location in the seating bowl should be considered.

11.1.4 Entrances and Plazas

There are two types of common entrance systems. The first is dedicated to addressing those fans in lines at ticket windows, Fig. 11-5. These are often split out from entrances as messages played to these patrons are often different, and a live microphone announcement will likely be directed only to those in those lines. Announcements to these patrons often involve informational messages to help expedite the ticketing transactions. Coverage of this system should be able to extend out at least 10 meters (m) (30ft).

Figure 11-4. Stadium audio control room.

Figure 11-5. Ticket window area showing overhead loudspeakers for announcements

The main entrance systems, however, provide sound and messages to fans entering the stadium, Fig. 11-6. A mix of pre-recorded messages (i.e., “no cans, no bottles”) radio broadcast and pre-game music and live announcements are a common programming mix for these systems. The system should be designed so that messages and announcements can be clearly heard at a distance of at least 40m (75ft).

At most stadiums, the entrance loudspeakers are limited to the primary public entrances. Staff, media, player, etc. entrances are generally not provided with systems. In some cases this is also true of VIP and premium entrances. Occasionally, there will be public access around all or most of the perimeter of the stadium with significant pedestrian traffic along this perimeter before and after games. In these cases, an entrance system which covers this public perimeter should be discussed.

Plazas are connected to the entrance(s), but are used differently by the public as gathering, meeting and entertainment spaces primarily before, but also after events. In these cases, audio systems may be temporary or permanently installed (depending on anticipated frequency of use), and utilized for musical performance, TV and radio broadcasters, live host promotions, etc. The concept is that the plaza becomes a pre and post-match gathering space with entertainment provided to enhance attendance, not merely provide information for patrons on their way into the stadium.

Figure 11-6. Stadium entry gate showing loudspeakers for crowd announcements.

11.1.5 Public Concourses

Most large stadiums include one or more public circulation concourses within the ticketed perimeter of the building. These concourses feature toilet rooms, food and beverage concessions and merchandise sales stands, along with access to other interior spaces and stadium facilities.

Due to acoustically “hard”, abuse and weather resistant architectural finishes and high peak crowds, the concourses can be very noisy environments, especially before and after games. These high noise levels can make intelligibility of messages and announcements difficult at other than annoyingly high sound levels.

For this reason, many facilities forsake continuous, blanket coverage of sound in the concourses in favor of a targeted approach to covering patrons standing in line, in food consumption areas or within the toilet rooms.

The non-continuous coverage can have life safety code ramifications which are discussed in section 11.4 Life Safety and Messaging Considerations.

11.1.6 Clubs and Premium Spaces

These spaces are defined as interior, without seating in the spectator bowl. The key design distinction for these rooms centers on if they are intended to serve solely as game day social, food and beverage locations or also will host non-game day or pre-match meeting Fig. 11-7.

Figure 11-7. Typical stadium club space.

In the former case, simple foreground music and broadcast/game audio systems are normally sufficient. For the latter use, the spaces may have built-in audio, video display/projection and presentation flexibility to rival any convention center meeting room or corporate board room. The level of AV services required has been found to vary considerably across buildings.

11.1.7 Suites

Three general classes of suites are commonly seen. The first is the “regular” suite, which features seating that is part of the spectator seating bowl and a rear flat floor area, for socializing or dining without a direct view to the playing field. The fixed stadium seating in the suites is treated the same as the rest of the seating bowl, hearing the seating bowl public address sound system. The rear, flat floor area may or may not have supplemental audio or AV systems. Many suites include large TV sets with reliance on the TV distribution system for any supplemental audio, without a need to hear PA announcements in this area. Suites that feature formal, sit down dining areas, will often include additional small, local audio systems as they are often occupied well in advance of the game start time.

The next category can be considered as ultra premium/founders/ or owner’s suites. These suites may also be much larger than a typical suite, hosting large groups. The audio and AV systems in these areas are often completely at the discretion of the suite holder and can range from nothing more than TV sets to elaborate home theatre style systems connected to stadium audio and video feeds with comprehensive remote control systems.

The third suite architectural category is the "bunker suite" which is a space that has no direct visual connection with the playing field and is often on the field or event level of the building. As such, the environment in these spaces is typically enhanced with the addition of local audio and AV systems, so that additional programming, including local and client provided, mobile sources can be utilized.

11.2 Seating Bowl Performance Requirements

Regardless of the size of the stadium, its geometry or sport being played there is a fundamental performance requirement that must be achieved by the seating bowl sound system design, installation, set up and operation. In many stadiums each of these four areas must be optimized to achieve a good result.

Maximum Loudness. There is considerable popular confusion over how loudly the stadium sound system must play. As with any successful voice communication system, the audio system must be able to play loudly enough to be heard over background noise, which includes all the sounds in the stadium environment, including existing ambient sounds (such as air, road and rail traffic) stadium mechanical systems and more importantly crowd noise.

The most common misconception is that the stadium sound system must be able to be understood over the loudest crowd noise. This assumption is incorrect for several reasons.

First, as one can experience easily for themselves, it is very difficult to understand voice communications when one is yelling.

Second with well documented, properly measured crowd sound levels at sports facilities in the range of 105dBA–115dBA, achieving the requisite 115dBA–125dBA public address (PA) sound levels to assure intelligibility (by conventional metrics) will result in the potential of claims against the stadium for hearing damage.

A third consideration is the significant incremental cost of the additional sound system loudspeakers and amplification required to achieve very high levels. When one considers the percentage of time during a match that crowd noise is at its most intense, it can be hard to justify the additional cost due to their limited benefit.

A more reasonable cost to performance ratio is a maximum sound level for the PA system of 105dBA. We have found that systems designed to this standard are well received and considered more than acceptably loud in the sporting environment.

Some events, such as baseball and cricket or facilities with smaller capacity have found that maximum sound levels in the 98dBA–102dBA range to be equally acceptable.

These recommended maximum sound levels are for the continuous levels the system is capable of sustaining.

Frequency Response. The frequency response required is dependent on the program to be played over the system. While more and more stadium game presentations (especially professional sports) depend on high energy, full bandwidth musical reproduction, some are for voice announcements only. Table 11-1 gives recommended starting points for design frequency responses, and may be appropriately altered depending on the requirements of the project, what is possible in a specific acoustical environment and budget.

Table 11-1. Recommended Frequency Response for Various Sound Systems

Voice Only Systems	150Hz –6000Hz ±3dB.
When good musical quality is required	50Hz–8,500Hz, ±3dB, with capability to be adjusted to achieve a frequency response of +12dB from 80Hz– 250Hz relative to 1000Hz. This allows for a contemporary, bass heavy musical sound to be achieved, if desired.

Uniformity of Loudness. This is another often misunderstood and applied metric in stadium sound system design. The ability for a sound system to be perceived as equally loud across the fixed seating is very important, as changes in level, up or down, will be perceived as such in all the seats. What is often not appreciated, however is that the ambient sound levels may vary considerably across the seating areas, most notably in under balcony areas, requiring higher sound pressure level (*SPL*) values to achieve the same perceived loud-ness (relative to the background noise) and intelligibility. For this reason, systems set up to have uniform sound levels in an unoccupied stadium will often experience poor intelligibility in high noise areas.

Where the seating and noise level conditions are the same, the desired direct sound uniformity target is ±3dBA and ±2dB in the 2000Hz octave band.

Multiple Arrivals and Intelligibility. Stadiums, due to their sheer size, often feature large sound reflecting surfaces along with multiple loudspeaker locations. These elements can create sound quality and speech intelligibility problems due to the delayed arrival of high intensity sound reflections relative to the direct sound at a given seat location. There are several useful design guidelines that address this issue, however it has been discovered that the nature of currently well accepted quantitative speech intelligibility measurements provides meaningful results in areas with multiple arrivals.

The most common, widely accepted, low cost and repeatable methods of making intelligibility measurements are the devices that measure STI-PA. While there is some concern that the correlation between perceived speech intelligibility may break down somewhat in large, reverberant spaces, STI-PA, both from a design and as an as-built performance standard is proving to be valuable and is a useful guide and documentation of performance in that systems that exhibit suitable STI-PA values are well accepted by patrons and facility owners. Additionally, large stadium sound systems that yield good STI-PA results, also exhibit highly acceptable subjective sound quality, along with speech intelligibility. For this reason STI-PA can be considered as a very useful single figure metric for both design and proof of performance evaluation.

Other well accepted intelligibility measurements are also useful and in some cases required by standards/codes, such as CIS[*].

As a design goal, it is recommended that the design calculations and/or modeling be based on a closed model and take into account at least 3rd order reflections to achieve a STI-PA value of not less than 0.55 over 95% of the fixed seating and known standing/berm type seating areas.

A measured value of 0.50 can be considered to be the minimum acceptable level and aspirations to higher values are recommended.

11.3 Typical Seating Bowl Loudspeaker

Configurations

While many variations exist, there are three primary loudspeaker configurations successfully employed for stadium seating bowl sound systems.

11.3.1 Point Source Cluster Systems

The first is the single point cluster. This approach has a long history, due to both function and cost reasons. The configuration places the majority of the loudspeakers into a single array, projecting sound, over very long (often up to 800ft/450m) distances. The main loudspeaker array is most commonly located at one end of the long axis of the stadium although successful designs have located the point cluster at the center of the short axis/sideline seating, Figs. 11-8 and 11-9.

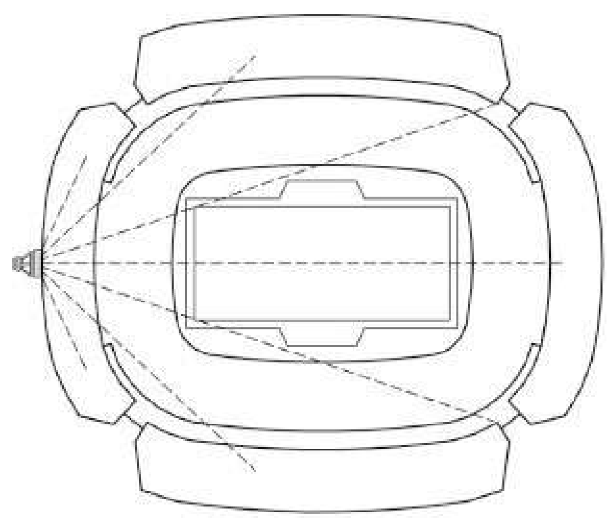


Figure 11-8. Stadium plan view showing placement of end zone point cluster.

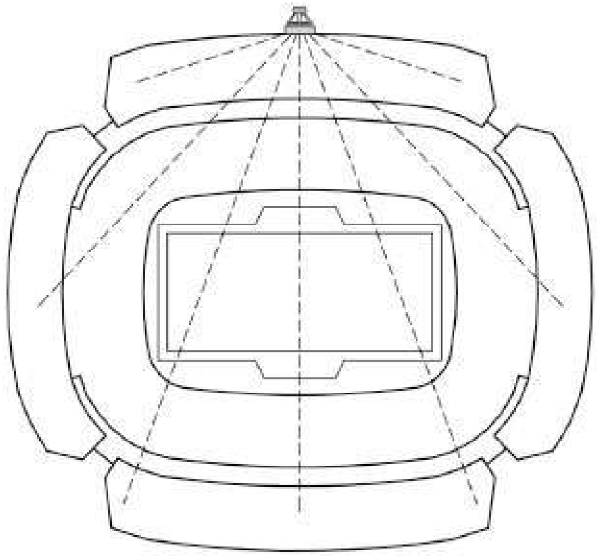


Figure 11-9. Stadium plan view showing placement of sideline point cluster.

In addition to the main cluster, supplemental loudspeakers are often required to “fill” seating areas that are blocked from direct line of sight from the main array, or are in locations where the noise level is substantially higher than in the general seating, such as underneath overhanging seating sections, despite line of sight to the primary cluster. To maintain synchronization with the point cluster, the fill loudspeakers would need to be signal delayed by appropriate amounts.

To achieve the desired sound levels and coverage in large stadia, the loudspeaker arrays are necessarily large, and have been successfully designed with conventional, high output loudspeaker devices, line arrays and specialty systems (i.e., EAW 900 loudspeaker) designed to be arrayed in large clusters with programmable steering, Figs. 11-10 and 11-11.

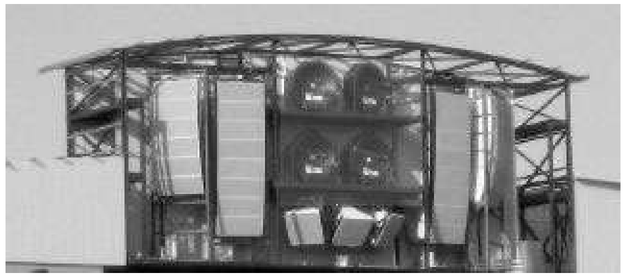


Figure 11-10. Collegiate end cluster located above scoreboard.

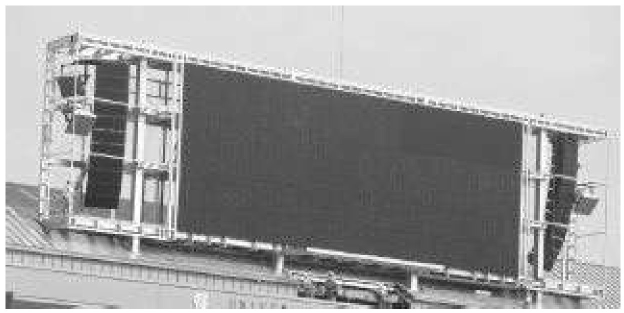


Figure 11-11. Collegiate end cluster located on either side of the scoreboard.

Point source clusters have a number of advantages; primarily that they are the lowest cost way of providing acceptable sound quality to large seating areas. They also require the least amount of structural support and ac power infrastructure compared to the other options.

Disadvantages of the point source or end cluster include:

- Lack of synchronization with video programming due to time of arrival differences for sound and light.
- Sound and intelligibility can be affected by wind currents and temperature gradients, although of these two, wind has the much greater impact.
- High frequency extension and sound quality is diminished by air absorption over long distances.
- If these venues are located near residential areas, audible sound

from the reinforcement system can extend into the neighborhoods and possibly be disturbing unless the loudspeaker placement can be oriented in a direction that projects the sound away from the more sensitive areas.

Despite these very difficult to overcome issues, point source clusters remain a viable and popular option, especially in USA collegiate football stadiums. Improvements in loudspeaker products have enhanced sound quality for these systems.

11.3.2 Atmospheric Effects

Sound propagation is subject to the vagaries of the medium in which it exists. The air in outdoor venues has variable temperature, wind, and relative humidity. The effect of wind is twofold. Wind speed near to the ground is ordinarily less than at a higher elevation. This causes sound waves propagating in a direction into the wind to be diffracted upward while sound waves propagating in the same direction as the wind to be diffracted downward. Cross-winds shift the azimuth of the propagation direction towards that of the wind. Thus wind can cause shifts in apparent loudspeaker aiming points. Additionally, sound will propagate greater distances with the wind than against the wind. A gusting or variable wind introduces a temporal quality to these properties. The effect on a listener is that the sound intensity appears to be modulated as the wind gust rises and falls, a layman's description being "it fades in and out."

Sound speed is influenced by air temperature with higher temperatures corresponding to increased sound speed. This relationship is given by

$$c = 20.06\sqrt{T} \tag{11-1}$$

where,

c is the sound speed in m/s,

T is the absolute temperature in K.

A fixed air temperature has no influence on propagation direction, but thermal gradients can be a source of further diffraction effects. Normal thermal gradients correspond to a temperature decrease with increasing elevation. Such a condition diffracts sound waves upward such that the apparent direction of propagation is elevated. A temperature inversion gradient has just the opposite effect producing an apparent depressed direction of propagation. The severity of these effects obviously depends on the size of the thermal gradients. Typically encountered stadium situations can result in shifts of 5° or more over a distance of 200m (650ft). These effects are illustrated in Fig. 11-12.

Atmospheric absorption of acoustical energy ultimately amounts to the conversion of the energy associated with a sound wave into heat energy associated with the random thermal motion of the molecular constituents of the air. Air is basically a gaseous mixture of nitrogen, oxygen, and argon with trace amounts of carbon dioxide, the noble gases, and water vapor. With the exception of argon and the other noble gases all of the constituent molecules are polyatomic and thus have complicated internal structures. There are three mechanisms contributing to the sound energy absorption process. Two of these, viscosity and thermal conductivity, are smooth functions of frequency and constitute what is called the classical absorption. The third or molecular effect involves transfer of acoustic energy into internal energy of vibration and rotation of polyatomic molecules and into the dissociation of molecular clusters. This third effect is by far the most dominant at audio frequencies and explains the complicated influence of water vapor on atmospheric absorption. The detailed behavior given in Fig. 11-13 is illustrative of these effects at a temperature of 20°C whereas the approximate behavior given in Fig. 11-14 is more useful for general calculations.

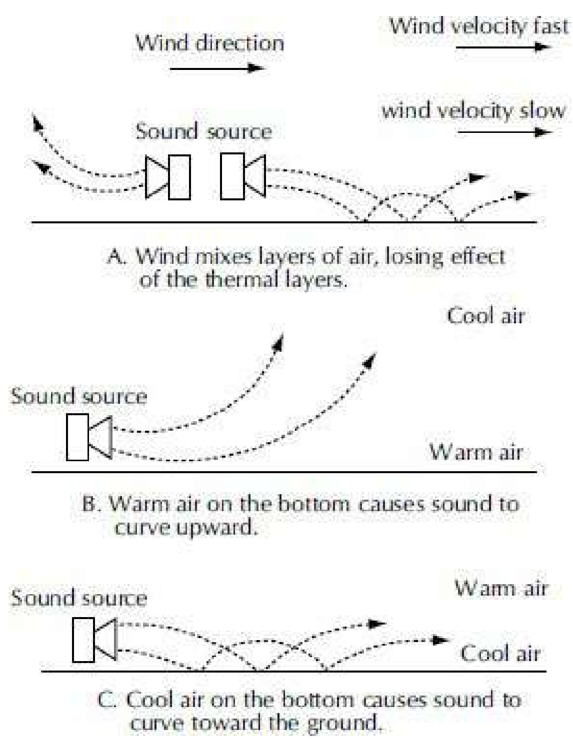


Figure 11-12. The effects of wind and thermal gradients on sound propagation.

Table 11-2 is extracted from Fig. 11-13 and illustrates the severity of the absorption effects.

Below 1kHz the attenuation is not significant even for a 200m (650ft) path length. The relative humidities encountered in practice usually lie in the 10% to 100% range and it can be seen that for frequencies below 5kHz that wetter air is preferable to drier air.

Table 11-2. The Entries Are Attenuation Values in dB/m for Various Values of Relative Humidity at 20°C

RH	0.1kHz	1kHz	2kHz	5kHz	10kHz	20kHz
0%	0.0012	0.0014	0.002	0.0052	0.019	0.07
10%	0.00053	0.018	0.053	0.11	0.13	0.20
100%	0.0003	0.0042	0.010	0.045	0.15	0.50

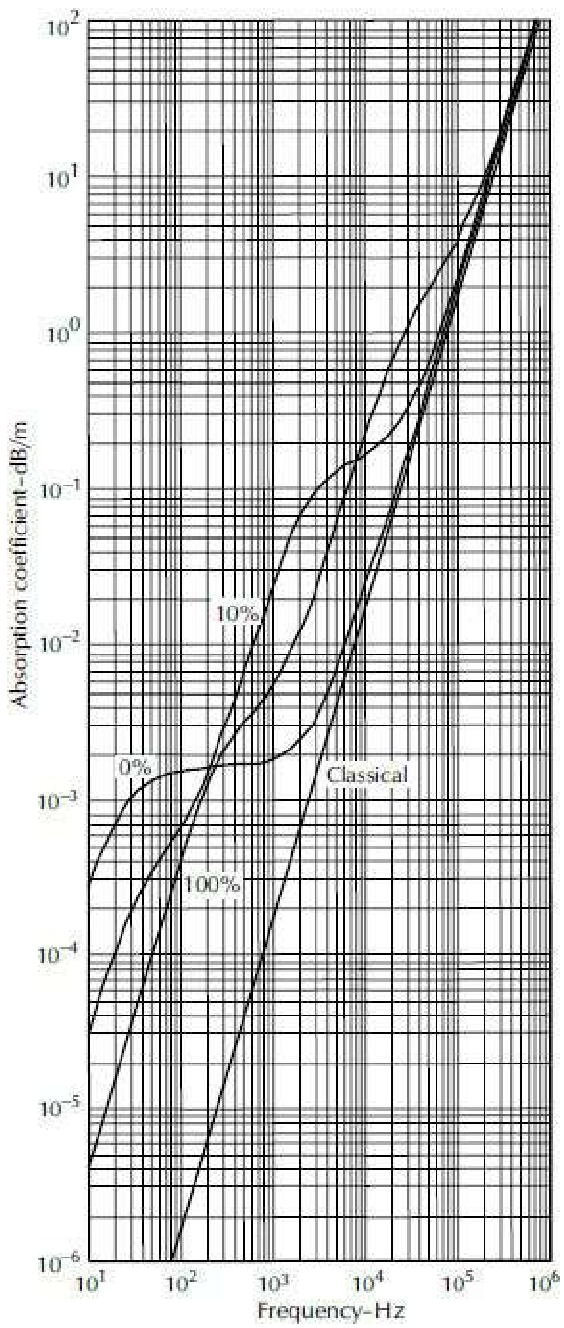


Figure 11-13. Absorption of sound in air at 20°C and one atmosphere for various relative humidities.

High-frequency equalization to compensate for air losses is usually possible up to about 4kHz with the amount of equalization required being dependent on the path length. Note that on a dry fall afternoon, the attenuation at 5kHz over a 200m path length is about 22dB. No wonder that a marching brass band on such a day loses its sparkle. As a consequence, long throws in a single source outdoor system are limited to about a 4kHz bandwidth.

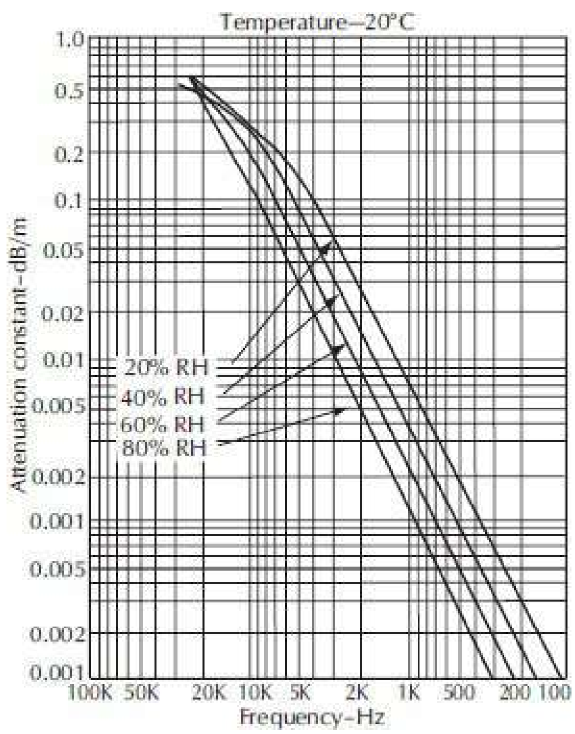


Figure 11-14. Absorption of sound for different frequencies and values of relative humidity.

11.3.3 Highly Distributed Loudspeaker Systems

A highly distributed loudspeaker system utilizes much smaller loudspeakers, placed primarily on existing stadium seating structure to cover the fixed seating areas. This results in, on average, approximately one loudspeaker (or a small group of loudspeakers) for each seating section, Fig. 11-15.

Highly distributed systems have been used as far back as the early 1960s for stadiums, with paging horn type devices being the most widely used transducer type. These fell out of favor as the rise of large scale amplified music systems developed better sounding, higher output loudspeaker devices, causing a shift to large, point source arrays as the dominant loudspeaker configuration for large stadiums. This began to change in the late 1980s, again partially due to the development of smaller, well controlled, wide bandwidth package loudspeaker systems that dramatically improved sound quality and output in comparison with paging horns. Coupled with the aesthetic and architectural desire to eliminate the large mass of the loudspeaker arrays of the era, a new interest in highly distributed systems was born, Fig. 11-16.

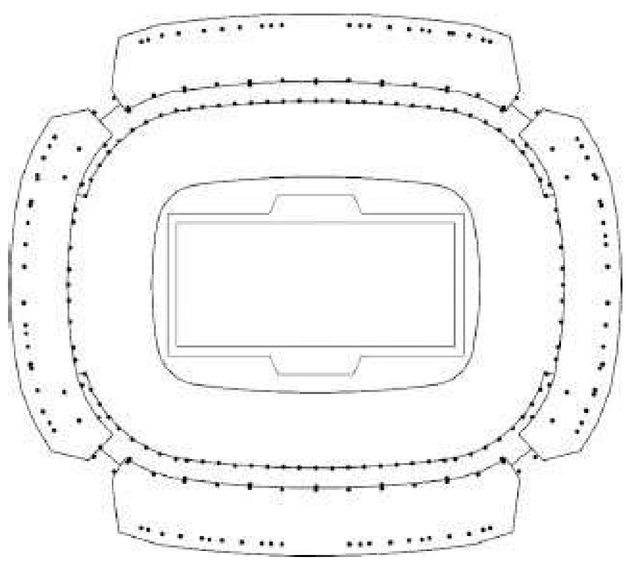


Figure 11-15. Stadium plan view showing location of distributed loudspeakers. Each dot represents one or two loudspeaker cabinets.

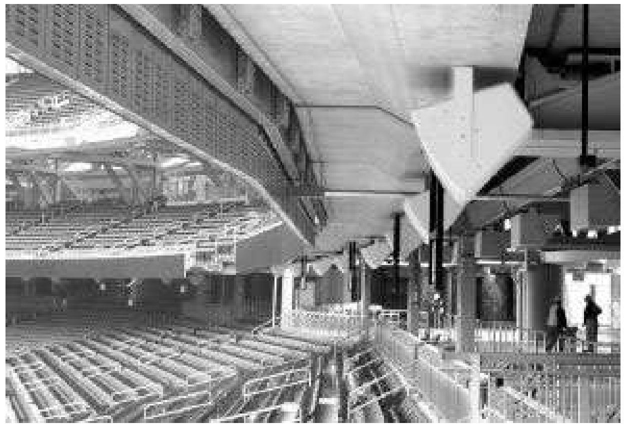


Figure 11-16. Distributed loudspeakers mounted below an overhang.

The success of a highly distributed system is often dependent on the ability of the designer to be able to persuade architects and owners to place the loudspeakers where they will perform acceptably well. As the designer is limited to placing loudspeakers on building structure, or in some cases, dedicated structures for loudspeakers above the uppermost rows of seating, achieving the uniformity of loudness across seating areas can be difficult, Fig. 11-17.

Figure 11-17. Distributed column loudspeakers mounted at back

of seating area.

Fig. 11-18 illustrates a typical cross-sectional geometry for a large stadium. As can be seen, there are a limited number of possible loudspeaker positions that will allow for good acoustical performance while avoiding sightline obstructions or expensive additional structure.

For these reasons and others including potential conflicts with signage locations (particularly on the front of seating sections), there are stadium geometries where a highly distributed system cannot be designed to have acceptable performance.

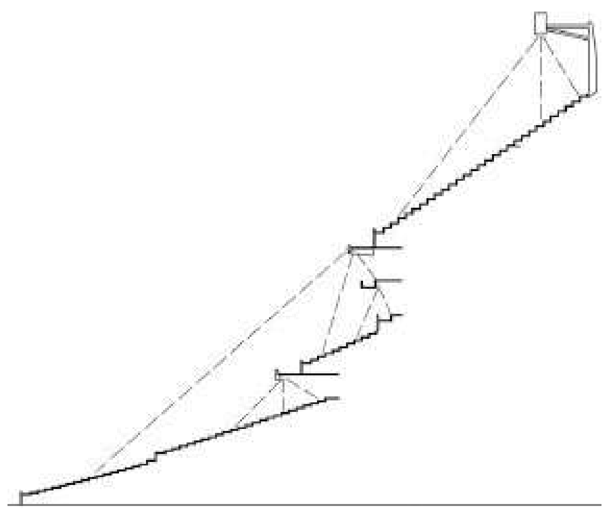


Figure 11-18. Section view through a stadium showing loudspeaker mounting points and general loudspeaker orientation.

As might be expected, the success of the system is highly dependent on the design configuration and system set up so that perceived uniform sound levels are maintained and, in particular, arrivals from multiple loudspeakers at a single listening location are managed in time and level so that acceptable speech intelligibility is achieved.

Another consequence of the large number of loudspeakers and associated amplifiers and processing equipment is a commensurate increase in installation and maintenance costs.

Given these potential design challenges, additional cost and maintenance burdens, why have highly distributed loudspeaker systems become so popular with owners and operators?

- Dramatically improved synchronization with video programming in comparison with point source cluster systems.
- Improved perception of fidelity. The proximity of loudspeakers allows for a more immediate sound quality that is preferred by many listeners.
- Elimination of the structure and aesthetic impact of a large array of loudspeakers, either as part of an end cluster or distributed cluster system.

11.3.4 Distributed Cluster Systems

This configuration is very common in indoor/domed stadiums and those with large roof canopies. The configuration becomes viable from a cost and design standpoint with the presence of structural attachment points created by the roof or canopy structure.

As might be imagined, the design challenges are a combination of both the point source and highly distributed loudspeaker system types. Cost is also, assuming similar equipment specifications, usually in the middle between the two other options.

The system can be conceived as breaking up a point source cluster and locating segments of that array around the circumference of the playing field or along the sides of the pitch. In reality, many more loudspeakers are required to adequately cover the seating bowl than would be required for a single point array, Figs. 11-19 and 11-20.

As with highly distributed systems, the design and system set up must manage the multiple direct sound arrivals at any seating location from the multiple cluster locations. For this reason, systems with a very small number of clusters, such as four arrays in the corners of a large stadium, often exhibit significant time arrival problems in the coverage overlap zones between clusters, which are necessarily asymmetric and cannot be fully addressed with electronic signal delay.

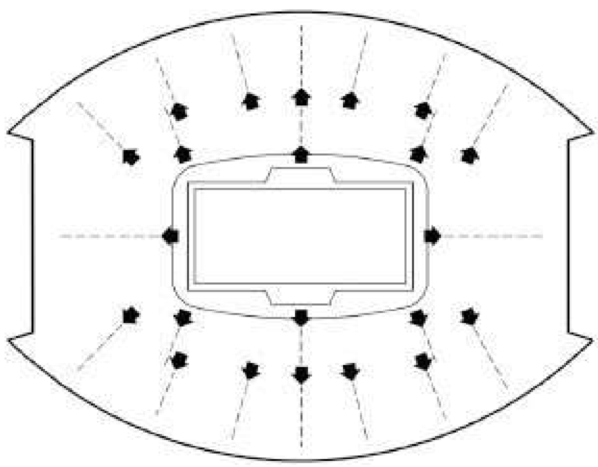


Figure 11-19. Retractable roof stadium showing locations of distributed line array clusters.

Figure 11-20. Line arrays suspended from structure within retractable dome stadium.

For large, indoor stadiums (and under very large roofs for "open air" stadiums as well), the design of the array locations must also take into account the often difficult acoustical conditions, which compound the multiple arrival challenges and make study of the calculated direct signal levels to all other arrivals and calculated STI values essential.

Fig. 11-21 is a simplified block diagram of a stadium distributed loudspeaker system.

11.4 Life Safety and Messaging Considerations

For many projects, there is a regulatory requirement for the seating bowl sound system to provide for life safety, evacuation, etc. messages. Requirements differ across jurisdictions, however there are elements which are common that should be investigated and incorporated or waived by the authorities, as required.

Many national and local regulations require a certain level of speech intelligibility if the system is to be used for voice notification. This requirement is generally listed in terms of STI or CIS. In some jurisdictions it may be necessary to determine if STI or STI-PA measurements will be accepted for proof of performance for the as-installed system. It is important for the system design to recognize any such requirements and that the modeled or calculated values for intelligibility consider the acoustical environment and any other "imperfections" that can degrade the as-built, measured values. Modeling and calculating solely the direct sound arrivals will almost certainly result in an overly optimistic predicted STI or CIS value.

Where used for life safety purposes, the applicable building codes may require that all the components of the system be rated for fire alarm use, include loudspeaker line supervision or other such features which are not common to high quality, professional sound systems. In most cases, these codes were not written in anticipation of the specialized nature and large size/capacity of large stadia and arenas. The available equipment that does feature the appropriate ratings to meet regulations does not offer the appropriate acoustical performance (i.e., bandwidth, sensitivity, directionality) suitable to achieve the performance goals of the project. In these cases, a waiver of the regulatory requirements for the equipment is often sought and granted with discussion of staffing (technical and fire department) during occupancy, system on/off, computer monitoring of lines, etc., required to achieve approval of the non-rated equipment.

Some regulations require the system to have the availability for zoned evacuation of the seating bowl, which requires the ability to direct a live voice announcement to selected, limited zones of the seating area. In these cases the system's electronic configuration and loudspeaker/signal processing zones must be configured to allow the desired addressability to particular zones.

The audio quality of live announcements for the notification/emergency can generally be controlled as part of the design and operation. This is not necessarily the case if the message is to be a pre-recorded file from a conventional life safety system. In some cases, the audio quality of such messages is poor and will not allow for good intelligibility, regardless of the measured STI or CIS values achieved by the system. Where such messages are to be used, care should be taken to determine the source of the message file and to preserve the audio quality of the transmission of the voice recording to the sound system to the extent possible.

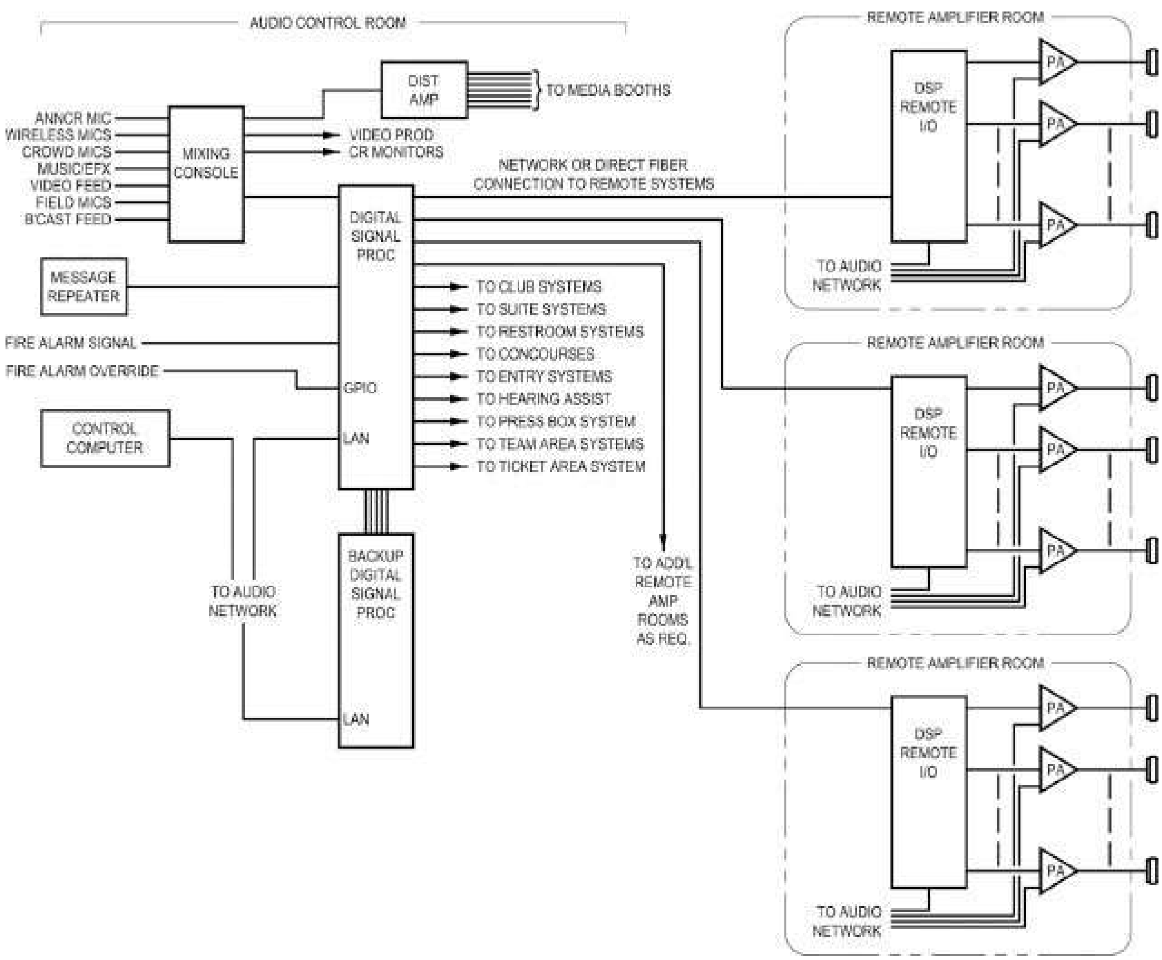


Figure 11-21. Simplified block diagram of a stadium distributed loudspeaker system.

11.4.1 Signal Transport

Given the potential for large numbers of loudspeakers and multiple individual subsystems, modern sports stadiums need a method to distribute audio signals to remote amplifier rooms and local equipment racks. While classic analog audio over shielded twisted pair cable is still available, audio designers today have several other alternatives that provide moderate to high channel density with good reliability and low latency.

One popular category of product utilizes standard Ethernet network hardware to transport dozens of channels of audio to multiple destinations. Initially popularized by CobraNet, today other multi-platform options are available such as Audinate Dante or Ravenna that offer even more capability. In addition to the multi-platform network products, there are also proprietary options such as Q-Lan offered by QSC.

Although network based audio delivery systems can provide a convenient method of moving audio signals over some distance, use of such systems does require familiarity with the specific system selected in order to deploy, setup and configure the components for optimum performance. If the intent is to use the building's data network rather than a separate, dedicated audio-only network, additional coordination will be needed to verify the building network hardware is capable of providing the needed bandwidth and performance while also delivering the other network services required for the stadium.

An alternative to Ethernet based audio delivery systems are proprietary fiber optic solutions that are purpose built for transporting large quantities of audio signals. These systems benefit from fiber's high channel count and extended transport distance without the complexity of dealing with data networks. Products such as Optocore, Riedel Rocknet and Stagetec Nexus are well suited to the requirements of stadium audio distribution.

11.4.2 Signal Processing

To accommodate the building-wide requirements for signal routing, gain adjustment, dynamics processing, equalization, crossovers and signal delay, it has become commonplace to utilize digital signal processing (DSP) platforms to provide these functions. These systems provide scalable processing capacity to match the needs of the venue, a flexible topology for system inputs and outputs, relatively easy adjustments to the internal signal flow, ability to interface third-party control systems and a centralized control computer to allow operator interaction with control screens and monitoring of system status. The ability to save multiple system configurations and settings to computer files, greatly simplifies the work required to setup a stadium for a different sport or type of show. By recalling a previously created preset, the audio system can be rapidly and reliably returned to a known configuration. Examples of products suitable for use in these types of venues include: BiAmp Tesira, BSS London, EV NetMax, Peavey MediaMatrix, and QSC Q-Sys.

11.4.3 Amplification

Modern stadiums typically use large numbers of amplifier channels to power the loudspeakers throughout the building. Although active loudspeakers with integral processing and amplification are increasing in popularity, the majority of stadiums continue to use remotely powered loudspeakers with amplifiers residing in dedicated rooms or equipment spaces. With some larger stadiums incorporating more than 800 channels of amplification, it becomes a major challenge for the AV staff to monitor amplifier health status, mute and unmute groups of channels, make precise level adjustments and power the units up or down. To aid in this management, it is recommended that amplifiers with full computer control connected through a local area network be used; allowing the operator to interact and control the amplifiers through one or more computer workstations or wireless tablets.

11.4.4 Redundancy

With the high profile nature of the sporting events held at major stadiums and because sound reinforcement systems are often utilized as part of the emergency announcement system, it is necessary to design these system to be as reliable as practically possible. A significant sound system outage could result in delaying an event or even possibly having to cancel an event if life safety is jeopardized.

In designing stadium systems, it is important to look at a variety of techniques to enhance the system's fault tolerance, such as:

- Evaluate how the failure of any major component will affect operation of the remainder of the system.
- Implement a system to automatically alert the operator when system faults occur with as much explanation of the nature of the fault as possible.
- Try to avoid situations where the failure of any single component will take down the entire system. Where it is not possible to avoid a single point of failure, consider providing a redundant unit that can easily (or automatically) be enabled or determine if it would be possible to simply bypass the failed component and restore operation.
- Consider redundant signal paths or feeds between the control room and remote amplifier rooms.
- If the system relies on a dedicated audio network, determine if a

backup network should be provided in case of a problem with the primary network.

- Connect the sound system to the building’s emergency generator and provide uninterruptible power supplies (UPS’s) for components such as control computers, DSP elements, console, outboard processing and any networking equipment.

By implementing these techniques, it should be possible to avoid many of the common failures that could plague a system during an event; however, for some venues with more demanding reliability requirements, it may be necessary to employ even more sophisticated precautions to avoid major outages.

* Common Intelligibility Standard.

Chapter 12

Surround Sound for Cinema

by Steve Barbar

12.1 Introduction

The merger of reproduced audio with film projection was transformative for the art of motion pictures. It also promoted numerous paradigm shifts for both the tools, and methods used for recording, processing, transmitting, and reproducing audio. The early development of audio for film employed many of the most notable scientists working in acoustics at the time. Bell Laboratories and RCA in the USA each had teams dedicated to the advancement of sound quality for cinema. Three channel stereo audio and rear sound effects channels—the precursor to 5.1 surround sound—was developed and demonstrated prior to WWII. The war effort, however, essentially halted the development of improved systems for the deployment of multi-channel audio and the means by which to practically distribute multi-channel media.

Post war developments—most notably magnetic tape and the television—created another paradigm shift in audio development favoring two channel stereo music reproduction. New and improved methods for audio recording, combined with stereo release media that was readily distributed, spawned the growth of the contemporary music industry. With several notable exceptions, limitations involving the practicality and cost of the multi-channel audio media or its distribution hampered further development of both high quality audio and surround sound for film. The advent of digital audio allowed for a significant improvement in quality of film sound. It also provided the means to encode more than two channels of audio onto film, and using a matrix decoder, derive something similar to the three channel stereo format developed by Bell Laboratories (Bell Labs)—with the addition of a "surround effects" channel. The introduction of an optical storage medium with sufficient capacity to digitize video at higher resolution than NTSC broadcast standards, in addition to two channels of CD quality audio, spawned another paradigm shift in home entertainment, "Home Theater". In its first iteration, a matrix decoder was employed to create left, center, right, and surround channels—and a subwoofer channel was created using the sum of left and right channel low frequency information. This in turn spawned new developments in television technology—first stereo broadcast and receivers, then a transformation of the entire broadcast system to digital signals that provide greater video resolution and improved audio quality. Enhancements to this technology (first DVD then Blu-ray) provide full bandwidth discreet multi-channel audio with resolution equal to CD quality.

Digital editing, digital post production, digital animation, digital transmission, and digital projection has created another paradigm shift in the art of motion pictures—Digital Cinema. The advent of these technologies has overcome the limitations of prior media and distribution methods that hindered the implementation of multi-channel audio. Digital cinema projection also provides enhanced resolution in the form of 4k projection, as well as the ability to release 3D movies. The current DCI standard provides for up to 16 channels with audio quality superior to CD format. On the horizon, however, is object oriented multi-channel audio. Harkening back to developments for the Bell Labs 3-channel stereo system, these systems have potential to create object localization and ambient environments that are independent from one another. Once again, audio development for film is on the verge of creating another paradigm shift for both cinema as well as home entertainment.

12.2 Origins of Sound for Cinema

One of the earliest examples of sound recording and synchronization with film was produced by William Kennedy Laurie Dickson. Dickson was working at Edison as the "official" photographer. Edison filed a preliminary patent claim in 1891 for a device that "would do for the eye what the phonograph does for the ear", and Dickson was charged with product development. In the span of five years, Dickson had created all that was required to develop the Kinetograph.[1] This included the first practical use of celluloid film that was developed by George Eastman, which Dickson decided should be 35mm—the basis for today's film standard. The Kinetograph was a motion picture camera that conveyed a strip of perforated film and used stop-expose-and go motion to take a rapid series of photographs. The developed film was installed on a cylinder that was illuminated by an Edison light source, and viewed through a window in a wooden cabinet housing the mechanism. This projection system dubbed the Kinetoscope which was first introduced at the Brooklyn Institute of Arts and Sciences on May 9, 1893. Although it was more of a peep show than projector, its development at Edison led to the next natural query—what would it take to marry sound with Kinetoscope? The answer is the "Dickson Experimental Sound Film" that was made in 1895. It consisted of Dickson playing violin into a megaphone while two Edison employees danced. The audio was recorded on Edison cylinder, and there was no means of synchronization. After leaving Edison, Dickson and other former Edison associates developed the Eidoloscope[2] projection system that was used in the first commercial movie screening on May 20, 1895.

The transition from silent motion pictures to a cinema experience that included "audio" began in the early 1900's in Paris. In 1903 French Engineer Leon Gaumont developed the "Chronophone".[3] Based on the "Gramaphone" it used a compressed air amplifier called the "Eglephone"[4] to increase loudness for audiences up to as many as four thousand. The Chronophone synchronized sound with film operating at 16FPS (frames per second) using a belt driven from a single motor that drove both the projector as two phonograph platters. The playing time of each gramophone disc was limited to approximately 200ft of film; hence the system relied on a skilled operator to change the compressed air feed from one disc to the other at the appropriate moment in order to maintain synchronicity with the visual image. Gaumont was the first to suggest that loudspeakers should be placed behind the screen. He also suggested that they should be "carried" about to enable sound to match the visual image.

The development of the moving coil loudspeaker ushered in a substantial improvement in audio quality. While Earnest Siemens first described the moving coil transducer in 1874,[5] the first practical application of moving-coil loudspeakers was developed by Peter L. Jensen and Edwin Pridham in 1925 calling the device: "Magnavox" (Latin for "great voice). Working at Bell Labs, Edward C. Wente had independently discovered the same moving coil principle, Wente and Albert L. Thuras developed the Westinghouse 555-W loudspeaker used in the Vitaphone sound system for film—the last system to use synchronized turntables. During the next year, Thuras patented the bass reflex loudspeaker,[6] and low frequency compression driver. Using these developments, Bell Labs developed the "split-range" (two way and three way) loudspeaker systems. Further developments at Bell Labs led to the "Fletcher System".

12.3 From Monaural to Stereo

Harvey Fletcher led the research team Bell Labs that included Went, Thuras, K. P. Secord, Joseph Maxwell, Henry Harrison Rogers Galt, Harold Black, Arthur Keller and others who were focused on electronic sound recording, reproduction and transmission. In 1931 Maxwell and Harrison developed a matched impedance audio recording system that incorporated a carbon microphone, tube amplifier and magnetic coil cutting head. This system was installed in the basement of the Philadelphia Academy of Music, and recordings of the Philadelphia Symphony Orchestra, conducted at that time by Leopold Stokowski, were made with his permission. The next several years brought about incremental improvements to this system—moving coil microphones, gold sputtered masters and cellulose pressings improved both signal to noise ratio as well as improved high frequency bandwidth. In 1933, Bell Labs developed a three channel stereophonic system that was used as part of a demonstration of sound transmission. Microphones captured the Philadelphia Symphony Orchestra playing in the Academy of Music in Philadelphia, PA, conducted by Alexander Smallens (the assistant conductor of the Philadelphia Orchestra) Signals from the microphones were transmitted to Constitution Hall in Washington D.C., and routed to the loudspeaker system that was concealed behind an acoustically transparent curtain. Stokowski was in Washington D.C. adjusting the system delivery to his satisfaction.

As part of the demonstration:

1. A singer walked across the stage in Philadelphia, and the stereo reproduction system matched the location in Washington.
2. A musician playing a trumpet on one side of the stage in Philadelphia was accompanied by a musician on the opposite side of the stage in Washington.
3. Finally, the orchestra performed.

To everyone's amazement, the curtain was lifted at the end of the demonstration revealing that the audience had been listening to loudspeaker reproduction and not a live orchestra.

John Hilliard was working in MGM's sound department and was familiar with the developments at Bell Labs that led to the Fletcher

System. He contacted Western Electric to engage them to build "marketable" systems for MGM based on the Fletcher System prototype. After one year passed with no progress made on the system by WE, Hilliard had a meeting with James Lansing and Dr. John Blackburn (considered the technical expert behind Lansing manufacturing). Lansing and Blackburn had recently attended a SMPTE (Society of Motion Picture and Television Engineers) meeting and were commenting on the numerous deficiencies in the quality of the audio systems used in cinema sound. This discussion and subsequent meetings formed the beginnings of the Shearer Horn Project.[7] Hilliard approached Don Shearer who was head of sound at MGM, with the idea of fabricating their own system, and Shearer approved the budget for the project. The result—the Shearer Horn—marked a significant improvement in sound quality for cinema and quickly became the industry standard, Fig. 12-1.

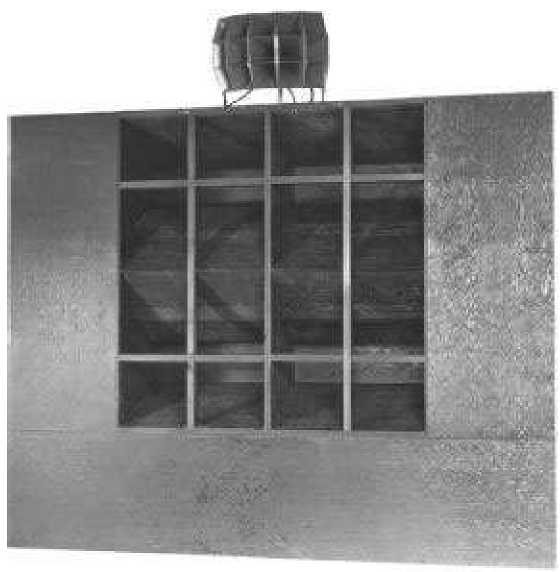


Figure 12-1. Shearer Horn. Courtesy JBL/Harman.

This advance in quality was considered so substantial that the Shearer Horn received an award in technical achievement at the

1936 Academy of Motion Picture Arts and Sciences ceremony. However, the improvement in playback fidelity unmasked high frequency noise inherent in the analog audio reproduction systems. To overcome this problem, Hilliard worked with the Academy of Motion Picture Arts and Sciences to establish a standard high frequency roll off filter that produced what became known as the “Academy Curve”.

12.4 From Stereo to Surround Sound

Walt Disney heard a performance of *The Sorcerer’s Apprentice* in the Hollywood Bowl and had the idea of using classical music with animated film that would ultimately become Fantasia.

Disney approached Stokowski, who was so taken by the idea that he offered to record the music for free, but insisted that the recording be made in stereo, and described the 1933 Bell Labs experiment to Disney. Disney envisioned taking this further, imagining *The Flight of the Bumblebee* to be localizable throughout the theater as well as across the screen. Although this was not included in *Fantasia*, the notion of “Surround Sound” was born. Disney also envisioned an immersive sound field that translated the conductor’s podium perspective to the audience. This required fundamentally different recording and playback methods than had ever been used for motion pictures. “Fantasound”[8] was developed by Disney engineers and incorporated groundbreaking concepts that are fundamental to modern audio production—including but not limited to; multi-track recording, the stereo Pan-Pot (panoramic potentiometer), overdubbing, signal routing switchers, the TOGAD (tone operated gain adjusting device that automatically increased gain during loud passages thus increasing the perceived dynamic range), and more.

There were ten variations of Fantasound—most involving incremental changes to automate dynamic level control and signal switching between loudspeakers. The first iteration used three loudspeakers behind the screen, and two loudspeakers at the rear corners of the venue. In this respect it was the forerunner of the modern 5.1 channel surround system. An important difference was that in its initial inception, Fantasound used only two audio channels—one fixed to the center channel, and one that could be panned to the other loudspeakers using a manually operated four channel pan pot.[9] This was expanded in a subsequent version to three audio channels, with the addition of three loudspeakers—two of which were located on the side walls, and one was located on the ceiling. Six way panpots that required more than one operator during screening were replaced by a pilot tone control track and Togad. The screening of *Fantasia* with Fantasound debuted as a "roadshow". The complexity of installing the equipment required the use of legitimate theaters that could accommodate the required down time necessary to install and calibrate the equipment. This limited the release to 13 major cities. The intrusion of WWII precluded the development of smaller, more portable equipment packages that would have made the "roadshow" more practical.

Post War Developments

Developments in communication systems made during WWII brought about significant changes to audio recording and reproduction systems. These included high quality permanent magnets that could be utilized in loudspeaker manufacturing, magnetic tape, and the television. Fred Waller created a new cinema format developed from film based gunner training simulators that was shot on three parallel strips of film, and presented in wide screen format. Dubbed “Cinerama”,[10] this new format incorporated seven discreet audio channels. Five channels were located behind the screen, and two channels were located at the rear corners of the theater. Cinerama was successful, but only a limited number of theaters could be equipped to show it primarily due to the nature of the very wide curved screen that was required. Based on the popularity of Cinerama, 20th Century Fox incorporated an anamorphic lens process that produced a wider screen image from one camera. Four tracks of magnetic audio were striped along the edges of a standard 35mm release print to create CinemaScope[11] in 1953, Fig. 12-2.

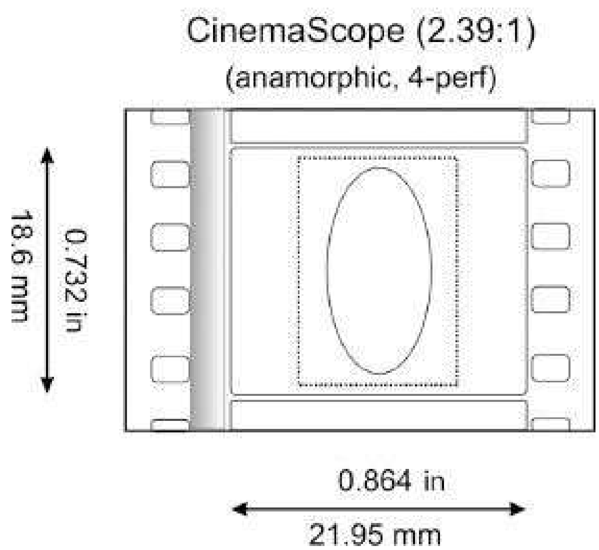


Figure 12-2. CinemaScope film showing compressed anamorphic image.

Three channels of audio were assigned to loudspeakers behind the screen, and one channel fed speakers located at the rear of the theater. Playback of magnetic audio tracks produced tape hiss, hence the Academy curve was employed to reduce high frequency noise in the primary audio channels. The effects channel used a 12kHz sine wave to switch the channel on and off, and a follow-on 12kHz notch filter removed the tone before the playback amplifier.

Michael Todd left Cinerama, and working with American Optical Company, developed 65mm cinema cameras operating at 30FPS (compared to film standard 24FPS) that shot in wide screen format. The film prints were 70mm and featured six channels of magnetically striped audio. The new format, was named for its inventors (Todd AO) and was described as “Cinerama from one hole”, i.e., a wide film format could replace the complexity of three projectors and operators required for Cinerama. Five audio channels were assigned to loudspeakers placed behind the screen, and one channel was used in the rear of the venue. Todd AO was debuted in 1955 with the release of *Oklahoma*. Like *Fantasia*, the Todd AO format was closely associated with “roadshow exhibitions” that were staged before the advent of multiplex cinemas. It was common for a Todd AO release to be screened exclusively in a large theater located in a major city with fixed seating assignments where it would run for months before it was released to neighborhood venues.

Multi-channel audio for film was hampered by the high cost of striping release prints—with each print requiring real time duplication. In the 1970’s the advent of Dolby Type A noise reduction prompted renewed interest in improving the quality of optical audio tracks printed to film. This process had not changed significantly since the 1930s, and the Academy curve had been instituted in order to reduce the noise and distortion associated with optical sound playback. In 1955 John Frayne (Westrex) proposed the means to optically record stereo audio on standard 35mm film.[12] Working at Kodak, Ronald Uhlig experimented with a 16mm print incorporating Frayne's optical encoding with Dolby A noise reduction.[13] Although this provided a significant improvement in audio quality, it was still two channel stereo that would feed loudspeakers positioned to the left and right behind the screen. It was missing the third center channel that was essential to maintaining sharp dialog localization to the center of the screen.

Dolby and Kodak collaborated in the development of a system[14] and employed a directional matrix decoder[15] designed for use in quadraphonic hi-fi systems in order to derive center channel information from the stereo signal. Dolby's implementation of the matrix used level and phase information to "steer" audio to the appropriate channel. The matrix was extended further to produce one channel of surround information. The system was called Dolby SVA, subsequently renamed Dolby Stereo, Figs. 12-3 and 12-4, and provided the same four channel audio layout as Cinemascope.

The elegance of the optical encoding afforded two key advantages. First, audio could be printed on the film, eliminating the time consuming and costly striping of magnetic audio channels. More importantly, however, was the backwards compatibility of the optical audio tracks. The same print could play in a theater with older monaural equipment as well as a venue equipped with a Dolby Stereo system, thus eliminating the requirement (and cost) for multiple prints for distribution.

Dolby Stereo enabled the next paradigm shift in theater sound. Initially it was deployed in theaters utilizing Cinemascope that had the audio infrastructure to support four channel playback. It was first used in 1976 with the release of *A Star Is Born* but within a few months the intrinsic benefits of the 4:2:4 matrix were made evident in the 1976 release of *Star Wars* followed soon after by *Close Encounters of the Third Kind*. The popularity of *Star Wars* made the addition of surround sound practical for many smaller theaters. As a result, many more patrons were exposed to the surround effects from a war in space (?), heightening expectations for an immersive cinema experience. However, the 70mm screening for *Star Wars* exposed a deficiency in headroom for magnetically recorded low frequency information. To overcome this, Gianluca Sergi and Steve Katz from Dolby remapped the six channel Todd AO audio format[16] to provide left, center, and right channels behind the screen, one surround channel, and dedicated LF channels. They demonstrated this for Gary Kurtz, (producer), who chose this remapping that was subsequently named "Baby Boom" for the 70mm *Star Wars* release. The low frequency output was combined with left center and right outputs and fed to the Todd AO loudspeaker arrays. *Close Encounters of the Third Kind* was the first release to use dedicated subwoofers[*] in theaters. This established matrix decoded surround sound as a standard for cinema. Subsequent developments included Dolby SR noise reduction that helped to both increase bandwidth as well as dynamic range. Stereo surround was introduced two years after the release of *Star Wars* with the release of *Superman*, the first time since Cinerama.

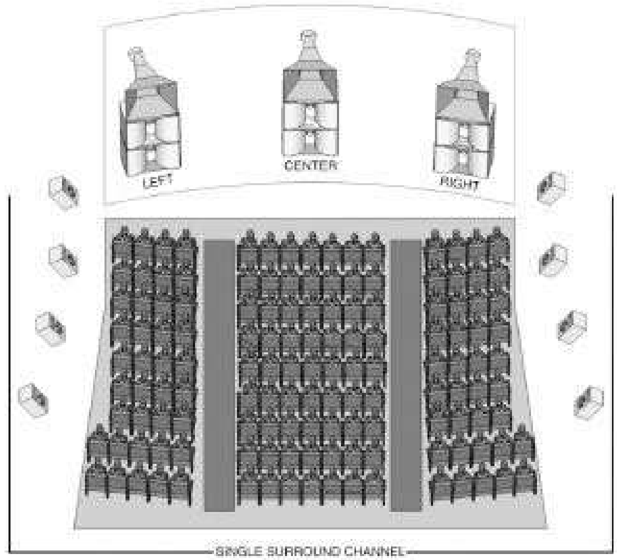


Figure 12-3. Dolby stereo loudspeaker layout.

Dolby Surround	Left	Right	Surround
Left	1	0	$-j\sqrt{\frac{1}{2}}$
Right	0	1	$j\sqrt{\frac{1}{2}}$

Dolby Pro Logic	Left	Right	Center	Surround
Left	1	0	$\sqrt{\frac{1}{2}}$	$-j\sqrt{\frac{1}{2}}$
Right	0	1	$\sqrt{\frac{1}{2}}$	$j\sqrt{\frac{1}{2}}$

Dolby Pro Logic	Left	Right	Center	Left Rear	Right Rear
Left	1	0	$\sqrt{\frac{1}{2}}$	$-j\sqrt{\frac{19}{25}}$	$-j\sqrt{\frac{6}{25}}$
Right	0	1	$\sqrt{\frac{1}{2}}$	$j\sqrt{\frac{6}{25}}$	$j\sqrt{\frac{19}{25}}$

Figure 12-4. Dolby encoding matrices.

With improvements to the 35mm format came re-invigorated interest in the 70mm film format. First Dolby A, then Dolby SR made improvements similar to those achieved for 35mm film sound. However, the 70mm format had the advantage of six discreet full bandwidth audio tracks. IMAX (ImageMAXimum)

70mm format was a introduced in 1970. It used a 70mm film format for both camera and projection, however, unlike standard 35mm projection, the film is transported horizontally using two large turntable platters. Unlike the Todd AO format, IMAX film does not contain audio, enabling the image to occupy more of the frame, Fig. 12-5.

Figure 12-5. A comparison of film size and format.

Initially IMAX used a 35mm magnetic film follower for audio that provided 6 discreet full bandwidth audio channels. In the 1990s a DTS transport replaced the 35mm follower.

12.5 The Advent of Home Theater

The development of helical scanning recording and playback technology by Ampex Corporation and is subsequent integration into VHS (Video Home System) tape machines by JVC allowed motion pictures to be viewed on television sets in the home. Initially these machines provided only monaural audio, however, the advent of stereo television broadcast promoted the ability to both record and play back stereo audio. The inherent limitations in NTSC broadcast video, as well as the size limitation of the linear magnetic audio tracks on VHS tape produced results that were substantially compromised by today's standards. However, the ability to schedule a recording of a broadcast that would otherwise be missed; and the ability to rent or own released motion pictures that could be viewed at a time of the viewers choosing in the comfort of their own home had enormous ramifications for both the motion picture and broadcast industries.

The development of the LaserDisc provided the means to encode both higher resolution video images as well as stereo audio with CD quality that was far superior to the linear analog tracks available on VHS tape. This allowed Dolby Stereo tracks from 35mm film to be incorporated with Laser Disc video releases. Hi-Fi VHS machines enabled audio to be recorded as part of the video using separate heads on the helical scanning drum. This provided sufficient dynamic range to record Dolby Stereo audio. Thus the market for a 4:2:4 decoder for the consumer marketplace was born—which was called Dolby Pro Logic. A Dolby Pro Logic Decoder provides four output channels that terminated to Left, Center, and Right loudspeakers located near the "screen" and one surround channel for loudspeakers located at the rear of a room. The first generation of the decoders was analog, and decoding accuracy suffered considerably with amplitude and phase errors. In 1988, David Griesinger (Lexicon Inc.) developed the first digital Pro-Logic decoder. This had the advantage of correcting phase and magnitude errors in the digital domain prior to decoding—which significantly improved steering accuracy. DVD format was introduced in 1997. DVD is capable of storing eight discreet full bitwidth audio channels. Video, however, was compressed. These channels are delivered as discreet analog signals from the DVD player. Blu-ray format was developed in 2000 by Sony and was commercially released in 2003. Blu-ray has larger storage capacity that allows for High Definition Video (1080p) as well as 8 discreet, full bitwidth, channels of audio. Like DVD, these audio channels are only available as analog outputs from a Blu-ray player, or from a pre-amplifier or decoder that accepts an HDMI (High Definition Multi-media Interface) signal from the Blu-ray player and provides discreet analog outputs.

12.6 The Digital Era

In 1987, a subcommittee of the Society of Motion Picture and Television Engineers (SMPTE) researched the means by which to encode digital audio on film. What emerged is a standard that became known as 5.1—the minimum number of channels that would create the desired aural experience. Three channels were assigned to loudspeakers behind the screen as Left, Center, and Right; two channels are used for surround sound Left Surround and Right Surround; and one channel is used for Low Frequency Effects (LFE), Fig. 12-6.

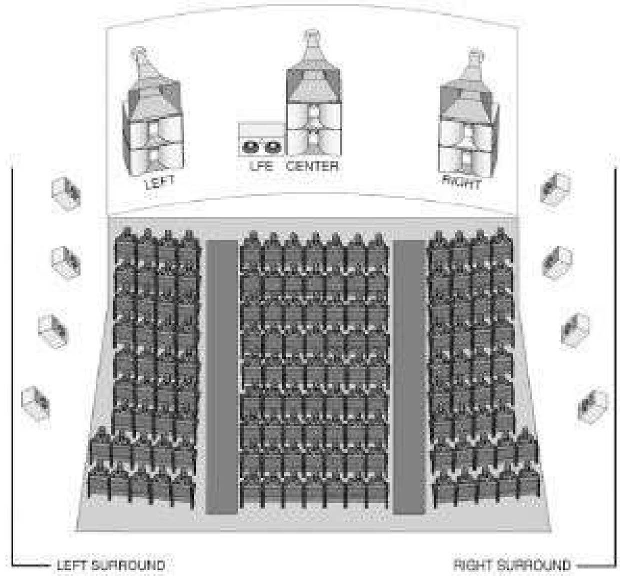


Figure 12-6. SMPTE 5.1 (ITU775) loudspeaker positions.

SMPTE describes these channels as discreet—however, at that time there was not sufficient space on either release prints, or on a single CD-ROM follower, to use conventional PCM digital audio coding. Three formats remain from the initial industry shakeout:

- Dolby Digital.
- Digital Theater Systems (DTS) (DTS can provide 8 channels).
- Sony Dynamic Digital Sound (SDDS) (SDDS has the capacity for 7.1 channels).

All of these systems incorporated some form of low bit data coding that reduced the bit count substantially compared to CD audio (which is 1,411,200 bits/s). The compromise of bit rate coding, in either print or CD follower form, made the development of digital audio for film possible. For the motion picture industry, the advantages of low bit rate coding for discreet multi-channel audio outweighed the need for higher quality two channel reproduction. The following describes the implementation of digital codecs in the approximate order of their development:

Dolby AC-3 which became known as Dolby Digital, prints digital audio information between the sprockets of a 35mm print. A data block of 78bits by 78bits is recorded between each perforation. There are four perforations per frame in a 35mm print operating at standard film rate of 24FPS. This yields 584064bits/s before synchronization and error correction, and the actual data rate is 320,000bits/s. The first release using Dolby Digital was *Batman Returns* in 1992.

DTS uses a CD follower system that synchronizes to the film. Instead of printing the digital audio on the film, a DTS time code track that includes the film title is printed on the film. The DTS decoder reads this time code and acts as an audio transport. DTS has a fixed 4:1 compression ration—this yields a data rate of 880,000 bits/s. The first film release using DTS was *Jurassic Park* in 1993.

Sony SDDS initially provided 8 channels of uncompressed audio placed on several portions of the print. In its final form, SDDS incorporates Adaptive Transform Acoustic Coding (ATRAC) that yields 220,000bits/s for eight channels, Fig. 12-7. The first film release with SDDS was *The Last Action Hero* in 1993.

Dolby Digital EX was developed with Lucasfilm THX and utilizes a matrix decoder to create a center rear channel from left surround and right surround channels. Like the Dolby Pro Logic process, steering is accomplished using amplitude and phase information, hence it is backwards compatible with Dolby Digital. Its first use was the release of *Star Wars Episode I: The Phantom Menace* in

1999.

DTS ES (Enhanced Surround) Matrix uses flags contained in the surround data to derive a center rear channel from left surround and right surround channels

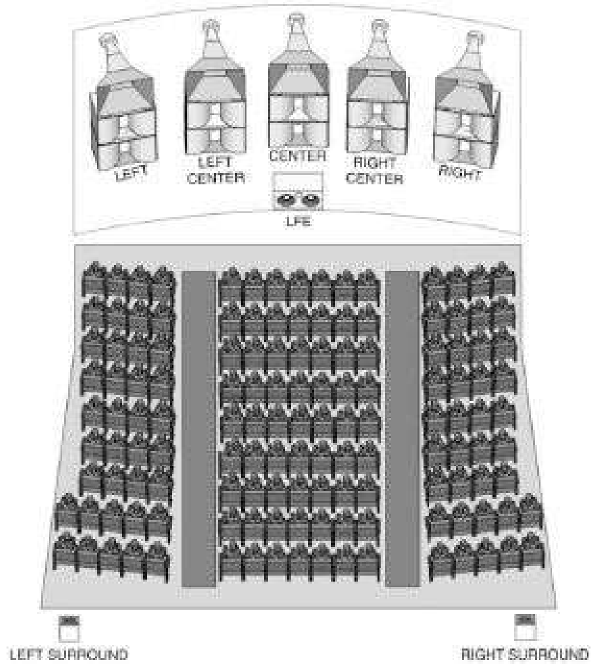


Figure 12-7. Sony SDDS loudspeaker positions.

DTS ES Discreet provides seven (6.1) independent channels with rear center channel information recorded discreetly. In order to maintain backwards compatibility with DTS 5.1, the discreet rear center channel information is matrixed to left surround and right surround channels. The DTS ES Discreet decoder, Fig. 12-8, removes this audio. Its first use was the release of *The Haunting* in 2000.

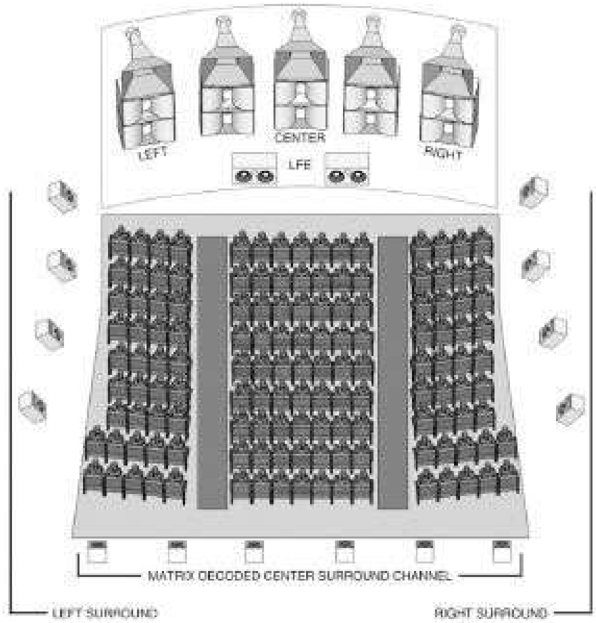


Figure 12-8. Dolby EX, DTS ES 6.1 loudspeaker positions.

12.7 Motion Pictures Go Digital

The first Digital Light Pipe (DLP) projection system was demonstrated in 1998 with the screening of *The Last Broadcast* in various cities throughout the USA.[17] This was the first film to be shot, edited, distributed, and screened digitally using technology developed by Texas Instruments. In 2000, SMPTE formed a committee (DC28) to explore standards for Digital Cinema. In 2002 Digital Cinema Initiatives was formed as a joint project involving several major motion picture studios (Disney, Paramount Pictures, Fox, MGM, Sony Pictures Entertainment, Universal Studios and Warner Brothers) to set standards for Digital Cinema. The DCI standard includes specifications for the Digital Cinema Package (the files that generate digital video and audio) that is "ingested" into the server that feeds the digital projection system and audio decoder, Fig. 12-9.

DCI specifies sixteen channels of full bandwidth digital audio operating at either 48 kHz or 96kHz sampling rate with 24 bits resolution. The audio file format is .wav per ITU Tech 3285 version 1 and is uncompressed. The “digital cinema package” can be transmitted directly to the server. This fundamentally changes the nature of film distribution. Once stored on the server, the encrypted files are processed by the Media Block which provides real time decryption and conversion into playback media that includes:

- Main Image.
- Sub Picture.
- Timed Text.
- Audio.

The Media Block provides a digital audio link that conforms to AES-3 protocol and delivers sixteen channels with 24 bit resolution operating at either 48kHz or 96kHz sampling rate to the Cinema Processor, which provides channel mapping and conversion from digital audio to analog audio (minimum channel mapping is 5.1), Figs. 12-10 thru 12-13.

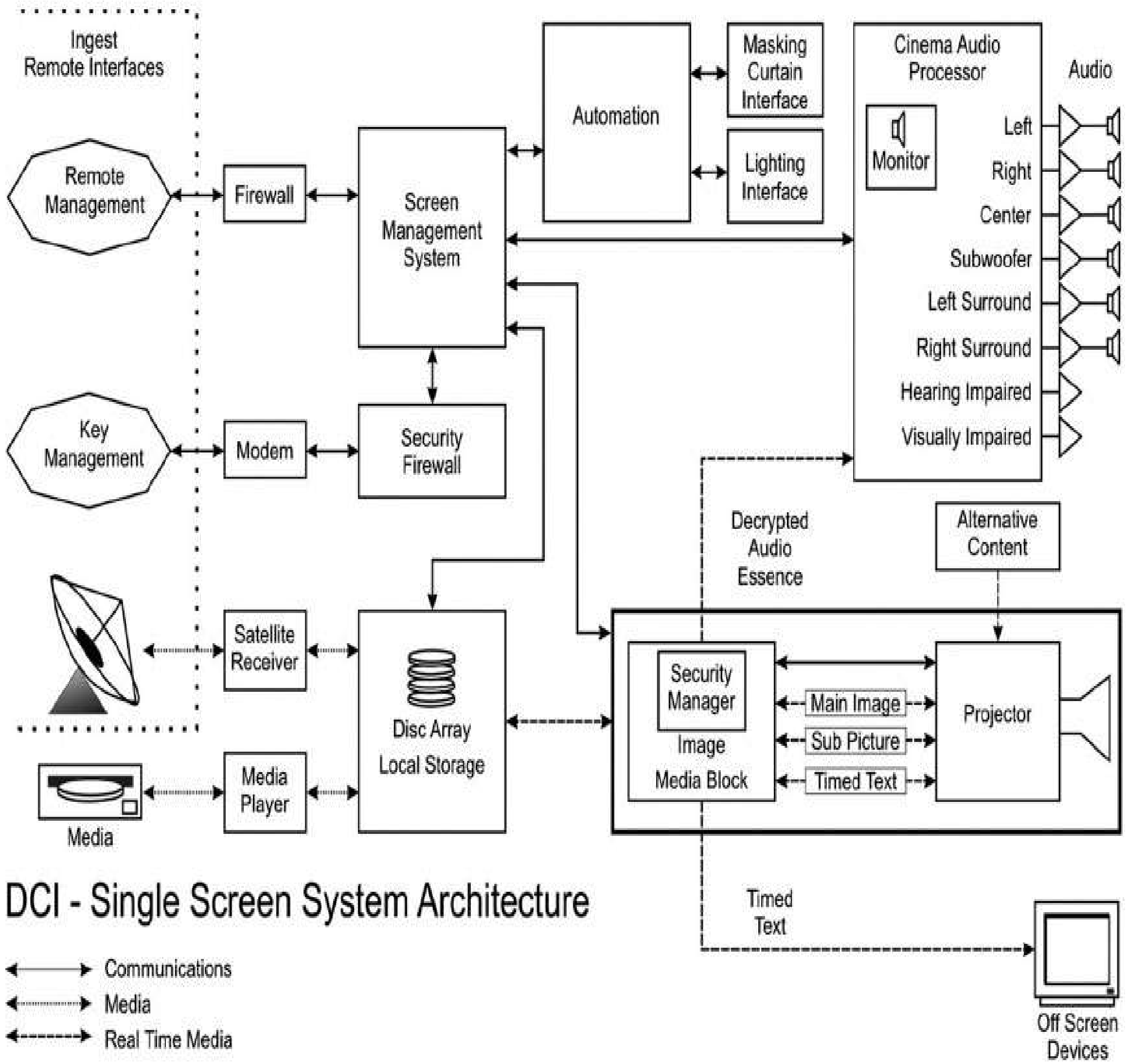


Figure 12-9. DCI single screen system architecture.

AES Pair#-Ch#	Channel #	Label/Name	Description
1-1	1	L - Left	Far Left Screen Loudspeaker
1-2	2	R - Right	Far Right Screen Loudspeaker
2-1	3	C - Center	Center Screen Loudspeaker
2-2	4	LFE-Screen	Screen Low Frequency Effects Subwoofer(s)
3-1	5	Ls - Left Surround	Left Wall Surround Loudspeakers
3-2	6	Rs - Right Surround	Right Wall Surround Loudspeakers
4-1	7	Lc - Left Center	Mid Left to Center Screen Loudspeaker
4-2	8	Rc - Right Center	Mid Right to Center Screen Loudspeaker
5-1	9	Cs - Center Surround	Rear Wall Mounted Surround Speakers
5-2	10		SMPTE Reserved
6-1	11		SMPTE Reserved
6-2	12		SMPTE Reserved
7-1	13		SMPTE Reserved
7-2	14		SMPTE Reserved
8-1	15		User Defined
8-2	16		User Defined

Figure 12-10. DCI nine channel audio assignments.

AES Pair#-Ch#	Channel #	Label/Name	Description
1-1	1	L - Left	Far Left Screen Loudspeaker
1-2	2	R - Right	Far Right Screen Loudspeaker
2-1	3	C - Center	Center Screen Loudspeaker
2-2	4	LFE-Screen	Screen Low Frequency Effects Subwoofer(s)
3-1	5	Ls - Left Surround	Left Wall Surround Loudspeakers
3-2	6	Rs - Right Surround	Right Wall Surround Loudspeakers
4-1	7	Lc - Left Center	Mid Left to Center Screen Loudspeaker
4-2	8	Rc - Right Center	Mid Right to Center Screen Loudspeaker
5-1	9		Unused
5-2	10		SMPTE Reserved
6-1	11		SMPTE Reserved
6-2	12		SMPTE Reserved
7-1	13		SMPTE Reserved
7-2	14		SMPTE Reserved
8-1	15		User Defined
8-2	16		User Defined

Figure 12-11. DCI eight channel audio assignments.

AES Pair#-Ch#	Channel #	Label/Name	Description
1-1	1	L - Left	Far Left Screen Loudspeaker
1-2	2	R - Right	Far Right Screen Loudspeaker
2-1	3	C - Center	Center Screen Loudspeaker
2-2	4	LFE-Screen	Screen Low Frequency Effects Subwoofer(s)
3-1	5	Ls - Left Surround	Left Wall Surround Loudspeakers
3-2	6	Rs - Right Surround	Right Wall Surround Loudspeakers
4-1	7		Unused
4-2	8		Unused
5-1	9	Cs - Center Surround	Rear Wall Surround Loudspeakers
5-2	10		SMPTE Reserved
6-1	11		SMPTE Reserved
6-2	12		SMPTE Reserved
7-1	13		SMPTE Reserved
7-2	14		SMPTE Reserved
8-1	15		User Defined
8-2	16		User Defined

Figure 12-12. DCI seven channel audio assignments.

AES Pair#-Ch#	Channel #	Label/Name	Description
1-1	1	L - Left	Far Left Screen Loudspeaker
1-2	2	R - Right	Far Right Screen Loudspeaker
2-1	3	C - Center	Center Screen Loudspeaker
2-2	4	LFE-Screen	Screen Low Frequency Effects Subwoofer(s)
3-1	5	Ls - Left Surround	Left Wall Surround Loudspeakers
3-2	6	Rs - Right Surround	Right Wall Surround Loudspeakers
4-1	7		Unused
4-2	8		Unused
5-1	9		Unused
5-2	10		SMPTE Reserved
6-1	11		SMPTE Reserved
6-2	12		SMPTE Reserved
7-1	13		SMPTE Reserved
7-2	14		SMPTE Reserved
8-1	15		User Defined
8-2	16		User Defined

Figure 12-13. DCI six channel audio assignments.

From Surround Sound to Immersive Audio

The availability of sixteen uncompressed audio channels in the digital cinema package has allowed for the expansion of the 5.1 cinema format, prompting the development of immersive audio for cinema. Instead of utilizing a matrix decode process to derive a center rear surround channel, discreet channels provide the ability to create surround "zones" and enabled precise panning between the zones. Dolby Surround 7.1, Fig. 12-14, provides eight discreet audio channels with four surround zones. Its first use was the release of *Toy Story III* in 2010.

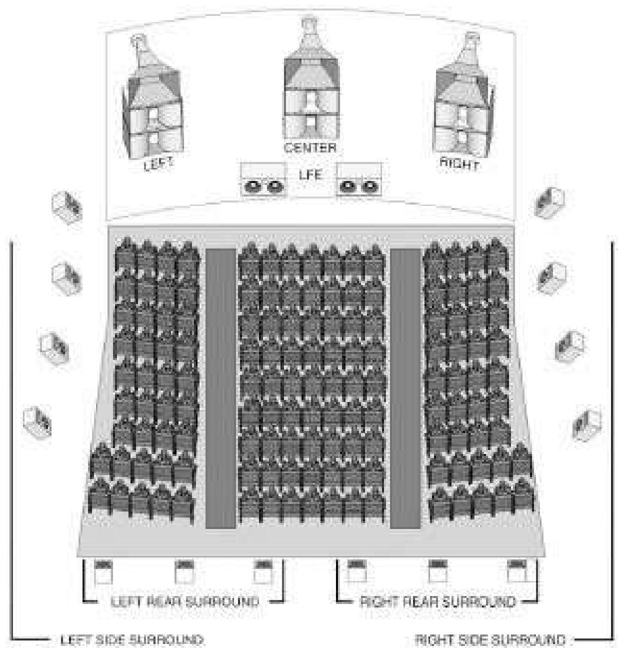


Figure 12-14. Dolby Surround 7.1 Loudspeaker Positions.

Auro was originally developed in 2005 by Wilfried Van Baelen of Galaxy Studios in Belgium. It is distributed exclusively by Barco, a leading manufacturer of digital cinema projectors. At minimum, Auro adds "height" loudspeakers to the left, left surround, right, and right surround locations forming a "dual quad" arrangement.[18] for a 9.1 channel configuration. The format expands to 13.1 adding a center channel and rear center channel height loudspeakers, and expends again to 18.1 with ceiling mounted loudspeakers, Fig. 12-15. While Auro complies with DCI specifications for 24 bit audio data, the Auro codec remaps the PCM bitstream to include additional audio channels. The primary 5.1 channels retain 20 bit resolution, and no codec is required for 5.1 channel playback. When an Auro codec is used, the additional channels are “unfolded” from the least significant bits of the PCM bitstream. Thus, Auro can provide multiple distribution releases in one standard DCI compliant 6 channel PCM master. The encode process for Auro uses “spatial panning” (x = width, y = depth, z = height), with the added “height” creating a “dimensional sound in space”. Its first use was the release of *Red Tails* in 2012.

In 2012, Dolby introduced “Atmos”, a multi-channel immersive audio system. The Atmos system utilizes the Dolby Surround 7.1 loudspeaker layout and adds front side surround loudspeakers, left and right surround subwoofers, optional left center and right center loudspeakers, and two rows of loudspeakers overhead. Atmos is much more integral in nature than previous codecs used for surround sound. It consists of a series of authoring software applications and plug-ins for the Avid Pro-Tools system. An Atmos mix consists of “Bed Audio” (channel based pre-mixes and stems that include specific multi-channel panning); “Object Audio” (mono or stereo soundtrack content that includes panning via Dolby Atmos panning software); and “Dolby Atmos Metadata” (the panning information for “objects”).

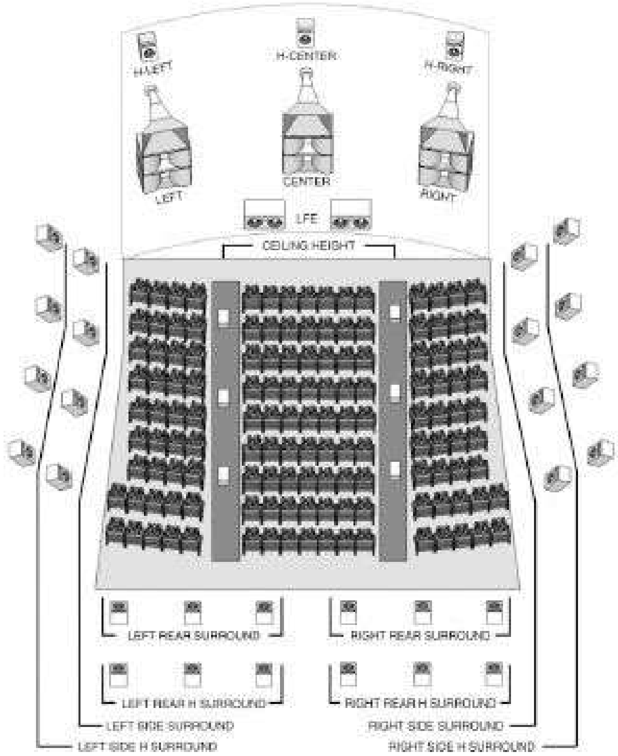


Figure 12-15. Auro 13.1 loudspeaker positions.

Dolby refers to “Beds” as channel based submixes and stems (like reverberation and ambient effects) that are fed to arrays of loudspeakers, citing the surround zones created in typical 5.1 and Dolby Surround 7.1 mixes. Two additional channels provide stereo overhead outputs. Objects are individual sound elements that can be static, or can be panned to any single channel and loudspeaker, or any grouping of loudspeakers. This is achieved using the Dolby Panner application that generates metadata which specifies the location of the object as rendered in 3D space.

Using an HD MADI interface, audio is transferred from Pro Tools to the Dolby Rendering and Mastering Unit (RMU) where it is combined with the metadata. In the Pro Tools system, 9.1 channels represent the bed, which is mapped to Dolby Surround 7.1 standard. Atmos objects occupy channels 11–128, Fig. 12-16.

Channel #	Label/Name
1	L - Left
2	R - Right
3	C - Center
4	LFE-Screen
5	Lss - Left Side Surround
6	Rss - Right Side Surround
7	Lsr - Left Surround Rear
8	Rsr - Right Surround Rear
9	Lts - Left Top Surround
10	Rts - Right Top Surround
11-128	C for Mono Object Tracks S for Stereo Object Tracks

Figure 12-16. Dolby Atmos channel mapping.

The Dolby RMU renders audio to loudspeakers located in the dubbing theater where "stems" are created. Each stem incorporates a channel based bed and several audio objects (along with their metadata). Stems are the means of creating groups (example, dialog, music, Foley, backgrounds, etc.). The RMU enables the engineers to hear how the beds and objects combine to create the final mix. The RMU also allows the engineers to hear 7.1 and 5.1 versions of the mix.

The final mastering session creates a Dolby Atmos package file, Fig. 12-17, as well as 7.1 and 5.1 renderings of the print master. These files are imported into the digital cinema packaging (DCP) along with the image file and subtitle files. Once in the DCP, the Main Audio MXF file is encrypted (with appropriate additional tracks appended) per SMPTE specifications. The Atmos MXF file is packaged as an auxiliary track file, and is optionally encrypted using a symmetric content key per SMPTE specifications. The single DCP can be distributed per DCI specifications, and ingested by any DCI compliant server. Any venues that are not equipped with Dolby Atmos will ignore the auxiliary file containing the Dolby Atmos sound track and use the Main Audio File for standard playback. A Venue equipped with the Atmos Cinema Processor will ingest the Auxiliary file, and render Atmos soundtrack information in accordance with the loudspeaker configuration stored in the unit. The Atmos B-chain audio processor provides up to 64 output channels that are configured by the Atmos Cinema Processor, Fig. 12-18. Its first use was the release of *Brave* in 2012.

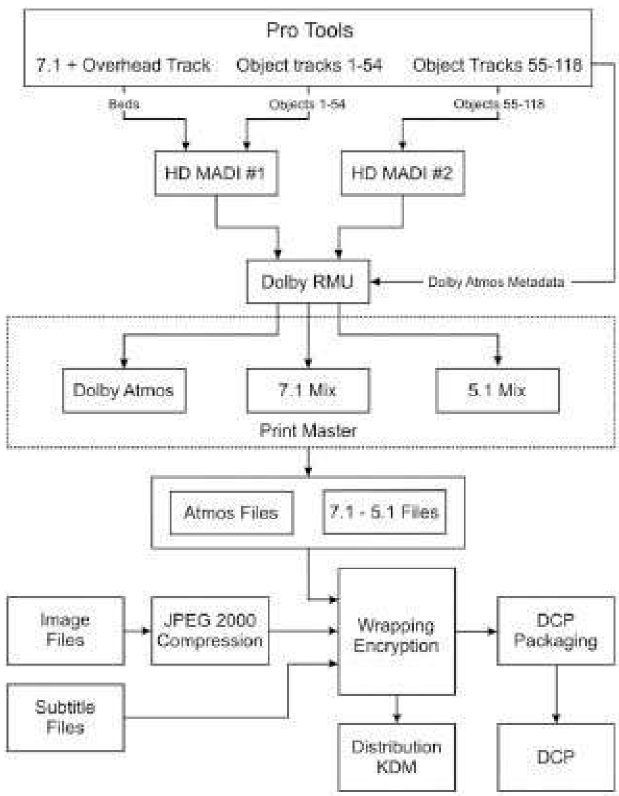


Figure 12-17. Dolby Atmos encoding.

The Multi-Dimensional Audio (MDA) group (consisting of DTS, Doremi Laboratories, Ultra-Stereo Laboratories, QSC, Barco and Auro Technologies) has also developed an immersive audio system. MDA is object based, allowing engineers to position audio in three dimensional space. MDA utilizes a plug in for digital audio workstations, and Avid's D-console. Each object, or group of objects, is assigned its own identity, allowing them to be addressed independently during the re-recording process. MDA uses standard PCM audio files that are uncompressed and unencrypted to record and deliver soundtracks. MDA creator generates three dimensional panning metadata for each PCM channel, and creates the Master Mix as an MDA interoperable file. This single mix is wrapped into a DCP along with an image file and subtitle file. Once ingested at the media block, the MDA file is unwrapped and delivered to the MDA Cinema processor. The MDA soundtrack is rendered in three dimensions in real time using panning metadata for each audio object. This allows the single mix to be scalable to both immersive sound formats as well as standard 7.1 and 5.1 formats.

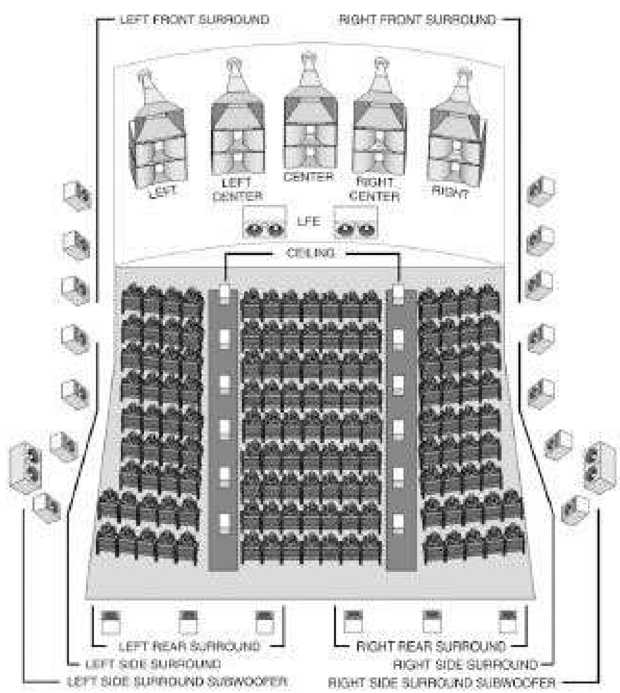


Figure 12-18. Dolby Atmos loudspeaker positions.

Currently there are upwards of 1000 theaters that have installed immersive surround sound systems. The popularity and growing interest in this sonic experience has prompted DCI to persuade

SMPTE to assist in the research and development of an open format, object based standard for immersive audio. SMPTE has formed a technical committee (25CSS) that has released its first report detailing progress toward this standard.[19] The greatest differentiation between standard and immersive sound technologies is that in standard systems, sound localization was processed in the A-chain. Immersive sound systems that are object based employ a rendering engine in the B-chain that uses metadata to create audio localization. The DCI specification that set the standard for digital cinema requires PCM audio coding that conforms to AES/EBU protocols. The Dolby Atmos system utilizes a lossless compression codec for the Atmos soundtrack. In general, however the committee finds more commonality than discord among competing immersive audio system loudspeaker placements for all but ceiling loudspeakers.

There is little doubt that immersive audio technology represents the next paradigm shift in both cinema as well as home audio systems for motion picture entertainment. The means by which this is encoded, transmitted, delivered and ingested will have ramifications throughout the communication industries on a world wide scale.

12.8 Monitoring Surround Sound

As you might expect, differing surround sound formats have unique requirements for loudspeaker placements. SMPTE, ITU, AES, and others have developed standards for these placements as follows:

Left, Center, Right, Surround (LCRS) is the prescribed configuration for Dolby AC-3 and Dolby Pro Logic Surround. The surround channel is monophonic, and has reduced high frequency content, Fig. 12-19.

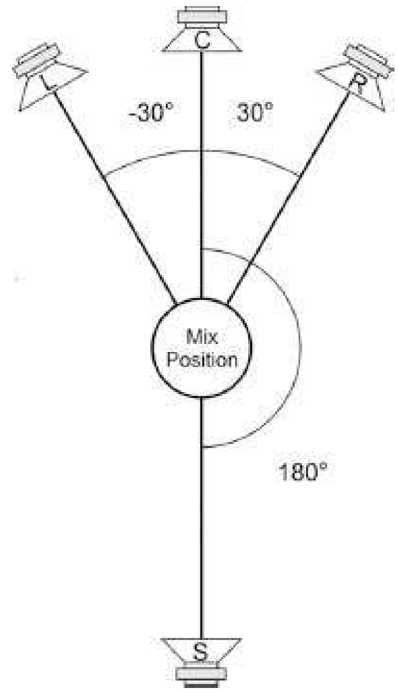


Figure 12-19. L, C, R, S – AC-3 matrix decoding.

5.1 conforms to ITU-775 as well as several other surround format standards. This is the most common surround format. The 5.1 channels are left, center, right, left surround (left rear), right surround (right rear), and LFE, Fig. 12-20.

6.1 is used for Dolby-EX or DTS-EX, Fig. 12-21. The 6.1 channels are left, center, right, left surround, surround, right surround, rear center surround and LFE. Dolby EX matrix decodes the rear center surround channel from the left surround and right surround channels. DTS ES matrix decodes the center rear channel using metadata encoded in the PCM bitstream. DTS ES discreet provides a discreet rear center surround channel, and uses metadata to alter the left surround and right surround channel as required when the center channel is active.

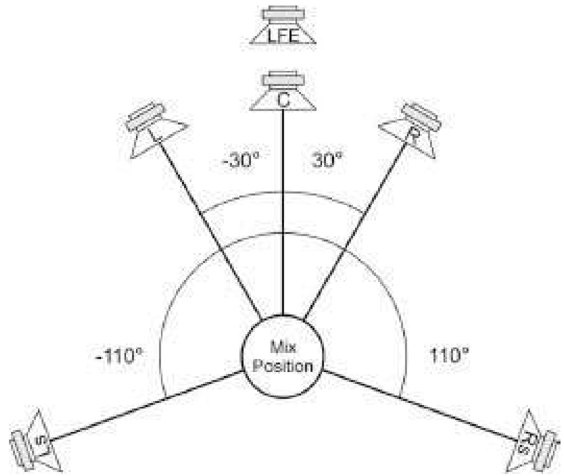


Figure 12-20. 5.1 (ITU-775).

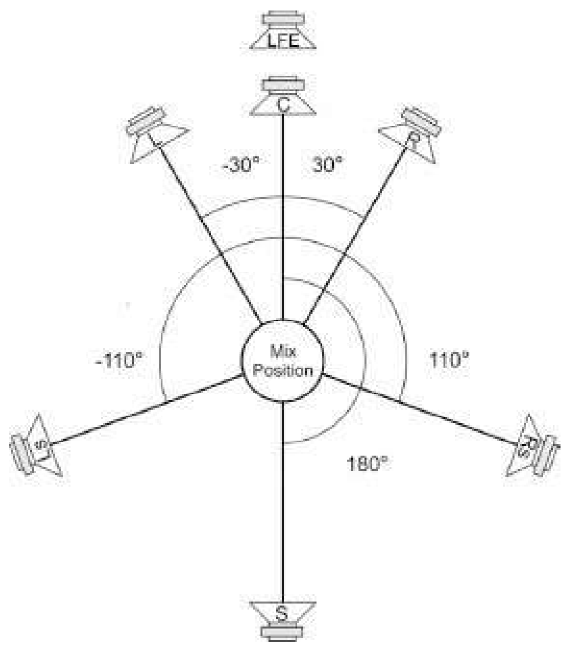


Figure 12-21. Dolby EX, DTS ES.

7.1 (3/4.1) uses the same loudspeaker configuration as 5.1, but adds two side channels, Ls (left side surround) and Rs (right side surround), placed at 70° and -70° from center. This conforms to Dolby Digital Surround format, Fig. 12-22.

7.1 SDDS or 7.1 Sony Dynamic Digital Sound, Fig. 12-23, adds two loudspeakers to 5.1, Lc (left center) and Rc (right center), and uses left surround and right surround loudspeakers positioned in the rear corners of the venue. This is designed for use in with Sony decoding and playback hardware.

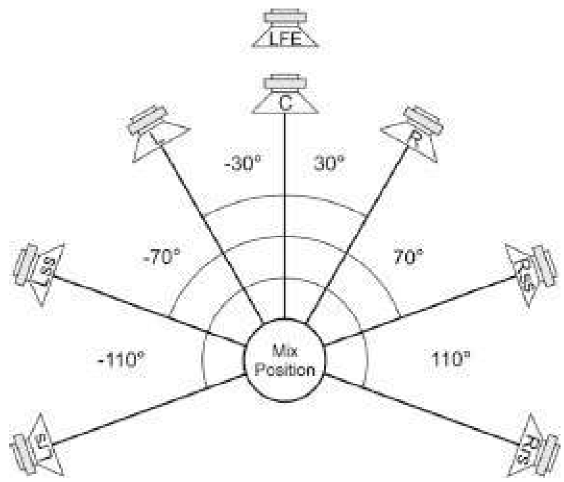


Figure 12-22. 7.1 (3/4.1).

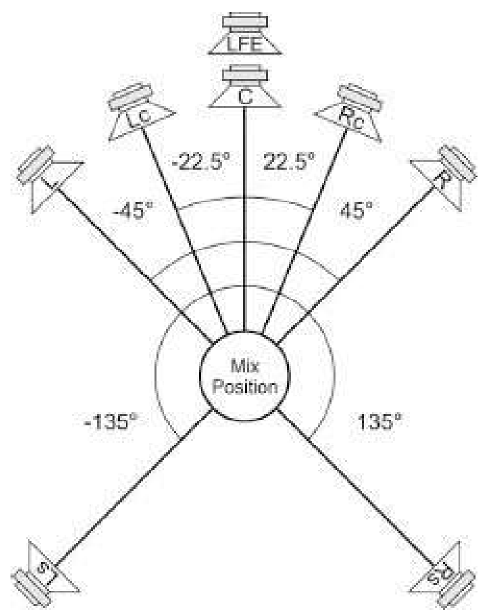


Figure 12-23. 7.1 SDDS.

12.9 Theater Audio System Calibration

The SMPTE Standard ST-202 prescribes the method for calibrating the B-chain equipment, adjusting the audio system in accordance with existing room acoustics, and adjusting the level of both individual components as well as the complete system. The B-chain signal flow is described in Fig. 12-24.

The stimulus specified is a wide band pink noise signal, preferably inserted at the input source selector prior to the system equalizer. SMPTE defines the referenced electrical level as the average measured voltage of a wide band pink noise signal with a bandpass filter of 22Hz to 22kHz using a recorded reference level and main “fader” set to nominal. The recording reference level should measure 20dB below 100% or full modulation. The absolute sound pressure level is the spatial average of a single channel of the system (L, C, or R) measured with broadband pink noise using the recording reference level as a stimulus.

Measurements should be made with a wide-band sound level meter set to C weighting and slow response. SMPTE also advocates for the use of a 1/3-octave band spectrum analyzer.

All sound pressure measurements should be made within the seating areas as shown in Fig. 12-25.

Sound pressure level should be measured in at least one position for each screen and surround loudspeaker, and the measurements for each channel should be spatially averaged. If only a single location can be selected, it should be position “S”. The subwoofer should be measured in at least four positions, with each measurement averaged over a time interval of a minimum of 30s.

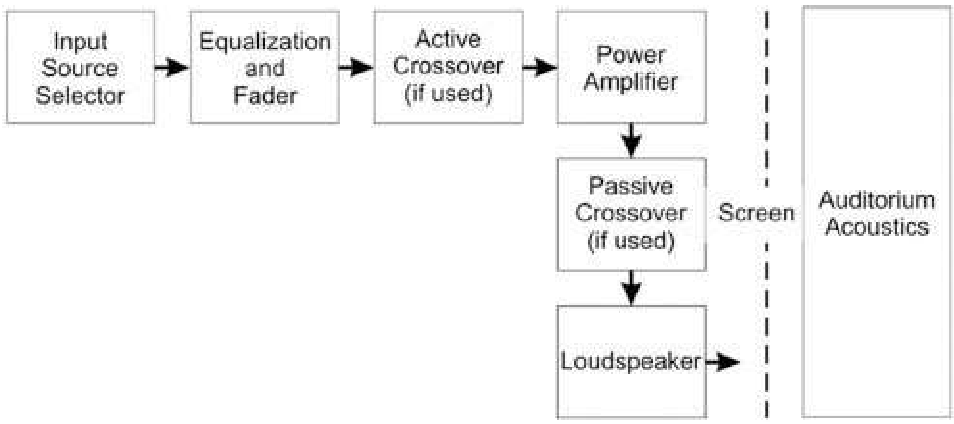


Figure 12-24. B-chain signal flow.

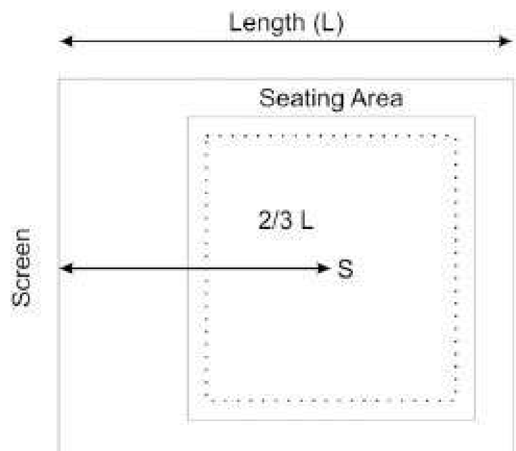


Figure 12-25. Recommended measurement locations.

12.9.1 Screen Channels

The relative sound pressure level of the screen channels should be adjusted to be within plus or minus 0.5dB of the absolute sound pressure level.

12.9.2 Surround Channels

If there is one surround channel, the sum of the loudspeakers should be adjusted to equal the absolute sound pressure level. When multiple independent surround channels are employed, their combined output should be equal to the absolute reference level when they are fed the same in-phase test signal.

12.9.3 Low Frequency

For discreet digital photographic soundtracks, or digital cinema soundtracks, the LFE channel should be adjusted to indicate 10dB of in-band gain. (using a real time spectrum analyzer is the only accurate means of measuring in-band gain) Thus, the subwoofer will have the same level in its passband as the level in the passband of a screen channel, Fig. 12-26.

When the soundtrack playback is analog photographic with bass extension processing, the bass extension subwoofer channel should show no in band gain when viewed on a real time spectrum analyzer, Fig. 12-27.

It is often difficult to determine the optimum polarity of a subwoofer by viewing cone excursion alone. Since the interaction of the subwoofer and main screen channels is of greatest interest, it is best practice to evaluate the center channel and subwoofer channels simultaneously using a real time spectrum analyzer to determine which polarity provides the best result.

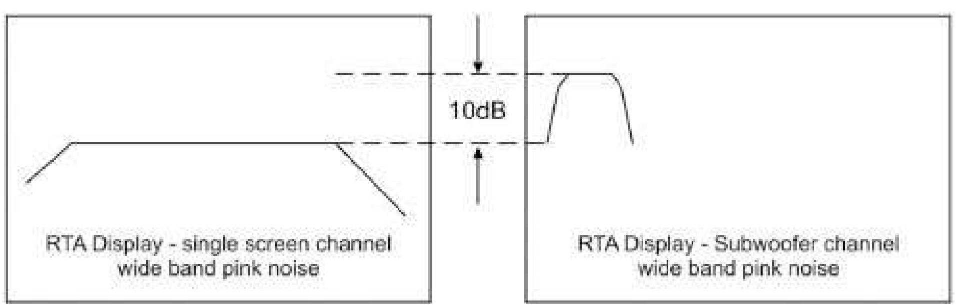


Figure 12-26. Measurement of subwoofer sound pressure level with digital LFE soundtrack using a RTA.

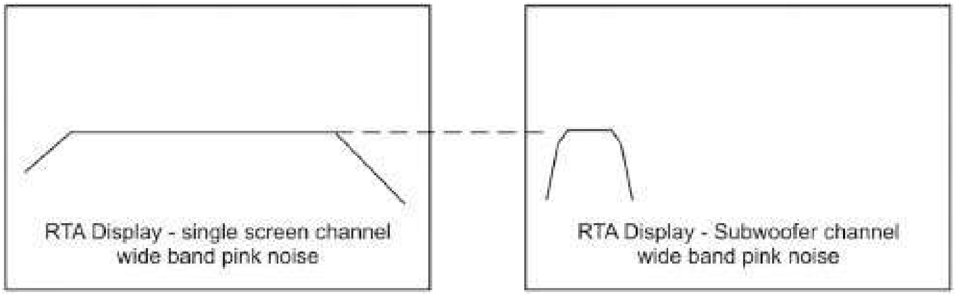


Figure 12-27. Measurement of subwoofer sound pressure level with analog photographic soundtrack with bass extension playback processing using a RTA.

12.9.4 Reference Level

The recommended reference level should be 85dBC for normal theater operation.

12.9.5 Immersive Audio Systems B-Chain Calibration

The ability to direct multiple sound sources to specific independent loudspeakers that comprise a multi-dimensional surround sound system will necessarily alter both the makeup of the components in such systems, as well as the methods used for system calibration. Each loudspeaker will potentially be required to deliver greater sound power and broader spectral content into the venue. Additionally, the sum power of specific loudspeakers can be used to form surround zones typically associated with 5.1 and 7.1 surround mapping. However, depending on the capabilities of the system utilized, these "zones" can be approximated in 3D space. Differing systems may have requirements for loudspeaker directivity that may conflict with one another.

With the mixing engineer having individual access to both independent loudspeaker channels as well as smaller groups of arrays, a calibration strategy specific to the system utilized becomes paramount to ensuring that artistic intent is preserved. Such systems will no doubt require greater attention to calibration on a channel by channel basis; including the means by which a given channel is equalized, and delayed with respect to both other independent channels as well as groups of channels that may comprise an array. Thus, while there is likely to be sufficient commonality among systems to devise the unified coding and distribution standard for immersive audio that DCI seeks to establish; it is also likely that each system will have its own preferred optimization and calibration strategy. For example, Dolby has provided specifications for the Atmos system that include loudspeaker power handling and sensitivity, as well as loudspeaker counts and placements as a function of seating area based on room volume.[20]

The development of an immersive audio standard is the first important step in translating the cinema audio distribution methodology to both a broadcasting standard, as well as the development of future media.

References

1. National Archives Identifier 561115.
2. USPTO US707934 Woodville Latham, Projecting Kinetoscope.
3. Richard Abel, “Encyclopedia of Early Cinema” ISBN 0-415-23440-9.
4. USPTO No 677.

5. German Patent 2355 E.W. Siemens – filed Dec. 14, 1877.

6. USPTO No 1,869,178 A. L. Thuras – Sound translating Device.

7. The Lansing Heritage Archive.

8. WM. E. Garity and J.N.A Hawkins (1941) Fantasound *Journal of the Society of Motion Picture and Television Engineers*.

9. WM. E. Garity and J.N.A Hawkins (1941) Fantasound *Journal of the Society of Motion Picture and Television Engineers*.

10. John W. Boyle ASC "And Now… Cinerama" (1952) *American Cinematographer*.

11. "Cinemascope – What It Is and How It Works" (1953) *American Cinematographer*.

12. J. G. Frayne, "A Compatible Stereophonic Sound System" (1955) *Journal of the Society of Motion Picture and Television Engineers*.

13. Ronald E. Uhlig, "Stereophonic Photographic Sound Tracks" (1973) Vol 82 *Journal of the Society of Motion Picture and Television Engineers*.

14. I. Allen, "The Production of Wide-Range, Low-Distortion Optical Sound Tracks Utilizing the Dolby Noise Reduction System," SMPTE Journal, Vol. 85, p. 720 (September, 1975).

15. Peter Scheiber, "Toward a More Accurate Spatial Environment". (December 1969). *Journal of the Audio Engineering Society* (AES) 17 (6): 690, 691.

16. Gianluca Sergi, "The Dolby Era: Film Sound in Contemporary Hollywood (2004) ISBN 071907066X.

17. Business Wire (Sept. 22, 1998).

18. Wilfried Van Baelen, Tom Bert, Brian Claypool, Tim Sinnaeve, “A new dimension in cinema sound” Auro White Paper.

19. Report of the Study Group on Immersive Audio Systems Cinema B-Chain and Distribution.

20. Dolby Atmos Specifications, Issue 2.

* ‘Technically, the film *Earthquake* was the first to use subwoofers—however, a pseudo-random noise signal was generated to feed the subwoofers, which was triggered by a control signal that was printed on the film.

Chapter 13

Acoustical Modeling and Auralization

by Dominique J. Chéenne, Ph.D.

13.1 Introduction

As it is often the case with branches of engineering dealing with the understanding and the prediction of complex physical phenomena, *modeling* has rapidly become an integral part of the acoustical design process. When dealing with indoor sound propagation the use of a suitable model may allow the designer to assess the consequences that a change in parameters such as room shape, material selection, or source placement will have on variables like sound pressure, reverberation time, or reflection ratios at specific points inside the room. Acoustical models can also be developed for outdoor sound studies as one may inquire what shape and height a highway barrier needs to be in order to attenuate a specific level of unwanted noise from highway traffic over a given area. In these instances, the model is expected to provide the designer with the answer to the fundamental "what if?" question that is at the genesis of an engineered design—i.e., one that leaves little to chance in terms of the sensible selection of its parameters in order to achieve a specific result. The answers to the question allow the designer to assess the performance or the cost-effectiveness of a design, based on a specific set of criteria prior to committing to it.

Although an experienced designer may be able to achieve a substantial understanding of the acoustics of a given environment simply by looking at the data that results from the modeling phase of the design, the acoustical model may also be used to provide an auditory picture of the data so that a qualitative evaluation of the acoustics can be performed by trained and/or untrained listeners. This phase of the design, known as *auralization*, aims at doing for one's ears what a picture does for one's eyes: present a description of an environment in a way that is best suited to the most appropriate sensor. In this instance, the basic goal is to use sound to demonstrate what a specific environment may sound like, just like the fundamental purpose of an image is to illustrate what an environment may look like. The challenges associated with virtual representation that exist in the visual world, such as accuracy, context, and perception, are also present in the aural environment, and the old engineering school adage that "a picture is worth a thousand words, but only if it is a good picture" rings also true in the world of acoustical modeling and of auralization.

The aim of this chapter is to provide the reader with a basic understanding of the various methodologies that are used in acoustical modeling and auralization, and the emphasis is on models that can be used for the evaluation of room acoustics. The reader is referred to the bibliography for further in-depth reading on the topic.

13.2 Acoustical Modeling

This section will review various acoustical modeling techniques from the perspective of theory, implementation, and usage.[*] The categorization and grouping of the modeling techniques into three general families (physical models, computational models, and empirical models) and in further subgroups as presented in Fig. 13-1 is provided as a means to identify the specific issues associated with each technique in a fashion that is deemed effective by the author from the standpoint of clarity. A brief mention of hybrid models that combine various techniques will also be introduced. The sections of this chapter are written as independently from each other as possible so the reader can go to topics of specific interest without reading the preceding sections.

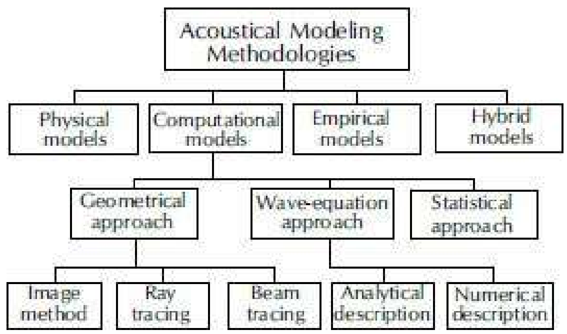


Figure 13-1. General classification of acoustical modeling methodologies.

13.2.1 Physical Models

This class of model uses a scaling approach to yield a 3D representation of typically large acoustical environments like theater or auditoria for the purpose of evaluation and testing. The models are constructed at geometrical scales ranging from 1:8 to 1:40, and an example of a 1:20 physical model is presented in Figs. 13-2 and 13-3. Physical modeling techniques became popular after WWII and reached their pinnacle in the early 1990s as computer-based methods allowing for detailed acoustical investigations started to become available to designers.

Figure 13-2. External view of a 1:20 scale physical model. Courtesy Kirkegaard Associates.

The scaled approach used in the geometrical construction of the model implies that in order to investigate some of the relevant physical parameters,[1] specific acoustic variables will also need to be scaled accordingly.

13.2.1.1 Frequency and Wavelength Considerations in Physical Models

The issues pertaining to frequency and wavelength in scale models are best presented with the use of an example. If a source of sound generating a frequency f = 1000Hz is placed inside a room, then under standard conditions of temperature (t = 20°C) where the velocity of sound is found to be c = 344m/s, the wavelength of the sound wave is obtained from

$$\lambda = \frac{c}{f} \qquad (13\text{-}1)$$

or in this example

l = 0.34 m

The wave number k can also be used to represent the wavelength since

$$k = \frac{2\pi}{\lambda} \tag{13-2}$$

so in our example we have

k = 18.3 m^{-1}

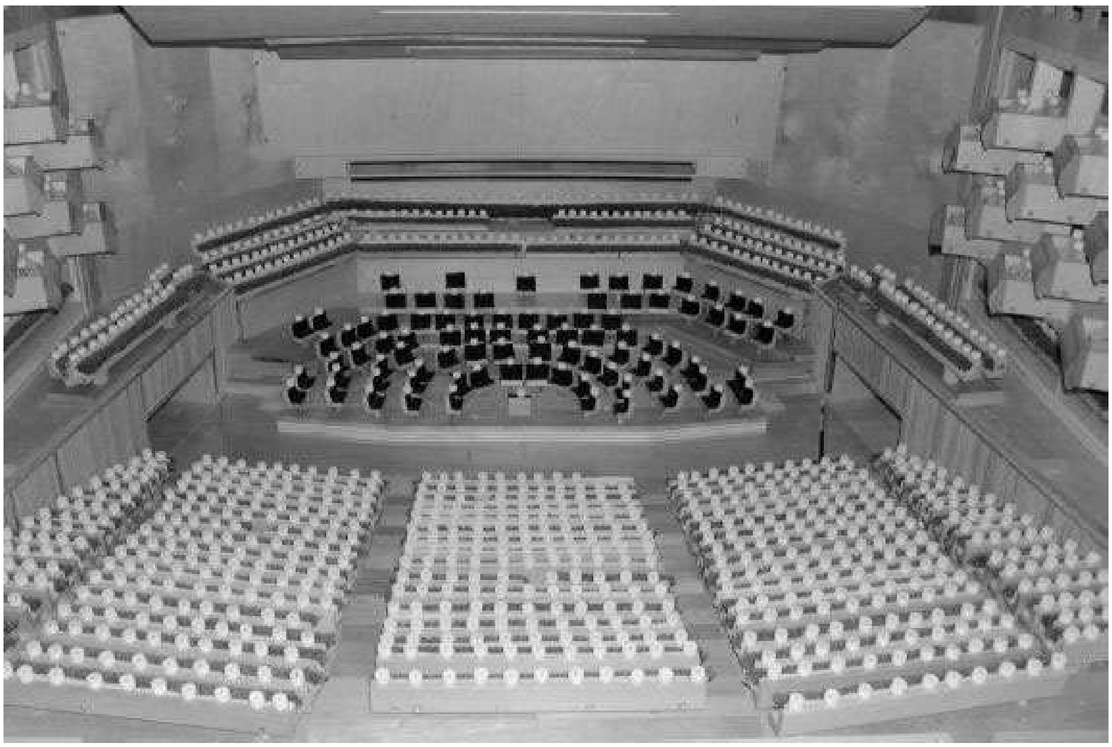


Figure 13-3. Internal view of a 1:20 scale physical model. Courtesy Kirkegaard Associates.

In the absence of acoustical absorption, the *relative* size of the wavelength of a sound wave to the size of an object in the path of the wave dictates the *primary* physical phenomenon that takes place when the sound waves reach the object. If the object can be considered acoustically hard (i.e., having a low absorption coefficient) so that the energy of the wave is not significantly reduced by absorption and if the characteristic dimension of the object (its largest dimension in the path of the sound wave) is denoted by a then the product ka can be used to predict how the sound waves will be affected by the presence of the object.

Table 13-1 shows a range of products ka and the resulting primary effect that the object will have on the sound waves.

Table 13-1. Effect of Wave Number and Object Dimension on the Propagation of Sound Waves

Value of *ka*	**Primary Phenomenon Taking Place When the Sound Waves Reach the Object (Not Including Absorption)**
$ka \leq 1$	*Diffraction*: the sound waves travel around the object without being affected by its presence. The object can be considered invisible to the waves.
$1 < ka < 5$	*Scattering*: the sound waves are partially reflected by the object in many directions and in a complicated fashion. This scattering phenomenon is associated with the notion of acoustical diffusion.
$ka > 5$	*Reflection*: the sound waves are deflected by the object in one or more specific direction(s) that can be predicted from application of basic geometry laws.

In our example if an acoustically hard object of dimension a = 0.5m is placed inside a full-size room, then the sound waves will be reflected by the object since ka = 9.1, a value clearly above the lower limit of 5 at which reflections become the primary phenomenon.

If a 1:10 scale model of the room is now created and the object is scaled by the same amount its dimension has now become a' = 0.05m. Under the earlier conditions where f = 1000Hz the product ka' now has a value of ka' = 0.91 and the conclusion that must be drawn from the guidelines presented in Table 13-1 is that the sound waves diffract around the object: in other words, the model has failed at predicting accurately the primary physical phenomenon of sound reflection that takes place when a 1000Hz sound wave strikes an acoustically hard object of 0.5m in size.

In order for the model to yield the proper conclusion, i.e., the sound waves are reflected by the object, the wavelength of the sound waves has to be scaled down by the same amount as the physical dimensions of the room, or looking at Eq. 13-1, and keeping the speed of sound a constant, the frequencies of the sound waves have to be scaled up by an amount equal to the inverse of the model's physical scale. In our example, a 10kHz frequency needs to be used in the model in order to assess the conditions that exist inside the full-size room under 1kHz excitation. If data pertaining to the room needs to be available from 50Hz to 20kHz, the use of a 1:10 scale physical model will require that frequencies from 500Hz to 200kHz be used during the investigation.

13.2.1.2 Time and Distance Considerations in Physical Models

A sound wave traveling at a velocity *c* will take a time *t* to cover a distance *x* according to the relation

$$x = ct \tag{13-3}$$

In a physical model the dimensions are scaled down and as a consequence the time of travel of the waves inside the model is reduced by the same factor. If time-domain information with a specific resolution is required from the model, then the required resolution in the time data must increase by a factor equal to the inverse of the scale in order to yield the desired accuracy.

As an example, if a sound source is placed at the end of a room with a length x = 30m and c = 344m/s, the application of relation, Eq. 13-3, shows that the sound waves can be expected to reach the other end of the room in 87.2ms. If an accuracy of ±10cm is desired in the distance information, then a time resolution of ±291µs is required for the time measurements.

In a 1:10 scale situation—and under the same conditions of sound velocity—the sound waves will now take 8.72ms to travel the length of the model and a resolution of ±29.1µs will be required in the time-measuring apparatus to yield the required distance resolution.

13.2.1.3 Medium Considerations in Physical Models

As a sound wave travels through a gaseous medium like air it loses energy because of interaction with the molecules of the media in a phenomenon known as *thermal relaxation*. Energy is also lost via spreading of the wave as it travels away from the source. The absorption in the medium as a function of distance of travel x and other parameters such as temperature and humidity can be represented by a loss factor K that is given by

$$K = e^{-mx} \tag{13-4}$$

where,

m is called the *decay index*, and it takes into account the temperature and the humidity of the air as a function of the frequency of the sound.

The value of m has been determined both analytically and experimentally for various conditions of temperature and humidity over a range of frequencies extending from 100Hz to 100kHz and an examination of the data shows that the absorption of air increases with increased frequencies and that for a given frequency the maximum absorption takes place for higher relative humidity.

Since the distance x traveled by sound waves in a physical model are scaled down in a linear fashion (i.e., by the scale factor), one cannot expect the attenuation of the waves inside the model to accurately reflect the absorption of the air since the loss factor K follows an exponential decay that is dependent upon the term m that is itself affected by the physical properties of the medium. In a scaled physical model this discrepancy is taken into account by either totally drying the air inside the model or achieving conditions of 100% relative humidity; in either case, the approach yields a simpler relation for m that becomes solely dependent on temperature and frequency. For example, in the case of totally dry air the decay index becomes

$$m = (33 + 0.2T)10^{-12}f^2 \qquad (13\text{-}5)$$

and under these conditions, it is clear that the dominant term that dictates the loss of energy in the sound wave is the frequency f since m is proportional to f^2 and varies only slightly with the temperature T.

Another available option to account for the differences in the air absorption between a scaled physical model and its full-size representation is to use a different transmission medium in the model. Replacing the air inside the model with a simple molecular gas like nitrogen will yield a decay index similar to that of Eq. 13-5 up to frequencies of 100 kHz but this is a cumbersome technique that limits the usability of the scale model.

13.2.1.4 Source and Receiver Considerations in Physical Models

To account for the primary phenomena (defined in Table 13-1) that take place over a range of frequencies from 40Hz to 15kHz inside a full-size room, one needs to generate acoustic waves with frequencies extending from 400Hz to 150kHz if a 1:10 model is used, and the required range of test frequencies becomes 800Hz to 300kHz in the case of a 1:20 scale model. The difficulties associated with creating efficient and linear transducers of acoustical energy over such frequency ranges are a major issue associated with the use of physical scale models in acoustics.

Transducers of acoustical energy that can generate continuous or steady-state waves over the desired range of frequencies and that can also radiate the acoustical energy in a point-source fashion are difficult to design, thus physical scale models often use impulse sources for excitation; in these instances the frequency information is derived from application of transform functions to the time-domain results. One commonly used source of impulse is a spark generator as shown in Fig. 13-4 where a high voltage of short duration (typically ranging from less than 20μs to 150μs) is applied across two conductors separated by a short distance. A spark bridges the air gap and the resulting noise contains substantial high-frequency energy that radiates over an adequately spherical pattern.

Figure 13-4. A spark generator used in scale physical models. Courtesy Kirkegaard Associates.

Although the typical spark impulse may contain sufficient energy beyond 30kHz, the frequency response of a spark generator is far from being regular and narrowing of the bandwidth of the received data is required in order to yield the most useful information. The bandwidth of the impulse $\Delta_{fimpulse}$ and its duration $\tau_{impulse}$ are related by the uncertainty principle

$$\Delta f_{impulse}\tau_{impulse} = 1 \qquad (13\text{-}6)$$

When dealing with impulses created with a spark generator, the received data must be processed via a bandpass filter of order to eliminate distortion associated with the nonlinearity of the spark explosion, but the filter must have a sufficient bandwidth Δf_{filter} to avoid limiting the response of the impulse.

$$\Delta f_{filter} \geq \frac{4}{\tau_{impulse}} \qquad (13\text{-}7)$$

Steady-state sound waves can be generated over spherical (nondirectional) patterns up to frequencies of about 30kHz by using specially designed electrostatic transducers. Gas nozzles can also be used to generate continuous high-frequency spectra although issues of linearity and perturbation of the medium must be taken into account for receiver locations that are close to the source.

Microphones are available with very good flatness (±2dB) over frequency responses extending beyond 50kHz. The main issues associated with the use of microphones in physical scale models are that the size of the microphone cannot be ignored when compared to the wavelength of the sound waves present in the model, and that microphones become directional at high frequencies. A typical half-inch microphone capsule with a 20cm housing in a 1:20 scale model is equivalent to a 25cm × 4m obstacle in a real room and can hardly be ignored in term of its contribution to the measurement; furthermore its directivity can be expected to deviate substantially (>6dB) from the idealized spherical pattern above 20kHz. Using smaller capsules (¼ in or even ⅛ in) can improve the omnidirectivity of the microphone but it also reduces its sensitivity and yields a lower *SNR* during the measurements.

13.2.1.5 Surface Materials and Absorption Considerations in Physical Models

Ideally, the surface materials used in a scale physical model should have absorption coefficients that closely match those of real materials planned for the full-size environment at the equivalent frequencies. For example, if a 1:20 scale model is used to investigate sound absorption from a surface at 1kHz in the model (or 50Hz in the real room) then the absorption coefficient α of the material used in the model at 1kHz should match that of the planned full-size material at 50Hz. In practice this requirement is never met since materials that have similar absorption coefficients over an extended range of frequencies are usually limited to hard reflectors where $\alpha <$ 0.02 and even under these condition, the absorption in the model will increase with frequency and deviate substantially from the desired value. The *minimum* value for the absorption coefficient of any surface in a model can be found from

$$\alpha_{min} = 1.8 \times 10^{-4}\sqrt{f} \qquad (13\text{-}8)$$

where,

f is the frequency of the sound wave at which the absorption is measured.

Thus at frequencies of 100kHz, an acoustically hard surface like glass in a 1:20 scale model will have an absorption coefficient of $\alpha_{min} = 0.06$, a value that is clearly greater than $\alpha < 0.03$ or what can be expected of glass at the corresponding 5 kHz frequency in the full-size space.

The difference in level between the energy of the *n*th reflected wave to that of the direct wave after *n* reflections on surfaces with an absorption coefficient a is given by

$$\Delta_{Level} = 10\log(1-\alpha)^{n} \qquad (13\text{-}9)$$

Considering glass wherein the model $\alpha = \alpha_{min} = 0.06$, the application of Eq. 13-9 above shows that after two reflections the energy of the wave will have dropped by 0.54dB. If the reflection coefficient is now changed to $\alpha = 0.03$ then the reduction in level is 0.26dB or a relative error of less than 0.3dB. Even after five reflections, the relative error due to the discrepancies between α and α_{min} is still less than 0.7dB, a very small amount indeed.

On the other hand, in the case of acoustically absorptive materials ($\alpha > 0.1$) the issue of closely matching the absorption coefficients in the models to those used in the real environment becomes very important. The application of Eq. 13-9 to absorption coefficients α in excess of 0.6 shows that even a slight mismatch of 10% in the absorption coefficients can result in differences of 1.5dB after only two reflections. If the mismatch is increased to 20% then errors in the predicted level in excess of 10dB can take place in the model.

Due to the difficulty in finding materials that are suitable for use in both the scaled physical model and in the real-size environment, different materials are used to match the absorption coefficient in the model (at the scaled frequencies) to that of the real-size environment at the expected frequencies. For example, a 10mm layer of wool in a 1:20 scale model can be used to model rows of seats in the actual room, or a thin layer of polyurethane foam in a 1:10 scale model can be used to represent a 50mm coating of acoustical plaster in the real space. Another physical parameter that is difficult to account for in scale physical model is stiffness, thus the evaluation of effects such as diaphragmatic absorption and associated construction techniques is difficult to model accurately.

The limitations pertaining to signal generation and capture, as well as the large approximations associated with the absorption properties of materials when frequencies are scaled are definite negatives when performing acoustical investigations with scaled models. Furthermore, the scattering phenomena associated with the propagation of the sound waves when encountering specific surfaces are only approximated in the context of strong (primary) interactions. For these reasons, large-scale physical models for acoustical investigations of indoor sound propagation are not commonly encountered anymore and have been supplanted by a wide range of computational models that allow for far more accurate investigations as well as an efficient way to answer the "what if?" questions that a designer may face when selecting structures, shapes, and materials.

13.2.2 Computational Models

This section presents models that create a mathematical representation of an acoustical environment by using assumptions that are based on either a geometrical, analytical, numerical, or statistical description of the physical phenomena (or parameters) to be considered, or on any combination of the afore-mentioned techniques, In all instances, the final output of the modeling phase is the result of extensive mathematical operations that are usually performed by computers. With the development of powerful and affordable computers and of graphical interfaces these modeling tools have become increasingly popular with acoustical designers. To various degrees, the aim of computational models is to ultimately yield a form of the impulse response of the room at a specific receiver location from which data pertaining to time, frequency, and direction of the sound energy reaching the receiver can be derived. This information can then be used to yield specific quantifiers such as reverberation time, lateral reflection ratios, intelligibility, and so on.

The inherent advantage of computational models is flexibility: changes to variables can be made very rapidly and the effects of the changes are available at no hard cost, save for that of computer time. Issues related to source or receiver placement, changes in materials and/or in room geometry can be analyzed to an infinite extent. Another advantage of computational models is that scaling is not an issue since the models exist in a virtual world as opposed to a physical one.

Computational models are themselves divided into subgroups that are fundamentally based on issues of adequacy, accuracy, and efficiency. An *adequate* model uses a set of assumptions based on a valid (true) description of the physical reality that is to be modeled. An accurate model will further the cause of adequacy by providing data that is eminently useful because of the high confidence associated with it. An *efficient* model will aim at providing fast and adequate results but maybe to a lesser—yet justified—extent in accuracy. Although issues of accuracy and of efficiency will be considered in this portion of the chapter, the discussion of the various classes of computational models will be primarily based on their adequacy.

13.2.2.1 Geometrical Models

The primary assumption that is being made in all geometrical models[2] applied to acoustics is that the wave can be seen as propagating in one or more specific directions, and that its reflection(s) as it strikes a surface is (are) also predictable in terms of direction; this is a very valid assumption when the wavelength can be considered small compared to the size of the surface and the condition $ka > 5$ presented in Table 13-1 quantifies the limit above which the assumption is valid. Under this condition, the propagating sound waves can be represented as straight lines emanating from the sources and striking the surfaces (or objects) in the room at specific points. The laws of optics involving angles of incidence and angles of reflection will apply and the term *geometrical acoustics* is used to describe the modeling technique.

The second assumption of relevance to geometrical acoustics models is that the wavelength of the sound waves impinging on the surfaces must be large compared to the irregularities in the surface, in other words the surface has to appear smooth to the wave and in this instance the irregularities in the surface will become invisible since the wave will diffract around them. If the characteristic dimension of the irregularities is denoted by b, then the condition $kb < 1$ is required using the criteria outlined in Table 13-1. This is a necessary condition to assume that the reflection is *specular*, that is that all of its energy is concentrated in the new direction of propagation. Unless this condition is met in the actual room the energy of the reflected wave will be spread out in a diffuse fashion and the geometrical acoustics assumption of the model will rapidly become invalid, especially if many reflections are to be considered.

Image Models. In this class of geometrical acoustics models, the assumption that is being made is that the only sound reflections that the model should be concerned about are those reaching the receiver, so the methodology aims at computing such reflections within time and order constraints selected by the user of the model while ignoring the reflections that will not reach the receiver. To find the path of a first-order reflection a source of sound S_0 is assumed to have an image, a virtual source S_1, located across the surface upon which the sound waves are impinging as presented in Fig. 13-5.

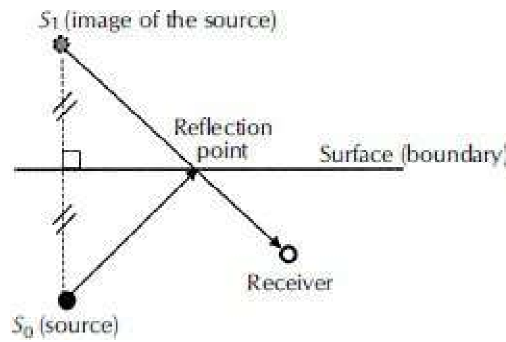


Figure 13-5. A source and its virtual image located across a boundary define the direction of the first-order reflection.

As long as the surface can be considered to be rigid, the image method allows for the prediction of the angles of reflections from the surface and can find all of the paths that may exist between a source and a receiver.[3] It also satisfies the *boundary conditions* that must take place at the surface, that is, the acoustical pressures have to be equal on both sides of the surface at the reflection point, and the velocity of the wave has to be zero at the interface. The image from the virtual source S_1 can also be used to determine where the second-order reflections from a second surface will be directed to since as far as the second surface is concerned, the wave that is impinging upon it emanated from S_1. A second order source S_2 can thus be created as shown in Fig. 13-6 and the process can be repeated as needed to investigate any order of reflections that constitutes a path between source and receiver.

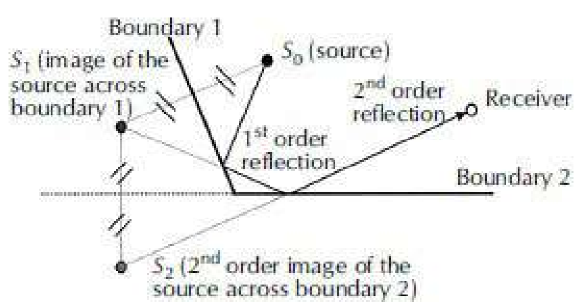


Figure 13-6. Higher-order reflections can be created by adding

virtual images of the source.

It is thus possible to collect the amplitude and the direction of *all* of the reflections at a specific location as well as a map of where the reflections emanate from. Even the reflection paths from curved surfaces can be modeled by using tangential planes as shown in Figs. 13-7 and 13-8.

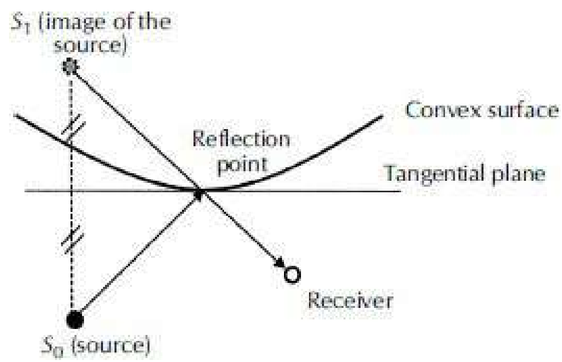


Figure 13-7. Image construction from a convex plane. Adapted from Reference 1.

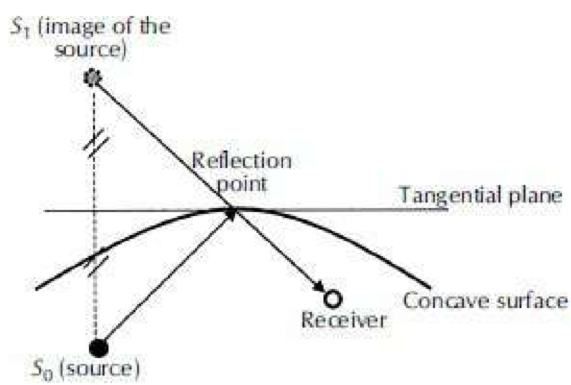


Figure 13-8. Image construction from a concave plane Adapted from Reference 1.

Since the speed of sound can be assumed to be a constant inside the room, the distance information pertaining to the travel of the reflections can be translated into time-domain information; the result is called a *reflectogram* (or sometimes an *echogram*) and it provides for a very detailed investigation of the reflections inside a space that reach a receiver at a specific location. A sample of a reflectogram is shown in Fig. 13-9.

Although it was originally developed solely for the determination of the low-order specular reflections taking place in rectangular rooms due to the geometric increase in the complexity of the computations required, the technique was expanded to predict the directions of the specular reflections from a wide range of shapes.[4] In the image method, the boundaries of the room are effectively being replaced by sources and there is no theoretical limit to the order of reflection that can be handled by the image methodology. From a practical standpoint, the number of the computations that are required tends to grow exponentially with the order of the reflections and the number of surfaces as shown in Eq. 13-10, where N_{IS} represents the number of images, N_W is the number of surfaces that define the room, and i is the order of the reflections.[5]

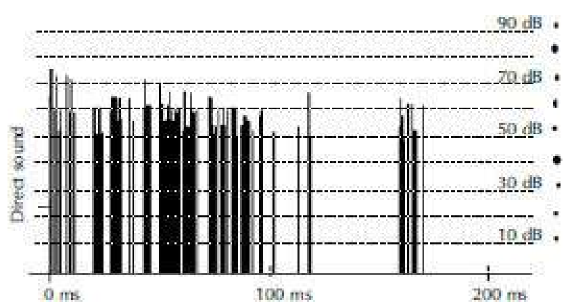


Figure 13-9. A reflectogram (or echogram) display of reflections at a specific point inside a room.

$$N_{IS} = \frac{N_W}{N_W - 2}[(N_W - 1)^i - 1] \qquad (13\text{-}10)$$

Furthermore, reflections must be analyzed properly in terms of their visibility to the receiver in the case of complicated room shapes where some elements may block the reflections from the receiver as shown in Fig. 13-10.

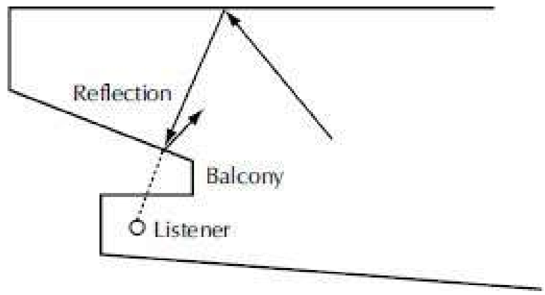


Figure 13-10. The reflection is not visible to the listener due to the balcony obstruction.

The model must also constantly check for the validity of the virtual sources to insure that they actually contribute to the real reflectogram by emulating reflections taking place inside the room and not outside its physical boundaries. Fig. 13-11 illustrates such a situation.

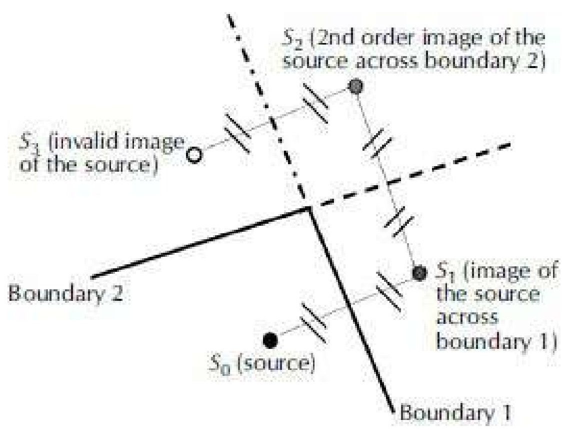


Figure 13-11. Invalid images can be created when a virtual source is reflected across the boundary used to create it. Adapted from Reference 4.

In Fig. 13-11 a real source S_0 creates a first-order image S_1 across boundary 1. This is a valid virtual source that can be used to determine the magnitude and the direction of first-order specular reflections on the boundary surface 1. If one attempts to create a second-order virtual source S_2 from S_1 with respect to boundary surface 2 to find the second order reflection, the image of this virtual source S_2 with respect to boundary 1 is called S_3 but it is contained outside the boundary used to create it and it cannot represent a physical reflection.

Once the map of all the images corresponding to the reflection paths has been stored, the intensity of each individual reflection can be computed by applying Eq. 13-9 introduced earlier. Since the virtual sources do represent the effect of the boundaries on the sound waves, the frequency dependence of the absorption coefficients of the surfaces is modeled by changing the power radiated by the virtual sources; thus, once the image map is obtained the model can be used to rapidly simulate an unlimited number of "what if" simulations pertaining to material changes as long as the locations of the sources and the receiver remain unchanged. A further correction for the air absorption resulting from the wave traveling over extended distances can also be incorporated at this time in the simulation. The same reasoning applies to the frequency distribution of the source: since the image map (and the resulting location of the reflections in the time domain) is a sole function of source and receiver position, the image model can rapidly perform "what if" simulations to yield reflectograms at various frequencies.

The image methodology does not readily account for variations in the absorption coefficient of the surfaces as a function of the angle of incidence of the wave. When taking into account all of the properties in the transmission medium, it can be shown that many materials will exhibit a substantial dependence of their absorption coefficient on the incidence angle of the wave, and in its most basic implementation the image method can misestimate the intensity of the reflections. It is however, possible to incorporate the relationship between angle of incidence and absorption coefficient into a suitable image algorithm in order to yield more accurate results, although at the expense of computational time.

In an image model the user can control the length of the reflection path as well as the number of segments (i.e., the order of the reflections) that comprise it. This allows for a reduction in the computational time of the process since virtual sources located beyond a certain distance from the receiver location can be eliminated while not compromising the fact that all of the reflections within a specific time frame are being recorded, and the image method can lead to very accurate results in the modeling of the arrival time of reflections at a specific location.

Efficient computer implementations of the image methodology have been developed[4] to allow for a fast output of the reflections while also checking for the validity of the images and for the presence of obstacles. Still the method is best suited to the generation of very accurate reflectograms of short durations (500ms or less) and limited number of reflections (fifth order maximum for typical applications). These factors do not negatively affect the application of the image method in acoustical modeling since in a typical large space—like an auditorium or a theater—the sound field will become substantially diffuse after only a few reflections and some of the most relevant perceived attributes of the acoustics of the space are correlated to information contained in the first 200ms of the reflectogram when dealing with a mid-to-large size space. This being said, the ongoing advances in computer power and memory have made the creation of reflectograms of almost any length a reality as the computational costs keep on coming down.

Ray-Tracing Models. The ray-tracing methodology follows the assumptions of geometrical acoustics presented at the onset of this section, but in this instance the source is modeled to emit a finite number of rays representing the sound waves in either an omnidirectional pattern for the most general case of a point source, or in a specific pattern if the directivity of the source is known. Fig. 13-12 shows an example of a source *S* generating rays inside a space and how some of the rays are reflected and reach the receiver location *R*.

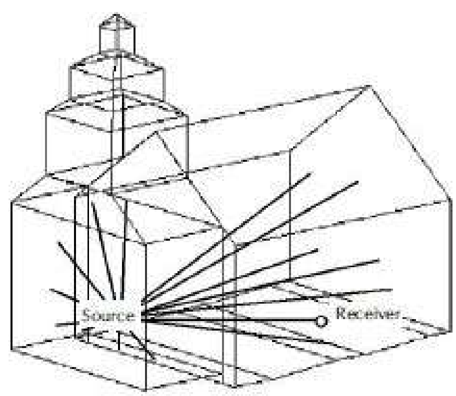


Figure 13-12. Rays are generated by a source, *S*. Some of the rays reach the receiver, *R*.

In this instance, the goal is not to compute all of the reflection paths reaching the receiver within a given time frame but to yield a high probability that a specified density of reflections will reach the receiver (or detector usually modeled as a sphere with a diameter selected by the user) over a specific time window.

In the case of the image method, the boundaries of the room are replaced by virtual sources that dictate the angle of the reflections of the sound waves. In a similar fashion, the ray-tracing technique creates a virtual environment in which the rays emitted by the source can be viewed as traveling in straight paths across virtual rooms until they reach a virtual listener as presented in Fig. 13-13.

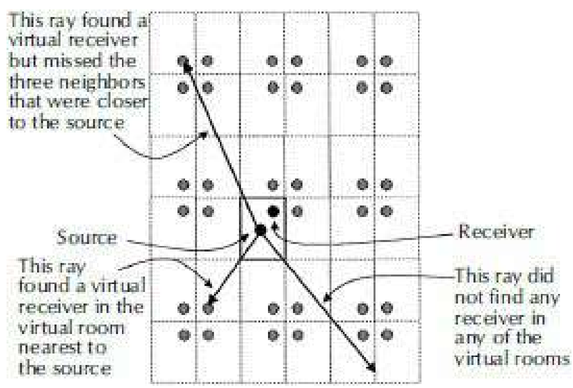


Figure 13-13. The rays can be seen as traveling in straight paths across virtual images of the room until they intercept a receiver. Adapted from Reference 4.

The time of travel and location of the ray are then recorded and can yield a reflectogram similar to that presented earlier in Fig. 13-9.

The main advantage of the ray-tracing technique is since the model is not trying to find all of the reflection paths between source and receiver, the computational time is greatly reduced when compared to an image technique; for a standard ray-tracing algorithm, the computational time is found to be proportional to the number of rays and to the desired order of the reflections. Another advantage inherent to the technique is that multiple receiver locations may be investigated simultaneously since the source is emitting energy in all directions and the model is returning the number and the directions of rays that are being detected without trying to complete a specific path between source and receiver. On the other hand, since the source is emitting energy in many directions and one cannot dictate what the frequency content of a specific ray is versus that of another, the simulations pertaining to the assessment of frequency-dependent absorption must be performed independently and in their entirety for each frequency of interest. Another substantial advantage of the ray-tracing approach is that individual rays can be “followed” throughout the model; this is of great value when trying to assess interferences as more and more reflections contribute to the creation of the sound field as the receiver moves away from the source. Because the time of travel for each ray is also recorded for each portion of its path, it is also possible to use the information to infer upon irregularities in the frequency response that are associated with delays, such as comb-filtering and potential “flanging” effects.

One problem associated with the ray-tracing technique is that the accuracy of the detection is strongly influenced by size of the detector. A large spherical detector will record a larger number of hits from the rays than another spherical detector of smaller diameter, even if the respective centers of the spheres are located at the exact same point in space. Furthermore, the ray-tracing method may lead to an underestimation of the energy reaching the detector (even if its size is considered adequate) unless large numbers of rays are used since the energy is sampled via rays that diverge as they spread from the source thus increasing the possibility that low-order reflections may miss the detector.

Techniques combining the image methodology and the ray-tracing approach have been developed.[5] The algorithms aim at reducing the number of images to be considered by using the more computationally efficient ray-tracing technique to conduct the visibility test required by the image method.

Beam-Tracing Models. The triangular area that is defined between two adjacent rays emanating from a source is called a 2D ray; more than two rays can also be used to define a 3D pyramidal or conical region of space in which the acoustical energy is traveling away from the source. In these instances, the source is viewed at emitting beams of energy, and the associated modeling techniques are known as *beam-tracing* methods. Fig. 13-14 shows an example of a beam and its reflection path from a surface.

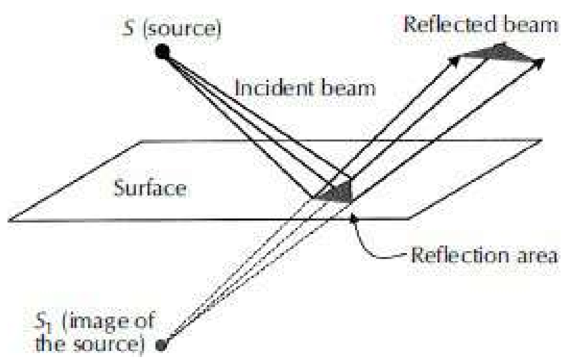


Figure 13-14. A 3D beam is emitted by a source S and reflects at a surface.

The beam-tracing technique offers the advantage of guaranteeing that the entire space defining the model will receive energy since the directions of propagations are not sampled as in the case of the traditional ray-tracing approach. Virtual source techniques are used to locate the points that define the reflection zones across the boundaries of the room. On the other hand, the technique requires very complex computations to determine the reflection patterns from the surfaces since the reflection cannot be viewed as a single point as in the case of the ray-tracing technique: when 2D beams are used, the reflections from the surfaces must be considered as lines, while 3D beams define their reflections as areas. Care must also be taken to account for overlapping of the beams by each other or truncation of the beams by obstacles in the room, however an assessment of the latter can yield to very valuable information when determining the impact of “acoustical shadows” that may have a noticeable impact in the overall sound pressure level as well as on the frequency content of a signal that reaches a listener.

Although the computational complexity of the model is substantially increased when it comes to assessing the direction of the reflections, the departure from the single point reflection model presents numerous advantages over the traditional image and/or ray-tracing technique. The issues associated with the divergence of the reflections as a function of increased distance from the source are naturally handled by the beam-tracing approach. Furthermore, the effects of acoustical diffusion can be modeled—at least in an estimated fashion—since the energy contained in the beams can be defined as having a certain distribution over either the length of the intersecting lines (for 2D beams) or areas (for 3D beams). For example, an adaptive beam-tracing model[6] that controls the cross-sectional shape of the reflecting beam as a function of the shape of the reflecting surfaces also allows for an evaluation of the diffuse and specular energy contained inside a reflecting beam. If the energy contained inside the incident beam is E_B and the energy reflected from a surface is E_R, then one can write

$$E_R = E_B(1-\alpha)(1-\delta) \qquad (13\text{-}11)$$

where,

α is the surface's absorption coefficient,

δ is the surface's diffusion coefficient.

The energy E_D that is diffused by the surface is found to be proportional to the area of illumination A and inversely proportional to the square of an equivalent distance L between the source and the reflection area

$$E_D \propto \frac{E_B A \delta(1-\alpha)}{4\pi L^2} \qquad (13\text{-}12)$$

The adaptive algorithm allows for a separate assessment of the specular and of the diffuse reflections from the same geometrical data set that represents the travel map of the beams inside the space. In this instance the diffused energy from a given surface is redirected to other surfaces in a recursive fashion via radiant exchange, a technique also used in light rendering applications. The diffuse and the specular portions of the response can then be recombined to yield a reflectogram that presents a high degree of accuracy, especially when compared to traditional ray-tracing techniques. Fig. 13-15 shows a comparative set of impulse response reflectograms obtained by the adaptive beam tracing, the image, the ray-tracing, and the nonadaptive beam-tracing techniques in a model of a simple performance space containing a stage and a balcony.

The adaptive model is able to yield a reflectogram that is extremely close to that obtained with an image method—i.e., it is able to generate all of the possible reflections paths at a single point in space. From the perspective of computing efficiency, the adaptive-beam tracing methodology compares favorably with the image methodology especially when the complexity of the room and/or the order of the reflections is increased.

Other variants of the beam-tracing approach have been developed. In the *priority-driven* approach,[7] the algorithms are optimized to generate a series of the most relevant beams from a psychoacoustics perspective so that the reflectogram can be generated very rapidly, ideally in real time and the model can be used in an interactive fashion. The beams are ranked in order of importance based on a priority function that aims at accurately reproducing the early portion of the reflectogram since it is by far the most relevant to the perception of the space from the perspective of psychoacoustics. The late portion of the reflectogram (the late reverberation) is then modeled by using a smooth and dense energy profile that emulates the statistical decay of the energy in a large space.

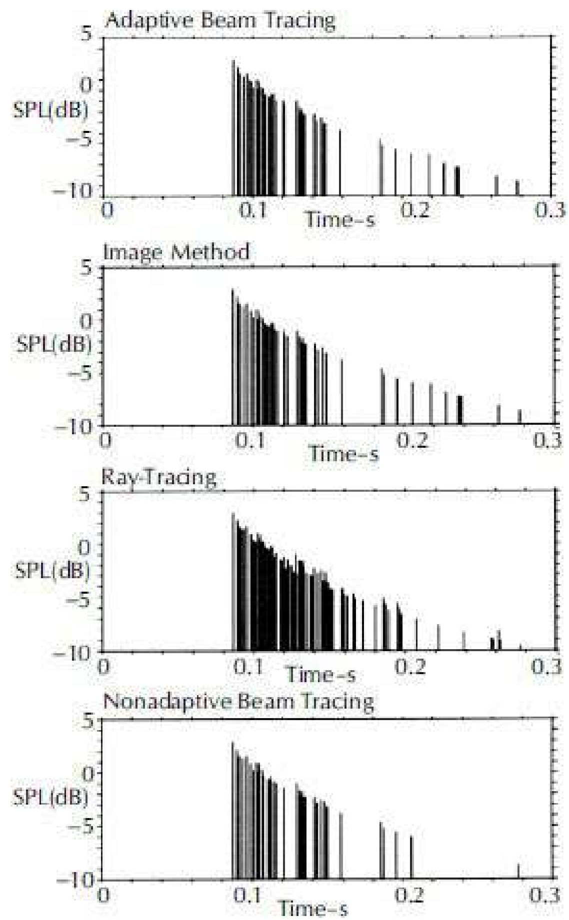


Figure 13-15. Comparative reflectograms for a simple room model. From Reference 6.

A note on diffusion: The issue of diffusion has been of prime interest to the developers of computer models since commercially available or custom-made diffusers are often integrated into room designs, sometimes at a high cost. Although diffusion is an essential qualitative part of the definition of a sound field, the quantitative question of "how much diffusion is needed?" is often answered using considerations that have little foundation in scatter theory and/or general acoustics. A concept as elementary as reverberation finds its classical quantitative representation (the Sabine/Eyring equation and associated variants) rooted into the notions that the sound field is assumed to be diffuse as well as even in terms of energy density throughout the space prior to the start of the decay, and unless these conditions are being met in reality, one will encounter substantial errors in predicting the reverberation time at various locations within a reverberant space. Today's advanced large room computer acoustic simulation software products incorporate the ability to model diffused reflections using either a frequency dependence function or an adaptive geometry that spreads out the incident energy of the sound ray over a finite area. This allows for a much more accurate correlation between predicted and test data especially in rooms that have geometry involving shapes and aspect ratios that are out of the ordinary, noneven distribution of absorptive surfaces, and/or coupled volumes.8 Under these conditions the incorporation of diffusion parameters into the model is necessary and a specular-only treatment of the reflections (even when using an efficient ray-tracing technique) will lead to potentially substantial differences between measured and predicted values.

13.2.2.2 Wave Equation Models

Wave equation models are based on an evaluation of the *fundamental wave equation*, which in its simplest form relates the pressure p of a wave at any point in space to its environment via the use of the 3D Laplacian operator ∇^2 and the wave number k

$$\nabla^2 p + k^2 p = 0 \qquad (13\text{-}13)$$

Solving the fundamental wave equation allows for an exact definition of the acoustic pressure at any specific point since appropriate boundary conditions defining the physical properties of the environment (surfaces, medium) can be used whenever required. As an example, in a model based on the wave equation the materials that comprise the environment (like the room) can be defined in terms of their *acoustical impedance* z given by

$$\mathbf{z} = \frac{\mathbf{p}}{\mathbf{U}} \qquad (13\text{-}14)$$

where,

p refers to the pressure of the wave,

U to its velocity in the medium.

When using the wave equation, issues having to do with diffraction, diffusion, and reflections are automatically handled since the phenomena are assessed from a fundamental perspective without using geometrical simplifications. The main difficulty associated with the method is found in the fact that the environment (surfaces and materials) must be described accurately in order for the wave equation to be applied: either an analytical or a numerical approach can be used to achieve this goal.

Analytical Model: Full-Wave Methodology

An analytical model aims at providing a mathematical expression that describes a specific phenomenon in an accurate fashion based on underlying principles and/or physical laws that the phenomenon must obey. Because of this requirement the analytical expression governing the behavior of a model must be free of correction terms obtained from experiments and of parameters that cannot be rigorously derived from—or encountered in—other analytical expressions.

The complexity of the issues associated with sound propagation has prevented the development of a single and unified model that can be applied over the entire range of frequencies and surfaces that one may encounter in acoustics; most of the difficulties are found in trying to obtain a complete analytical description of the scattering effects that take place when sound waves impinge on a surface. In the words of J.S. Bradley,[9] one of the seminal researchers in the field of architectural acoustics:

> *Without the inclusion of the effects of diffraction and scattering, it is not possible to accurately predict values of conventional room acoustics parameters [...]. Ideally, approximations to the scattering effects of surfaces, or of diffraction from finite size wall elements should be derived from more complete theoretical analyses. Much work is needed to develop room acoustics models in this area.*

In this section, we briefly present the full-wave methodology,[10] one analytical technique that can be used for the modeling of the behavior of sound waves as they interact with nonidealized surfaces, resulting in some of the energy being scattered, reflected, and/or absorbed as in the case of a real space. Due to the complexity of the mathematical foundation associated with this analytical technique, only the general approach is introduced here and the reader is referred to the bibliography and reference section for more details.

The full-wave approach (originally developed for electromagnetic scattering problems) meets the important condition of acoustic reciprocity requiring that the position of the source and of the receiver can be interchanged without affecting the physical parameters of the environment like transmission and reflection coefficients of the room's surfaces. In other words if the environment remains the same, interchanging the position of a source and of a receiver inside a room will result in the same sound fields being recorded at the receiver positions. The full-wave method also allows for exact boundary conditions to be applied at any point on the surfaces that define the environment, and it accounts for all scattering phenomena in a consistent and unified manner, regardless of the relative size of the wavelength of the sound wave with that of the objects in its path. Thus the surfaces do not have to be defined by general (and less than accurate) coefficients to represent absorption or diffusion, but they can be represented in terms of their inherent physical properties like density, bulk modulus, and internal sound velocity.

The full-wave methodology computes the pressure at every point in the surface upon which the waves are impinging, and follows the shape of the surface. The coupled equations involving pressure and velocity are then converted into a set of equations that separate the forward (in the direction of the wave) and the backward components of the wave from each other, thus allowing for a detailed analysis of the sound field in every direction. Since the full-wave approach uses the fundamental wave equation for the derivation of the sound field, the model can return variables such as sound pressure or sound intensity as needed.

The main difficulty associated with the full-wave method is that the surfaces must also be defined in an analytical fashion. This is possible for simple (i.e., planar, or curved) surfaces for which equations are readily available, but for more complicated surfaces—such as those found in certain shapes of diffusers—an analytical description is more difficult to achieve, and the methodology becomes restricted to receiver locations that are located at a minimum distance from the surfaces upon which the sound waves are impinging. Still for many problems the full-wave methodology is a very accurate and efficient way to model complicated scattering phenomena.

13.2.2.3 Numerical Model: Boundary Element Methodology

The *boundary element analysis* (BEA) techniques are numerical methods that yield a quantitative value of the solution to the problem under investigation. BEA techniques[11,12,13,14] can be used in solving a wide range of problems dealing with the interaction of energy (in various forms) with media such as air and complex physical surfaces, and they are well suited to the investigation of sound propagation in a room. Although the method is based on solving the fundamental differential wave equation presented earlier, the BEA methodology makes use of an equivalent set of much simpler algebraic equations valid over a small part of the geometry, and then expands the solution to the entire geometry by solving the resulting set of algebraic equations simultaneously. In essence, the BEA technique replaces the task of solving one very complex equation over a single complicated surface by that of solving a large quantity of very simple equations over a large quantity of very simple surfaces. In a BEA implementation the surface is described using a meshing approach as shown in Fig. 13-16.

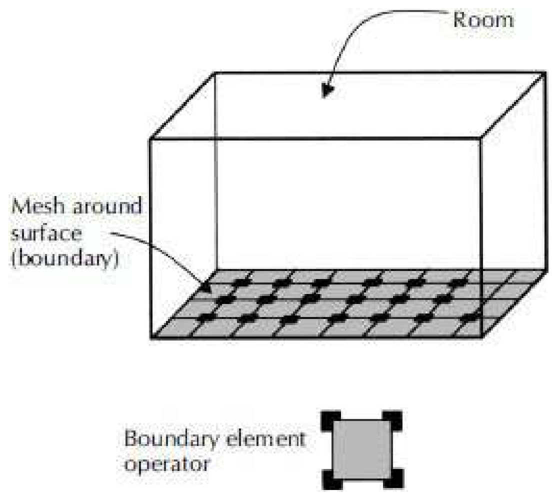


Figure 13-16. A mesh describes a boundary in the BEA method.

In the BEA method the analytical form of the solution over the small domain (area) of investigation is not directly accessible for modification. The use exercises control over the solution by properly specifying the domain (geometry) of the problem, its class (radiation or scattering), the parameters of the source (power, directivity, location), and, of course, the set of boundary conditions that must be applied at each area defined by the mesh. It is thus possible to assign individual material properties at every location in the mesh of the model in order to handle complex scattering and absorption scenarios, if needed. Although it can be adapted to solving acoustical problems in the time domain the BEA technique is better suited to providing solutions in the frequency domain since the characteristics of the materials are considered to be time-invariant but frequency dependent.

The main issue that is associated with the use of BEA methodology for the investigation of acoustical spaces is that the size of the elements comprising the mesh representing the surfaces dictates the accuracy of the solution. A small mesh size will, of course, allow for a very accurate description of the surfaces, both geometrically and in terms of its materials, but it will also drastically affect the computational time required to yield a solution. On the other hand, a large mesh size will yield very fast results that may be inaccurate because the algebraic equations that are used in lieu of the fundamental wave equation improperly being applied over large surfaces. A comparison of the accuracy yielded by BEA techniques over very simple geometries indicates that a minimum ratio of seven to one (7:1) must exist between the wavelength of the sound and the element size in order to bind the dependence of the BEA analysis on the size of its element to less than a ±0.5dB resolution. In other words, the wavelengths considered for analysis must be at least seven times larger than the largest mesh element in order for the methodology to be accurate. For this reason the BEA methodology is very efficient and accurate to model sound propagation at low frequencies (below 1000Hz), but it becomes tedious and cumbersome at higher frequencies since in this instance the mesh must be modeled with better than a 50mm resolution. Still the technique can be shown to yield excellent results when correlating modeled projection and actual test data from complicated surfaces such as diffusers.[13]

Numeric Model: Finite Difference Time-Domain Methodology. As mentioned earlier, the BEA techniques are best suited to yielding data in the frequency domain, although they can be adapted to provide time-domain information albeit at a cost in computing efficiency. Another numerical methodology that uses a discrete representation of the acoustical environment is known as *Finite-Difference Time-Domain* (FDTD) and it is very efficient in terms of computational speed and storage while also offering excellent resolution in the time domain. It has been demonstrated[15] that the technique can be used to effectively model low-frequency problems in room acoustics simulations and the results are suitable for the generation of reflectograms.

In the finite difference (FD) approach instead of describing the surface with a mesh (as with the BEA technique), a grid is used and the algebraic equations are solved at the points of the grid as shown in

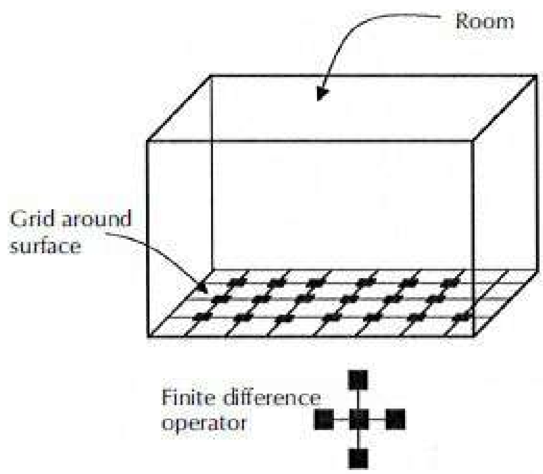


Figure 13-17. A grid describes a surface in the FD method.

In this instance the size of the grid can be made as small as needed to provide a high degree of resolution when needed and the grid points can be defined using the most effective coordinate system for the application. For example, a flat surface could be defined with a (x, y, z) Cartesian grid system while cylinders (for pillars) and spheres (for audience's heads) could be expressed with cylindrical and spherical systems respectively.

13.2.2.4 Statistical Models

The use of statistics in acoustical modeling is primarily reserved for the study of the behavior of sound in rectangular and rigid rooms where the dominant phenomena that are taking place are related to modes. The issues of modal frequencies, modal density, and mode distributions are presented, along with the appropriate descriptive equations in Chapter 5, Small Room Acoustics.

Another application of statistics in acoustical modeling can be found in situations where resonance effects take place at high frequencies, as opposed to the traditionally low frequencies associated with room modes. A technique known as Statistical Energy Analysis[16] (SEA) can be used to accurately account for the effect of modal resonance effects that take place in systems such as partitions and walls, by analyzing the kinetic energy and the strain energy associated with vibrating structures. An SEA model will describe a vibrating system (such as a wall) with mass and spring equivalents and will allow for the analysis of the effect that adding damping materials will have on the vibration spectrum. SEA techniques are optimized for frequency-domain analysis and the output cannot be used for time-domain applications to add information to the impulse response of a room, or to yield a reflectogram; still, the main advantage of SEA is that real materials such as composite partitions with different degrees of stiffness, construction beams, and acoustical sprays can be modeled in a precise manner (i.e., not only in terms of a unique physical coefficient) over an extended range of frequency. Because SEA techniques can take a complex acoustical environment in terms of material properties (different stiffness, densities, absorption parameters, etc.) and divide the overall model into distinct sub-systems, the impact of specific materials and/or structures on the overall model can be assessed in an efficient fashion. This makes SEA a tool of choice when dealing with complex acoustic environment such as the interior of an automobile, where a wide range of surfaces and finishes are present and where other stimuli such as vibrations and exterior noise will impact the overall sound field.

13.2.2.5 Small Room Models

A room that is acoustically small can be defined as one in which classically defined reverberation phenomena (using the assumption of a diffuse sound field) do not take place, but rather that the sound decays in a nonuniform manner that is a function of where the measurement is taken. The use of diffusion algorithms in large room models that rely on either ray-tracing, image source, or adaptive algorithms has vastly improve the reliability of the prediction models in a wide range of spaces, however accurate predictions of the sound field can be made in small rooms considering the interference patterns that result from modal effects. Figs. 13-18 and 13-19 shows the mapping[17] of the interference patterns resulting from modal effects in an 8m × 6m room where two loudspeakers are located at B1 and B2. In the first instance, a modal effect at 34.3Hz creates a large dip in the response at about 5.5m, while the second case shows a very different pattern at 54.4Hz. Such models are very useful to determine the placement of low-frequency absorbers into a room in order to minimize the impact of modal effects at a specific listening location, and they are a good complement to the large room models that typically do not investigate the distribution of the sound field at the very low frequencies. One of the chief advantage of small room models is that they are easy to generate as the computational algorithms are very simple. Modal densities and distributions can lead to valuable insight about the effects that a small room will have when it comes to "coloration" of the sound in the mid-bass and in the midrange.

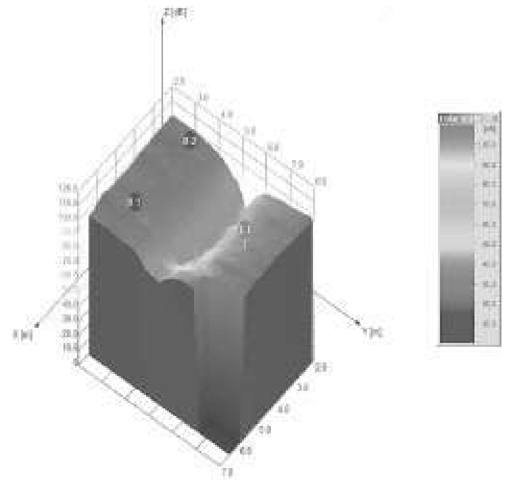


Figure 13-18. A mapping of the sound field resulting from modal interference patterns at 34.3Hz. From CARA.

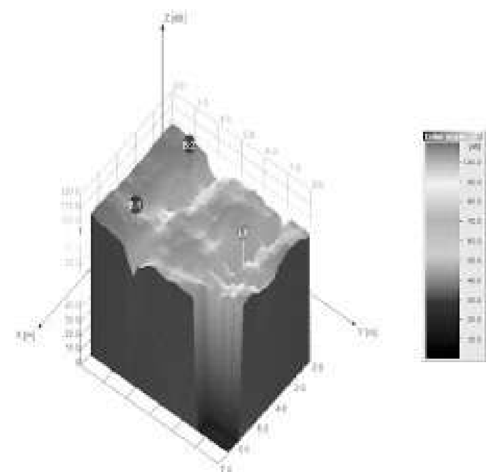


Figure 13-19. A mapping of the sound field resulting from modal interference patterns at 54.4Hz. From CARA.

13.2.3 Empirical Models

Empirical models are derived from experiments and are described by equations that typically follow curve-fitting procedures of the data obtained in the observations. No analytical, geometrical, and/or statistical expression is developed to fully explain the interdependence of variables and parameters in the model, but a general form of a descriptive expression may be constructed from underlying theories. Empirical models have been extensively used for many years in acoustical modeling due to the large quantity of variables and parameters that are often present when dealing with issues of sound propagation in a complicated environment, and this section will present only a couple of examples.

13.2.3.1 Gypsum Cavity Wall Absorption

There is numerous test data available pertaining to the sound transmission class (STC) of various wall constructions, but little has been investigated regarding the sound absorption of walls constructed of gypsum (dry-wall) panels. The absorption of a composite wall panel is partly diaphragmatic (due to the mounting), partly adiabatic (due to the porosity of the material and the air), and some energy is lost inside the cavity via resonance. The complicated absorption behavior of gypsum walls has been described[18] using an empirical model that takes into account absorption data acquired in reverberation chamber experiments. The mathematical model is fitted to the measured data to account for the resonant absorption of the cavity by assuming that the mechanical behavior of the wall can be modeled by a simple mechanical system.

In this model, the resonance frequency at which the maximum cavity absorption takes place is given by

$$f_{MAM} = P\sqrt{\frac{m_1 + m_2}{d(m_1 m_2)}} \qquad (13\text{-}15)$$

where,

m_1 and m_2 are the mass of the gypsum panels comprising the sides of the wall in kg/m^2,

d is the width of the cavity expressed in mm,

P is a constant with the following values:

If the cavity is empty (air), P = 1900,

If the cavity contains porous or fibrous sound-absorptive materials, P = 1362.

The empirical model combines the maximum absorption α_{MAM} taking place at the resonant frequency given by Eq. 13-15 with the high-frequency absorption α_S into a form that fits data obtained experimentally, to give an equation that allows for the prediction of the absorption coefficient of the wall as a function of frequency

$$\alpha(f) = \alpha_{MAM}\left(\frac{f_{MAM}}{f}\right)^2 + \alpha_{MAM} \qquad (13\text{-}16)$$

Although it does not take into account all of the construction variables (stud spacing, bonding between layers) the model still provides accurate prediction of the sound absorption parameters of various gypsum wall constructions.

13.2.3.2 Absorption from Trees and Shrubs

When dealing with issues related to outdoor noise propagation one may need to predict the anticipated noise reduction that can be expected from vegetation. In this instance, some of the general attributes of a tree-barrier such as height and width can be modeled from geometry, but others like leaf density, wind resistance, or diffraction effects from trunks may prove very difficult to describe either analytically or geometrically. In this instance an empirical model that fits experimental data to polynomial equations based on statistical regression is the most appropriate[19] to yield the sound pressure level at various distances from a sound source while taking into account tree height, width of the tree barrier, wind velocity, and tree type. An example of such an equation is presented below, and it is shown to give excellent (±1dB) accuracy between predicted and observed levels for distances extending 150ft to 400ft from the source that is assumed to be truck noise on an interstate highway. The receiver is shielded from the traffic by a belt of conifer trees planted along the interstate.

$$L_{dB} = 81.65 - 0.2257H - 0.0229W + 0.728V - 0.0576D \qquad (13\text{-}17)$$

where,

L_{dB} is the predicted sound level behind the tree belt,

H is the height of the tree belt, expressed in ft,

W is the width of the tree belt, expressed in ft,

V is the wind velocity component in the direction of the sound propagation, expressed in mph,

D is the distance from the receiver to the tree belt.

Other equations are available for different sources and different types of trees. In this class of empirical models, no attempt is made to support the equation by analytical expressions but this does not affect the usefulness or the accuracy of the model.

13.2.4 Hybrid Models

As the name implies *hybrid models* use a combination of techniques to yield results and the choice of the technique may be based on a specific need such as fast output, accuracy, range of applicability, etc. A hybrid model can combine the inherent accuracy of the image method for the determination of reflection arrival time in the specular case, with an adaptive beam-tracing approach when diffusion is required, and may also incorporate some BEM computations for complicated materials wherever required. A hybrid model can also rely on empirical approaches to provide a confidence factor for results obtained from physical scale models or from statistical approaches.

An example of hybrid techniques can be found in models that are aimed at assessing outdoor noise propagation.[20] In this instance, the objects that are in the path of the sound waves are typically buildings or large natural obstacles and can be considered to be much larger than the wavelength, except for the lowest frequencies, and as such, the geometrical acoustics assumptions apply very well; as such, the image method is very appropriate to compute reflection paths between obstacles. On the other hand one cannot ignore the fact that outdoor noise may contain a lot of low frequencies and that diffraction effects will take place; in this instance the model must use an appropriate description of diffraction such as the one presented in Chapter 7, Acoustical Noise Control, and the model may also be refined from empirical data table to represent complicated sources such as car traffic, aircraft, and trains since point source assumptions become invalid and the sources are also moving. Figs. 13-20 and 13-21 shows the type of noise prediction maps that can be obtained from such a model; in the first instance the noise sources are a combination of street traffic and large mechanical systems, and the model takes into account the diffraction effects of various buildings. In the second instance, the model is used to assess the difference in expected noise levels between different types of pavements (asphalt vs. concrete) based on traffic data on a segment of road that is surrounded by residences.

Hybrid models are also found in construction applications, where they combine analytical techniques based on specific equations with databases of test results obtained in the laboratory and in the field. As an example, a simple model could be developed using the Mass Law in order to predict the sound transmission between two spaces and yield an estimate of the Sound Transmission Class (STC) of the partition, however, the results would not be very useful because they would extensively be influenced by the construction technique and the presence of flanking paths. With a model that takes into account the construction techniques of the partition,[21] the results are much more accurate and provide the designer with valuable insight on the weak links of the construction as they pertains to noise transmission.

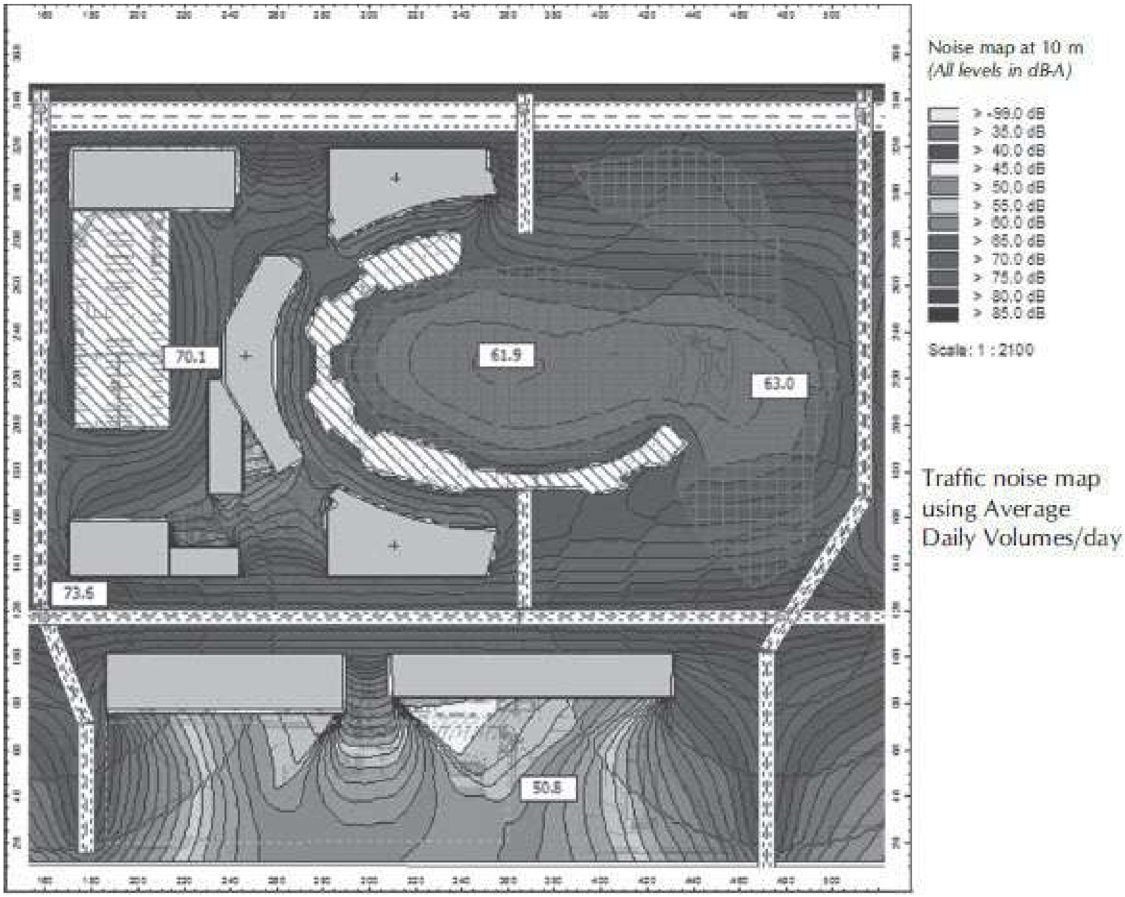


Figure 13-20. A noise map from an outdoor propagation model. CadnaA by DataKustik GmbH.

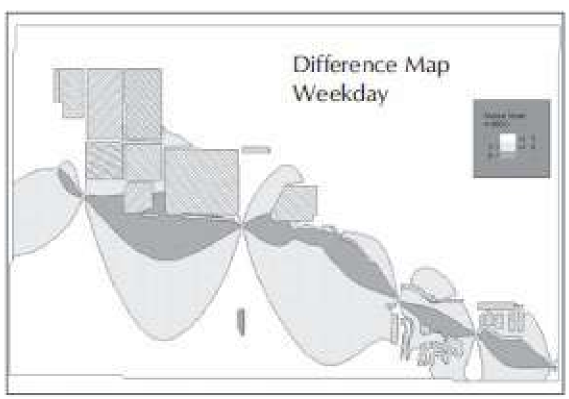


Figure 13-21. Difference maps for noise generated by different types of road pavements. Courtesy SoundPLAN, by SoundPLAN LLC.

Hybrid models are rapidly becoming a new standard for a wide range of applications pertaining to indoor and outdoor noise propagation. These advanced models use a combination of image methods to accurately localize reflections, and of ray-tracing techniques to model the radiation pattern of specific sources in terms of spatial and frequency characteristics. The physical parameters of the materials that may interact with the sound waves are described using various metrics (mass, density, absorption coefficients, measured transmission loss) and the results can be analyzed over a wide scale, from a "global" presentation, i.e., environment, to a more "local" perspective such as how sound would travel from one space to another. In the context of outdoor noise propagation the models can also be adapted to take into account complex effects such as meteorological conditions[22] (humidity, temperature, wind) although at the time of this writing there are limited options available to accurately describe and model the effect of complex phenomena such as temperature inversions in a practical format that can be easily applied to "real-world" scenarios.

13.3 Auralization

> *Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.*[23]

Auralization systems have been in existence since the 1950s.[24] During the early experiments, researchers used a physical 1:10 scale model in which a tape containing speech and music samples was played back at scaled-up speed through a scaled omnidirectional source while also taking into account the air absorption and scaling the reverberation time of the model. A recording of the sound was made at the desired receiver locations using a scaled dummy head and the results were played back at a scaled-down speed under anechoic conditions using two speakers. The sound of the model was then subjectively assessed and compared to that perceived in the real room.

The technique—or variants of it—was used for the prediction of the acoustics of both large and small rooms throughout the 1970s, however, with computer systems becoming increasingly faster and more affordable auralization techniques based on computational models have been developed to yield an audible representation of the sound field at any specified receiver location by using the results of the acoustical modeling phase. Various implementations of auralization have been put into place[25,26,27,28] at the time of this writing but because of the tremendous developments that are taking place in the field of auralization this section will only explore the general concepts associated with the topic of auralization since it is safe to say that specific implementations of auralization techniques will be subject to changes and additions dictated by new technological advances and/or by market demands.

13.3.1 The Basic Auralization Process

The basic auralization process associated with an acoustical model is illustrated in Fig. 13-22.

The process starts with the reflectograms representing the *impulse response* (IR) of the model obtained at a specific receiver location for various frequencies. The reflectograms contain the information pertaining to the intensity and the direction of arrival of the reflections over a period of time that is deemed suitable to record the desired order and length of the reflections, and they are obtained from any of the methodologies presented in the modeling portion of this chapter. The reflectograms are then convolved—or mixed—with a dry (anechoic) recording of speech or music that can be played back under controlled conditions, using either headphones or loudspeakers, for the purpose of subjective evaluation.

13.3.2 Implementation

The energy reaching the listener is comprised of the direct sound, of the early reflections, and of the late reflections as shown in Fig. 13-23.

The direct sound is easily found and modeled accurately in the reflectogram since it represents the energy traveling from source to receiver in a direct line of sight. The only concern for the accurate auralization of the direct sound is to insure that the attenuation follows the inverse-distance spreading law as dictated by the source configuration and directivity. The early reflections are also obtained from the modeling phase but the reflectogram must be limited in length—or in the order of the reflections—because of computational constraints. The late portion of the reflectogram is usually modeled from a dense and random pattern of reflections with a smooth decay and a frequency content patterned after the reverberation time of the room estimated at various frequencies.

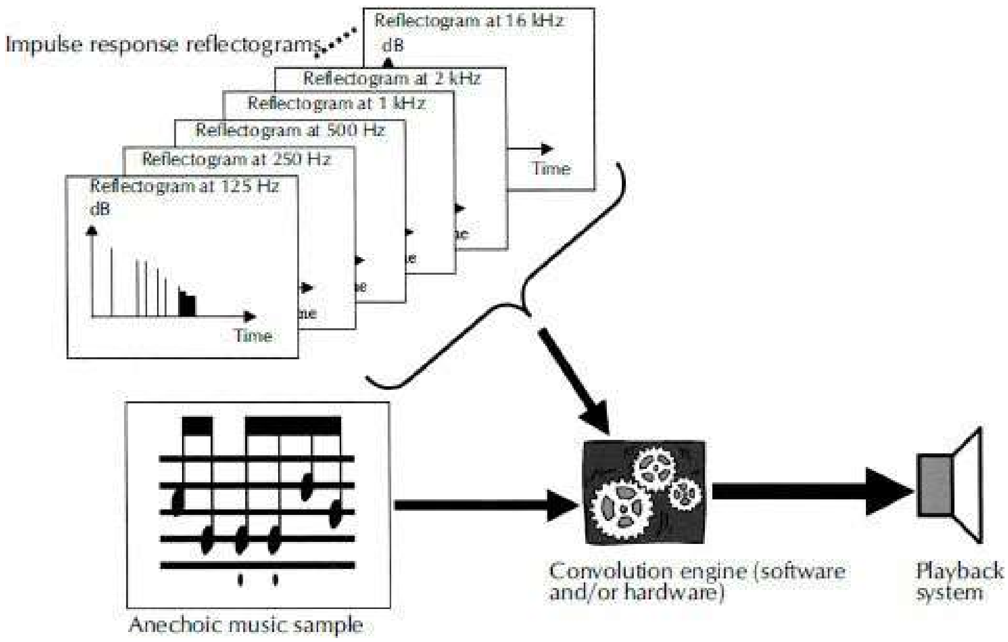


Figure 13-22. The basic auralization process.

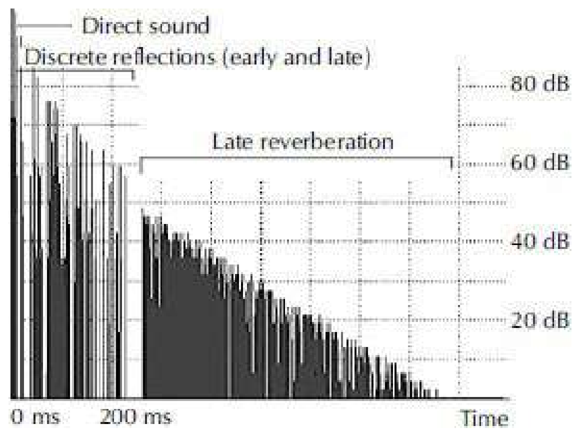


Figure 13-23. An example of a complete reflectogram at 1000Hz.

Since the reflectogram typically represents the impulse response at a single point (or within a small volume) in the modeled space, it must be modified in order to represent the binaural sound that would be reaching the eardrums of a listener by and at this point, two separate approaches are available.

13.3.2.1 Binaural Reproduction Using Loudspeakers

The impulse response is divided into its left and right components corresponding to the vertical left and right planes crossing the receiver location, and thus yielding the *binaural impulse response* (BIR) of the room for a listener at the receiver location. The anechoic signal is convolved separately for the left and the right channel, and the result is presented under anechoic and/or near field conditions to a listener using loudspeakers as shown in Fig. 13-24.

This technique has the advantage of being efficient from a computational standpoint since the process is limited to the separation of the IR into the BIR and the resulting convolution into left and right channels for the playback system. The drawback of the technique is that the playback requires a controlled environment where the listener has to maintain a fixed position with respect to the playback system and the crosstalk between the loudspeakers must be very small in order to yield the proper sense of spatial impression.

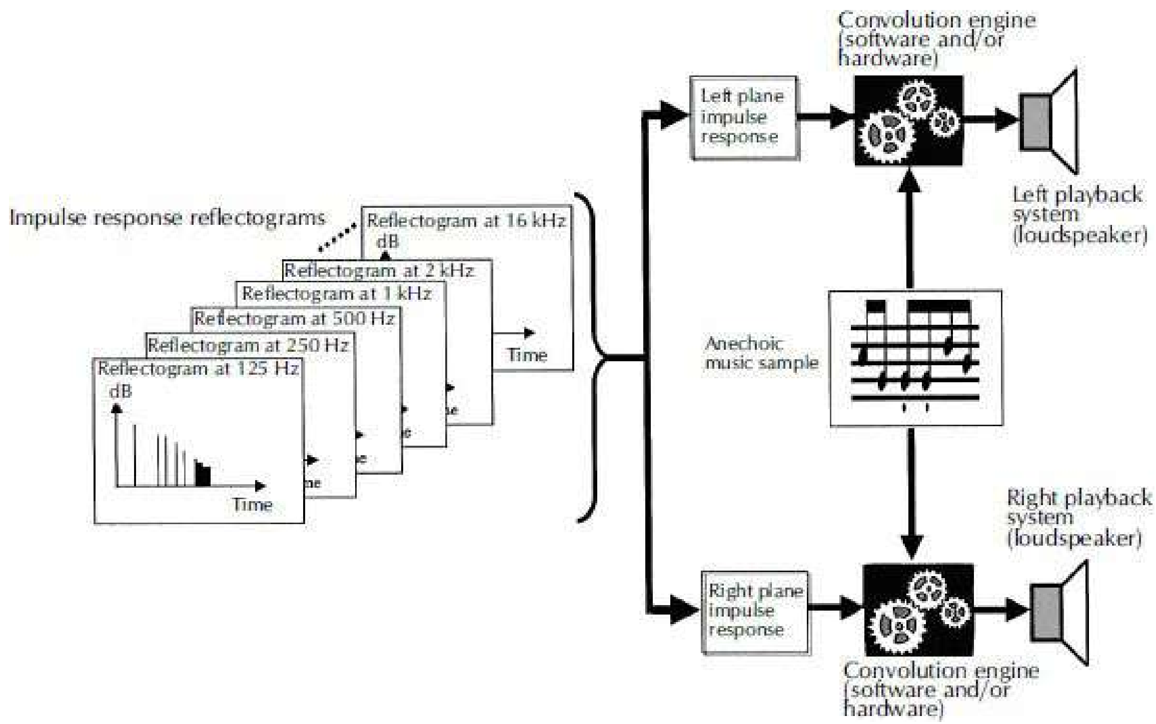


Figure 13-24. Auralization with a binaural impulse response and speaker presentation.

13.3.2.2 Binaural Reproduction Using Headphones

In this approach, the BIR is further modified by the application of *head-related transfer functions* (HRTF) that represent the effects that the head, torso, shoulders, and ears will have on the sound that reaches the eardrums of the listener. It has been shown[29,30] that these parameters have a drastic influence on the localization of the sound and on its overall subjective assessment. As shown in Fig. 13-25, the reproduction system must now use headphones since the effects of the body and head shape of the listener have already been taken into account. The advantage of this approach is that the playback system is very simple; good quality headphones are readily available and no special setup is required. The drawback is that the implementation of the modified BIR takes time due to the computational requirements for the application of the HRTF. It must also be noted that the HRTF may not accurately describe the specific parameters that a given listener experiences, although current HRTF research has yielded accurate composite data for a wide segment of the test results. Another issue of concern when using headphone reproduction is that the apparent source location will move with the listener's head movements, something that does not take place in the real world.

13.3.2.3 Multichannel Reproduction Using Loudspeakers

In this instance, the impulse response of the room is divided into components that correspond to the general locations in space from where the reflections originate as shown in Fig. 13-26. Various systems have been developed[31,32] throughout the years ranging from just a few speakers to hundreds of units driven by dozens of separate channels. The advantage of the technique is that the system relies on the listener's own HRTF while also allowing for head tracking effects. From the perspective of efficiency the approach can be implemented with minimal hardware and software since the reflections can be categorized in terms of their direction of arrival while the IR is being generated. The multichannel reproduction technique can actually be implemented from a physical scale model without the need for computer tools by using delay lines and an analog matrix system. The reproduction system is, of course, rather complicated since it requires substantial hardware and an anechoic environment. One manufacturer[33] has developed a system that relies on a near-field presentation of the auralized material to the listener using speakers that can yield a full-range response. In this context, the constraints associated with the anechoic environment are eliminated; however, the relative positioning of the listener with respect to the playback system is essential in order to achieve an accurate presentation of the results.

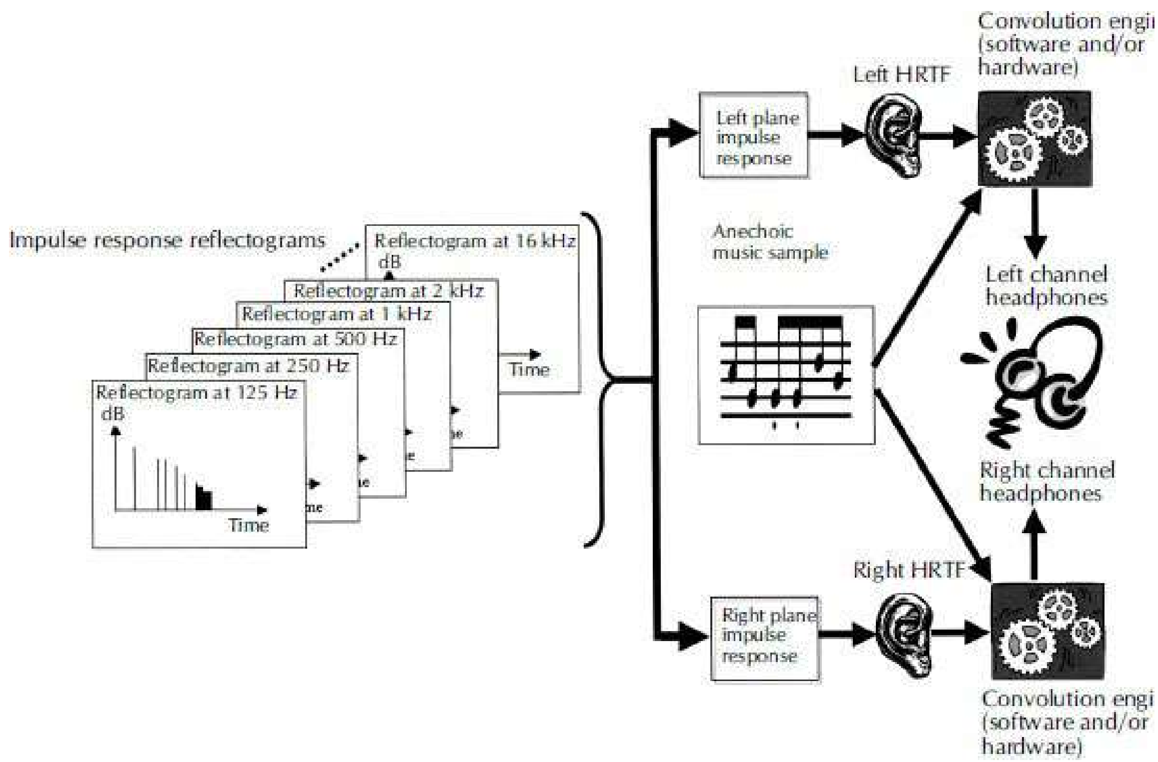


Figure 13-25. Auralization with a HRTF binaural impulse response and headphone presentation.

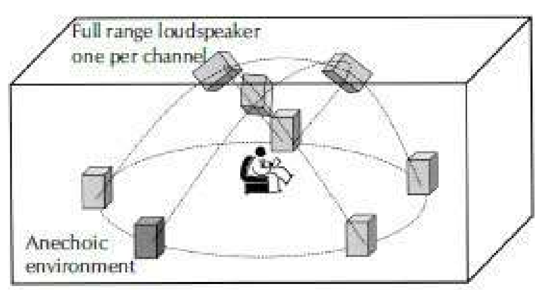


Figure 13-26. Speaker arrangement for multichannel presentation. Adapted from Reference 29.

13.3.3 Real-Time Auralization and Virtual Reality

A real-time auralization system allows the user to actually move within the room and to hear the resulting changes in the sound as they actually happen. This approach requires the near-instantaneous computation of the impulse response so that all parameters pertaining to the direct sound and to the reflections can be computed. In a recent implementation[34] the space is modeled using an enhanced image method approach in which a fast ray-tracing preprocessing step is taken to check the visibility of the reflections at the receiver location. The air absorption and the properties of the surface materials are modeled using efficient digital filters, and the late reverberation is described using techniques that give a smooth and dense reflection pattern that follows the statistical behavior of sound in a bounded space. The technique yields a *parametric room impulse response* (PRIR) in which a combination of real-time and nonreal-time processes performs a modeling of the physical parameters that define the space. A diagram of the modeling and auralization process of this system is presented in Fig. 13-27.

In this approach, known as *dynamic auralization*, the presentation of the sound field can be done either via binaural headphones or by multi-channel speaker techniques and the auralization parameters must be updated at a fast rate (typically more than ten times per second) in order for the rendering to be of high quality. The impulse response that is used for the convolutions can be a combination of an accurate set of binaural responses (that map head-tracking movements) to account for the early portion of the reflections with a simpler static impulse response that provides the foundation for the calculation of the late part of the sound field. This approach is moderately efficient in terms of computational time and memory consumption and recent developments[35] have been aimed at making use of an efficient means to process the impulse response of a space. Using an approach known as Ambisonics B-format[36] the sound information is encoded into four separate channels labeled W, X, Y and Z. The W channel would be equivalent to the mono output from an omnidirectional microphone while the X, Y and Z channels are the directional components of the sound in front-back (X), left-right (Y), and up-down (Z) directions. This allows a single B-format file to be stored for each location to account for all head motions at this specific location and to produce a realistic and fast auralization as the user can move from one receiver location to the other and experience a near-seamless simulation even while turning his/her head in the virtual model.

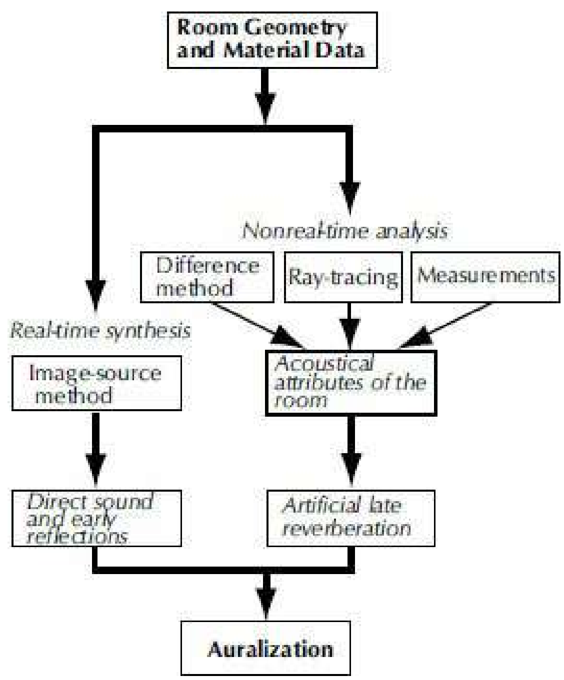


Figure 13-27. A real-time interactive modeling and auralization system. Adapted from Reference 32.

13.4 Conclusion

Acoustical modeling and auralization are topics of ongoing research and development. Originally planned for the evaluation of large rooms, the techniques have also been used in small spaces,[37] in outdoor noise propagation studies, and in complex indoor environments. One is now witnessing the standard use of these representation tools in a wide range of applications aimed at assessing complex acoustical quantifiers associated with both sound and with noise. Even simple digital processing systems such as those offered as plug-ins for audio workstations can be used to illustrate the effect of frequency-dependent transmission loss from various materials using simple equalization and level settings corresponding to octave or third-octave band reduction data.

Further work is needed in the representation and modeling of complicated sources such as musical instruments, automobiles, trains, and other forms of transportation; work is also ongoing in the definition of materials and surfaces so that the effect of vibrations and stiffness is accounted for. Still, the models are rapidly becoming both very accurate and very efficient and they are demonstrating their adequacy at illustrating the complicated issues that are associated with the propagation of acoustic waves and with the perception of sound, both wanted and unwanted.

References

1. Leonid I. Makrinenko, "*Acoustics of Auditoriums in Public Buildings*." First Publication: 1986 (Russian). Acoustical Society of America (1994) Translated by R.S. Ratner. J.S. Bradley, *Editor*.

2. Allan D. Pierce, *Acoustics. "An Introduction to Its Physical Principles and Applications."* Acoustical Society of America, 1991, 2nd printing.

3. J. B. Allen, and D. A. Berkley, “Image Method for Efficiently Simulating Small-Room Acoustics.” *Journal of the Acoustical Society of America*, 65 (4). 943–950, 1979.

4. J. Barish, “Extension of the Image Method to Arbitrary Polyhedra.” *Journal of the Acoustical Society of America*, 75 (6). 1827–1836, 1984.

5. M. Vorländer, “Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image Source Algorithm.” *Journal of the Acoustical Society of America*, 86 (1). 172–178, 1989.

6. I. A. Drumm, and Lam, Y. M. “The Adaptive Beam Tracing Algorithm.” *Journal of the Acoustical Society of America*, 107 (3). 1405–1412, 2000.

7. P. Min, and T. Funkhouser, “Priority-Driven Acoustic Modeling for Virtual Environments.” *Eurographics* 2000, 19 (3), 2000.

8. Bengt-Inge Dalenbäck, “The rediscovery of diffuse reflection in room acoustics prediction.” A presentation to the 144th Meeting of the Acoustical Society of America, Cancun, 2002.

9. J. S. Bradley, “*Architectural research today*.”

10. E. Bahar, “Acoustic Scattering by Two-Dimensionally Rough Interfaces between Dissipative Acoustic Media. Full-Wave, Physical Acoustics, and Perturbation Solutions.” *Journal of the Acoustical Society of America*, 89 (1). 19–26, 1991.

11. J. Kane, *Boundary Elements Analysis*, Prentice Hall, 1994.

12. COMET/Acoustics, *User Document V.2.1*. Automated Analysis Corporation, 1994.

13. T. J. Cox, “Predicting the Scattering from Diffusers Using 2-D Boundary Element Methods.” *Journal of the Acoustical Society of America*, 96 (1).

874–878, 1994.

14. P. D'Antonio and T. Cox, "Two Decades of Sound Diffusor Design and Development. Part 2: Prediction, Measurements, and Characterization." *Journal of the Audio Engineering Society*, 46 (12) 1075–1091, 1998.

15. D. Botteldooren, "Finite-Difference Time-Domain Simulations of Low-Frequency Room Acoustics Problems." *Journal of the Acoustical Society of America*, 8 (6). 3302–3309, 1995.

16. N. Lalor, "Statistical Energy Analysis and Its Use as an NVH Analysis Tool." *Sound & Vibration*, 30 (1). 16–20, 1996.

17. CARA (Computer-Aided-Room-Acoustics) software, from www.cara.de, Version 2.2.

18. J. S. Bradley, "Sound Absorption of Gypsum Board Cavity Walls." *Journal of the Audio Engineering Society*, 45 (4), 253–259, 1997.

19. D. Cook, and D. Van Haverbeke, *"Trees and Shrubs for Noise Abatement."* Research Bulletin 246. The Forest Service, U.S. Department of Agriculture cooperating with the University of Nebraska College of Agriculture, 1971.

20. W. Probst, and B. Huber, "Calculation and Assessment of Traffic Noise Exposure." *Sound and Vibration*, 07/00 pp.16–20, 2000.

21. BASTIAN software, from DataKustic, Germany.

22. P. Economou, and P. Charalampous, "A comparison of ISO 9613 and advanced calculation methods using Olive Tree Lab-Terrain, an outdoor sound propagation software application: Predictions versus experimental results," *Meeting on Environmental Noise Propagation - Definitions, measuring and control aspects*. Royal Society of London (2012).

23. M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization — An

Overview." *Journal of the Audio Engineering Society*, 41 (11): 861–875, 1993.

24. Adapted from reference 1, Chapter 4, section 4, pp. 106, endnote [43].

25. CATT-Acoustics, from B.I. Dalenbäck.

26. ODEON, from Ørsted DTU.

27. EARS, from ADA (Acoustic Design Ahnert).

28. DIVA, from the DIVA Group.

29. F. L. Wightman, and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *Journal of the Audio Engineering Society*, 85 (2). 858–867, 1989.

30. F. L. Wightman, and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychological Validation." *Journal of the Audio Engineering Society*, 85 (2). 868–878, 1989.

31. M. Camras, "Approach to Recreating a Sound Field." *Journal of the Acoustical Society of America*, 43 (6). 1425–1431, 1968.

32. K. Ono, S. Komiyama, and K. Nakabayashi, "A Method of Reproducing Concert Hall Sounds by Loudspeaker Walls." *Journal of the Acoustical Society of America*, 46 (11). 988–995, 1998.

33. Bose Corporation, Auditioner® System.

34. L. Savioja, *Modeling Techniques for Virtual Acoustics*. Ph.D. Dissertation, 2000.

35. Bengt-Inge Dalenbäck and M. Strömberg, "Real time walkthrough auralization – The first year." Proceedings of the Institute of Acoustics,

28 (2), 2007.

36. From Wikipedia reference on Ambisonics.

37. E. Granier, M. Kleiner, B. I. Dalenbäck, and Svensson, P. “Experimental Auralization of Car Audio Installations.” *Journal of the Audio Engineering Society*, 44 (10): 835–849, 1996.

* Model (mäd’l) n. [Fr. modèle < It modello] **1**. a) a small copy or imitation of an existing object made to scale b) a preliminary representation of something, serving as the plan from which the final, usually larger, object is to be constructed c) archetype d) a hypothetical or stylized representation e) a generalized, hypothetical description, often based on an analogy, used in analyzing or explaining something.

From *Webster New World College Dictionary*. 4th Edition (2000).

Part 3

Electronic Components

Chapter 14

Resistors, Capacitors, and Inductors

by Glen Ballou

14.1 Resistors

Resistance is associated with the phenomenon of energy dissipation. In its simplest form, it is a measure of the opposition to the flow of current by a piece of electric material. Resistance dissipates energy in the form of heat; the best conductors have low resistance and produce little heat, whereas the poorest conductors have high resistance and produce the most heat. For example, if a current of 10A flowed through a resistance of 1Ω, the heat would be

100W. If the same current flowed through 100Ω, the heat would be 10,000W, which is found with the equation

$$P = I^2R = \frac{V^2}{R} \tag{14-1}$$

where,

P is the power in W,

I is the current in A,

R is the resistance in Ω.

Changing the voltage, while holding the resistance constant, changes the power by the square of the voltage. For instance, a voltage change from 10V to 12V increases the power 44%. Changing the voltage from 10V to 20V increases the power 400%.

Changing the current while holding the resistance constant has the same effect as a voltage change. Changing the current from 1A to 1.2A increases the power 44%, whereas changing from 1A to 2A increases the power 400%.

Changing the resistance while holding the voltage constant changes the power linearly. If the resistance is decreased from 1kΩ to 800Ω and the voltage remains the same, the power will increase 20%. If the resistance is increased from 500Ω to 1kΩ, the power will decrease 50%. Note that an increase in resistance causes a decrease in power.

Changing the resistance while holding the current constant is also a linear power change. In this example, increasing the resistance from 1kΩ to 1.2kΩ increases the power 20%, whereas increasing the resistance from 1kΩ to 2kΩ increases the power 100%.

It is important in sizing resistors to take into account changes in voltage or current. If the resistor remains constant and voltage is increased, current also increases linearly. This is determined by using Ohm's Law, Eq. 14-1.

In a pure resistance—i.e. one without inductance or capacitance—the voltage and current phase relationship remains the same, so the voltage drop across the resistor is

$$V = IR \tag{14-2}$$

where,

V is the voltage in V,

I is the current in A,

R is the resistance in Ω.

and the current through the resistor is

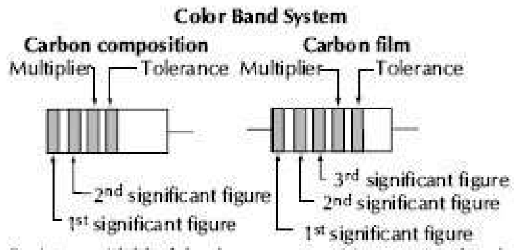
$$I = \frac{V}{R} \tag{14-3}$$

Resistors can be fixed or variable, have tolerances from 0.5% to 20%, and power ranges from 0.1W to hundreds of watts.

14.1.1 Resistor Characteristics

Resistors will change value as a result of applied voltage, power, ambient temperature, frequency change, mechanical shock, or humidity.

The values of the resistor are either printed on the resistor, as in power resistors, or are color coded on the resistor, Fig. 14-1.



Resistors with black body are composition, noninsulated.
Resistors with colored body are composition, insulated.
Wirewound resistors have the 1st color band double width.

Color	Digit	Multiplier	Tolerance	Failure Rate
Black	0	1	–	–
Brown	1	10	±1%	1.0
Red	2	100	±2%	0.1
Orange	3	1000	±3%	0.01
Yellow	4	10,000	±4%	0.001
Green	5	100,000	–	–
Blue	6	1,000,000	–	–
Violet	7	10,000,000	–	–
Gray	8	100,000,000	–	–
White	9	–	–	Solderable*
Gold		0.1	±5%	–
Silver		0.01	±10%	–
No Color			±20%	

Figure 14-1. Color codes for resistors.

Voltage Coefficient. The *voltage coefficient* is the rate of change of resistance due to an applied voltage, given in percent parts per million per volt (%ppm/V). For most resistors the voltage coefficient is negative—that is, the resistance decreases as the voltage increases. The voltage coefficient of very high valued carbon-film resistors is rather large and for wirewound resistors is usually negligible. *Varistors* are resistive devices designed to have a large voltage coefficient.

Temperature Coefficient of Resistance. The *temperature coefficient of resistance* (*TCR*) is the rate of change in resistance with ambient temperature, usually stated as parts per million per degree Celsius (ppm/°C). Many types of resistors increase in value as the temperature increases, while others, particularly hot-molded carbon types, have a maximum or minimum in their resistance curves that gives a zero temperature coefficient at some temperature. Metal film and wirewound types have temperature coefficient values of less than 100ppm/°C. Thermistors are resistance devices designed to have a large temperature coefficient. The percent temperature coefficient of resistance is

$$TCR = \frac{(R-r)100}{(T_R - T_T)R} \qquad (14\text{-}4)$$

where,

TCR is the temperature coefficient in percent per °C,

R is the resistance at reference temperature,

r is the resistance at test temperature,

T_R is the reference temperature in °C,

T_T is the test temperature in °C.

It is better to operate critical resistors with a limited temperature rise.

Noise. *Noise* is an unwanted voltage fluctuation generated within the resistor. The total noise of a resistor always includes Johnson noise, which depends only on the resistance value and the temperature of the resistance element. Depending on the type of element and its construction, total noise may also include noise caused by current flow and by cracked bodies and loose end caps or leads. For adjustable resistors, noise is also caused by the jumping of the contact over turns of wire and by an imperfect electrical path between the contact and resistance element.

Hot-Spot Temperature. The *hot-spot temperature* is the maximum temperature measured on the resistor due to both internal heating and the ambient operating temperature. The maximum allowable hot-spot temperature is predicated on the thermal limits of the materials and the resistor design. The maximum hot-spot temperature may not be exceeded under normal operating conditions, so the wattage rating of the resistor must be lowered if it is operated at an ambient temperature higher than that at which the wattage rating was established. At zero dissipation, the maximum ambient temperature around the resistor may be its maximum hot-spot temperature. The ambient temperature for a resistor is affected by surrounding heat-producing devices. Resistors stacked together do not experience the actual ambient temperature surrounding the outside of the stack except under forced cooling conditions.

Carbon resistors should, at most, be warm to touch, 40°C (140°F), while wirewound or ceramic resistors are designed to operate at temperatures up to 140°C (284°F). Wherever power is dissipated, it is imperative that adequate ventilation is provided to eliminate thermal destruction of the resistor and surrounding components.

Power Coefficient. The *power coefficient* is the product of the temperature coefficient of resistance and the temperature rise per watt. It is given in percent per watt (%/W), and is the change in value resulting from applied power.

ac Resistance. The *ac resistance* value changes with frequency because of the inherent inductance and capacitance of the resistor plus the skin effect, eddy current losses, and dielectric loss.

Ambient Temperature Effect. When operating a resistor in free air at high ambient temperature, the power capabilities must be derated, Fig. 14-2. Free air is operation of a resistor suspended by its terminals in free space and still air with a minimum clearance of one foot in all directions to the nearest object.

Grouping. Mounting a number of resistors in close proximity can cause excessive temperature rise requiring derating the power capabilities, Fig. 14-3. The curves are for operation at maximum permissible hot spot temperature with spacing between the closest points of the resistors. Derating could be less if operated at less than permissible hot spot temperature.

Enclosure. Enclosures create a rise in temperature due to the surface area, size, shape, orientation, thickness, material and ventilation. Fig. 14-4 indicates the effects on a resistor enclosed in an unpainted steel sheet metal box, 0.32 in thick without vents. Determining the derating is often by trial and error.

Forced Air Cooling. Resistors and components can be operated at higher than rated wattage with forced air cooling, Fig. 14-5. The volume of cooling air required to keep the resistor temperature within limits can be found with the equation

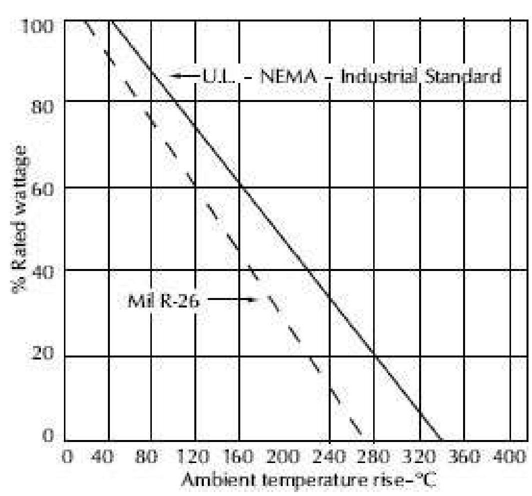


Figure 14-2. Resistors derating for elevated ambient temperature.

Courtesy Ohmite Mfg. Co.

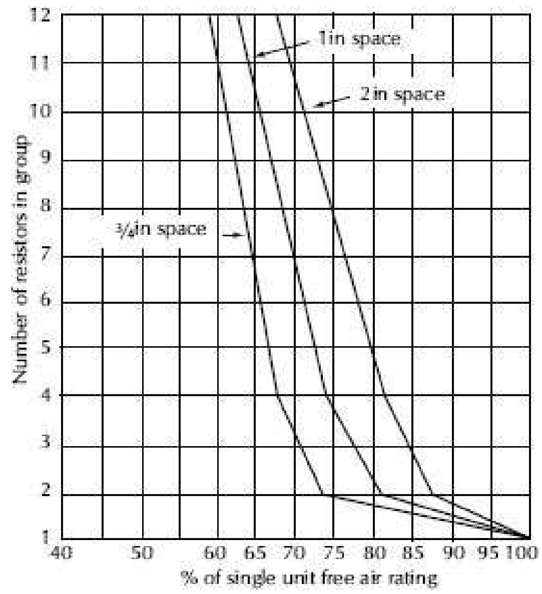


Figure 14-3. Power derating for grouping resistors. Courtesy Ohmite Mfg. Co.

$$\textit{Volume of air} = \frac{3170}{\Delta T} W \qquad (14\text{-}5)$$

where,

Volume of air is in ft^3/min,

ΔT is the permissible temperature rise in °F,

W is the power dissipated inside the enclosure in kW.

Air density at high altitudes causes less heat to be dissipated by convection so more forced air would be required.

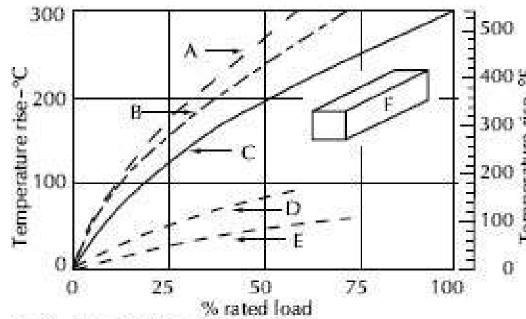


Figure 14-4. Effect of the size of an enclosure on a 500W 0.75 in × 6.5in resistor. Courtesy Ohmite Mfg. Co.

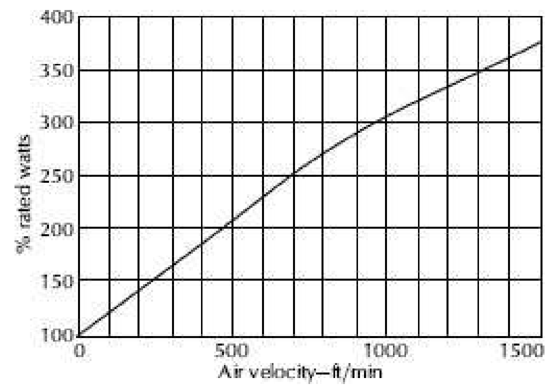


Figure 14-5. Percent of free air rating for a typical resistor cooled by forced air. Courtesy Ohmite Mfg. Co.

Pulse Operation. A resistor can usually be operated with a higher power in the pulse mode than in a continuous duty cycle. The actual increase allowed depends on the type of resistor. Fig. 14-6 is the percent of continuous duty rating for pulse operation for a wirewound resistor. Fig. 14-7 is the percent of continuous duty rating for pulse operation for typical NEMA duty cycles.

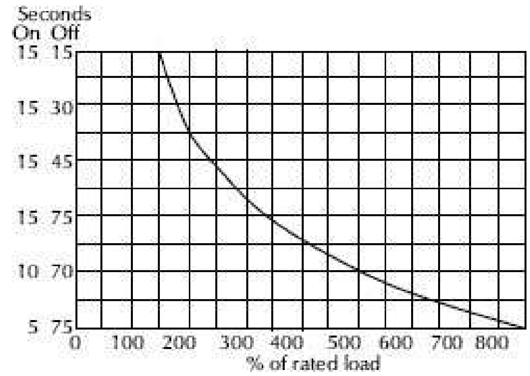


Figure 14-6. Effect of pulse operation on wirewound resistors. Courtesy Ohmite Mfg. Co.

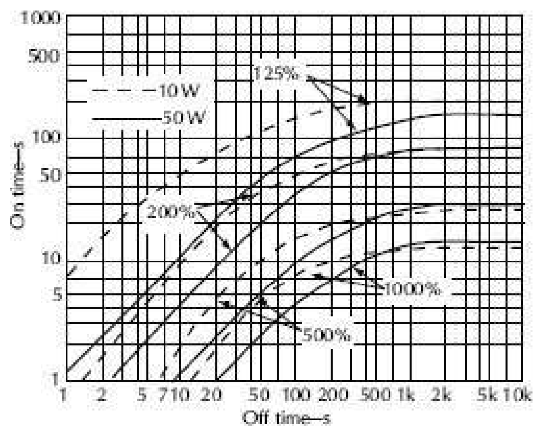


Figure 14-7. Percent of continuous duty rating for pulse operation of small and medium size vitreous enameled resistors. Courtesy Ohmite Mfg. Co.

14.1.2 Combining Resistors

Resistors can be combined in series or parallel or series/parallel.

Resistors in series. The total resistance of resistors connected in series is the summation of the resistors.

$$R_T = R_1 + R_2 + \ldots R_n \qquad (14\text{-}6)$$

The total resistance is always greater than the largest resistor.

Resistors in Parallel. The total resistance of resistors in parallel is

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \ldots \frac{1}{R_n}} \qquad (14\text{-}7)$$

If two resistors are in parallel use

$$R_T = \frac{R_1 \times R_2}{R_1 + R_2} \qquad (14\text{-}8)$$

When all of the resistors are equal, divide the value of one resistor by the number of resistors to determine the total resistance. The total resistance is always less than the smallest resistor.

To determine the value of one of the resistors when two are in parallel and the total resistance and one resistor in known, use

$$R_2 = \frac{R_T \times R_1}{R_1 - R_T} \qquad (14\text{-}9)$$

14.1.3 Types of Resistors

Every material that conducts electrical current has resistivity, which is defined as the resistance of a material to electric current. Resistivity is normally defined as the resistance, in ohms, of a 1cm per side cube of the material measured from one surface of the cube to the opposite surface. The measurement is stated in ohms per centimeter cubed (Ω/cm^3). The inverse of resistivity is conductivity. Good conductors have low resistivity, and good insulators have high resistivity. Resistivity is important because it shows the difference between materials and their opposition to current, making it possible for resistor manufacturers to offer products with the same resistance but differing electrical, physical, mechanical, or thermal features. Table 14-1 gives the resistivity of various materials:

Table 14-1. Resistivity of various materials

Material	Resistivity
Aluminum	0.0000028
Copper	0.0000017
Nichrome	0.0001080
Carbon (varies)	0.0001850
Ceramic (typical)	100,000,000,000,000 or (10^{14})

Carbon-Composition Resistors. Carbon-composition resistors are the least expensive resistors and are widely used in circuits that are not critical to input noise and do not require tolerances better than ±5%.

The carbon-composition, hot-molded version is basically the same product it was more than 50 years ago. Both the hot- and cold-molded versions are made from a mixture of carbon and a clay binder. In some versions, the composition is applied to a ceramic core or armature, while in the inexpensive version, the composition is a monolithic rigid structure. Carbon-composition resistors may be from 1Ω to many megohms and 0.1 to 4W. The most common power rating is ¼W and ½W with resistance values from 2Ω to 22MΩ.

Carbon-composition resistors can withstand higher surge currents than carbon-film resistors. Resistance values, however, are subject to change upon absorption of moisture and increase rapidly at temperatures much above 60°C (140°F). Noise also becomes a factor when carbon-composition resistors are used in audio and communication applications. A carbon-core resistor, for example, generates electrical noise that can reduce the readability of a signal or even mask it completely.

Carbon-Film Resistors. Carbon-film resistors are leaded ceramic cores with thin films of carbon applied. Carbon film resistors offer closer tolerances and better temperature coefficients than carbon composition resistors. Most characteristics are virtually identical for many general purpose, noncritical applications where high reliability, surge currents, or noise are not crucial factors.

Metal Film Resistors. Metal film resistors are discrete devices formed by depositing metal or metal oxide films on an insulated core. The metals are usually either nichrome sputtered on ceramic or tin oxide on ceramic or glass. Another method of production is to screen or paint powdered metal and powdered glass that is mixed in an ink or pastelike substance on a porous ceramic substrate. Firing or heating in an oven bonds the materials together. This type of resistor technology is called *cermet technology*.

Metal film resistors are most common in the 10Ω to 1MΩ range and ⅛W to 1W with tolerances of ±1%.

The *TCR* is in the ±100ppm/°C range for all three technologies. Yet there are subtle differences:

- Cermet covers a wider resistance range and handles higher power than nichrome deposition.
- Nichrome is generally preferred over tin oxide in the upper and lower resistance ranges and can provide *TCR*s that are lower than 50ppm/°C.
- Tin oxide is better able to stand higher power dissipation than nichrome.

Wirewound Resistors. *Wirewound resistors* have resistive wire wound on a central ceramic core. One of the oldest technologies, wirewounds provide the best known characteristics of high temperature stability and power handling ability. Nichrome, Manganin, and Evanohm are the three most widely used wires for wire-wound resistors.

Wirewound resistors are usually in the 0.1Ω–250kΩ range. Tolerance is ±2% and *TCR* is ±10ppm/°C.

Wirewound resistors are generally classed as power or instrument-grade products. Power wirewounds, capable of handling as much as 1500W, are wound from uninsulated coarse wire to provide better heat dissipation. Common power ratings are 1.5W, 3W, 5W, 8W, 10W, 20W, 25W, 50W, 100W, and 200W.

Instrument-grade precision wirewound resistors are made from long lengths of finely insulated wire. After winding, they are usually coated with a ceramic material.

All wirewound resistors are classed as air-core inductors and the inductive reactance alters the high frequency resistive value. This problem is directly proportional with frequency. Special windings are useful to cancel reactance at audio frequencies. Because of the severity of the problem, these resistors cannot be used at high frequencies.

Noninductive Resistors. Non-inductive resistors are used for high frequency applications. This is accomplished by utilizing the Ayrton-Perry type of wiring, i.e. two windings connected in parallel and wound in opposite directions. This keeps the inductance and distributed capacitance at a minimum. Table 14-2 is a comparison of MEMCOR-TRUOHM type FR10, FR50, VL3 and VL5 resistors.

Resistor Networks. With the advent of printed circuit boards and integrated circuits, *resistor networks* became popular. The resistive network may be mounted in a single-in-line package (SIP) socket or a dual-in-line package (DIP) socket—the same as the ones used for integrated circuits. The most common resistor network has 14 or 16 pins and includes 7 or 8 individual resistors or 12 to 15 resistors with a common terminal. In most resistor networks the value of the resistors are the same. Networks may also have special value resistors and interconnections for a specific use, as shown in Fig. 14-8.

The individual resistors in a thick-film network can have a resistance value ranging from 10Ω to 2.2MΩ and are normally rated at 0.125W per resistor. They have normal tolerances of ±2% or better and a temperature coefficient of resistance ±100ppm/°C from -55°C to +125°C (-67°F to +257°F).

Thin-film resistors are almost always specialized units and are packaged as DIPs or flatpacks. (Flatpacks are soldered into the circuit.) Thin-film networks use nickel chromium, tantalum nitride, and chromium cobalt vacuum depositions.

Table 14-2. Inductance Comparison of Standard and Non-Inductive Windings.

		Approximate Frequency Effect		
		Stock inductive winding	Non-inductive winding	
Type	Resistance (Ω)	L_S (μH)	L_S (μH)	C_P ($\mu\mu$F)
FR10 (10W)	25	5.8	0.01	–
	100	11.0	0.16	–
	500	18.7	0.02	–
	1000	20.8	–	0.75
	5000	43.0	–	1.00
FR50 (50W)	25	6.8	0.05	–
	100	>100.0	0.40	–
	500	>100.0	0.31	–
	1000	>100.0	–	1.10
	5000	>100.0	–	1.93
VL3 (3W)	25	1.2	0.02	–
	100	1.6	0.07	–
	500	4.9	–	0.47
	1000	4.5	–	0.70
	5000	3.0	–	1.00
VL5 (5W)	25	2.5	0.08	–
	100	5.6	0.14	–
	500	6.4	–	0.03
	1000	16.7	–	0.65
	5000	37.0	–	0.95

Courtesy Ohmite Mfg. Co.

Variable Resistors. *Variable resistors* are ones whose value changes with light, temperature, or voltage or through mechanical means.

Photocells (Light-Sensitive Resistors). *Photocells* are used as off–on devices when a light beam is broken or as audio pickups for optical film tracks. In the latter, the sound track is either a variable density or variable area. Whichever, the film is between a focused light source and the photocell. As the light intensity on the photocell varies, the resistance varies.

Photocells are rated by specifying their resistance at low and high light levels. These typically vary from 600Ω to 110kΩ (bright), and from 100kΩ to 200MΩ (dark). Photocell power dissipation is between 0.005W and 0.75W.

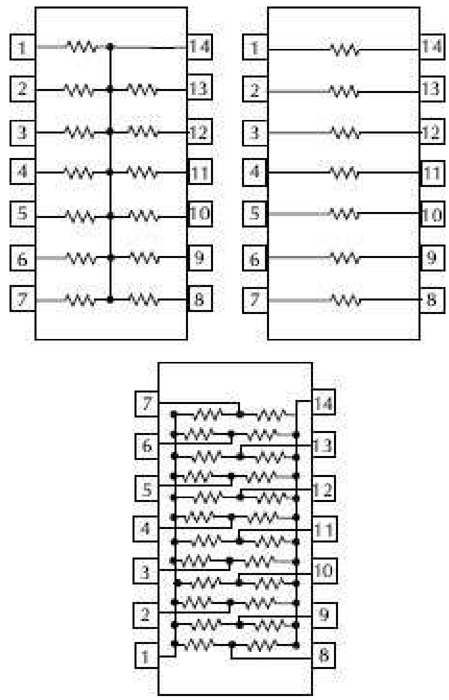


Figure 14-8. Various types of resistor networks.

Thermistors. *Thermistors*, thermal-sensitive resistors, may increase or decrease their resistance as temperature rises. If the coefficient of resistance is negative (NTC), the resistance decreases as the temperature increases; if positive, the resistance increases with an increase in temperature. Thermistors are specified by how their resistance changes for a 1°C change in temperature. They are also rated by their resistance at 25°C and by the ratio of resistance at 0°C and 50°C. Values vary from 2.5Ω to 1MΩ at room temperature with power ratings from 0.1 to 1W.[1]

Thermistors are normally used as temperature-sensing devices or transducers. When used with a transistor, they can control transistor current with a change in temperature. As the transistor heats up, the emitter-to-collector current increases. If the power supply voltage remains the same, the power dissipation in the transistor increases until it destroys itself through thermal runaway. The change in resistance due to temperature change of the thermistor placed in the base circuit of a transistor can be used to reduce base voltage, reducing the transistor emitter to collector current. By properly matching the temperature coefficients of the two devices, the output current of the transistor can be held fairly constant with temperature change.

Varistors. *Varistors* (voltage-sensitive resistors) are voltage-dependent, nonlinear resistors which have symmetrical, sharp breakdown characteristics similar to back-to-back Zener diodes. They are designed for transient suppression in electrical circuits. The transients can result from the sudden release of previously stored energy—i.e., electromagnetic pulse (*EMP*)—or from extraneous sources beyond the control of the circuit designer, such as lightning surges. Certain semiconductors are most susceptible to transients. LSI and VLSI circuits, which may have as many as 20,000 components in a 6.35mm × 6.35mm area (0.25in × 0.25in), have damage thresholds below 100μJ.

The *varistor* is mostly used to protect equipment from power-line surges by limiting the peak voltage across its terminals to a certain value. Above this voltage, the resistance drops, which in turn tends to reduce the terminal voltage. Voltage-variable resistors or varistors are specified by power dissipation (0.25 to 1.5W) and peak voltage (30 to 300V).

Thermocouples. While not truly a resistor, thermocouples are used for temperature measurement. They operate via the Seebeck

Effect which states that two dissimilar metals joined together at one end produce a voltage at the open ends that varies as the temperature at the junction varies. The voltage output increases as the temperature increases. Thermocouples are rugged, accurate, and have a wide temperature range. They don't require an excitation source and are highly responsive.[1]

Thermocouples are "tip" sensitive, measuring temperatures at a very small point of reference. Their output is quite nonlinear, which means they require external linearization in a form of cold-junction compensation. Cold junction compensation is crucial if accurate temperature measurements are needed. Thermocouple output voltage is quite low, in the tens of hundreds of microvolts, requiring careful wiring layout techniques to minimize noise and drift. One way to reduce noise is to place resistors in series with the thermocouple and a capacitor across the thermocouple leads to form a filter.

Never use copper wire to connect a thermocouple to the measuring device as that constitutes another thermocouple.

Resistance Temperature Detectors. RTDs are very accurate and stable. Most are made of platinum wire wound around a small ceramic tube. They can be thermally shocked by going from 100° C to –195° C 50 times with a resulting error less than 0.02°C.[2]

RTDs feature a low resistance-value change to temperature (0.1Ω/1°C. RTDs can self heat, causing inaccurate readings, therefore the current through the unit should be kept to 1mA or less. Self heating can be controlled by using a 10% duty cycle rather than constant bias or by using an extremely low bias which can reduce the *SNR*. The connection leads may cause errors if they are long due to the wire resistance.

Potentiometers and Rheostats. The resistance of *potentiometers* (pots), and *rheostats* is varied by mechanically varying the size of the resistor. They are normally three terminal devices, two ends and one wiper, Fig. 14-9. By varying the position of the wiper, the resistance between either end and the wiper changes. Pots may be wirewound or nonwirewound. The nonwirewound resistors usually have either a carbon or a conductive plastic coating. Pots may be 300° single turn or multiple turn, the most common being 1080° three turn and 3600° ten turn.

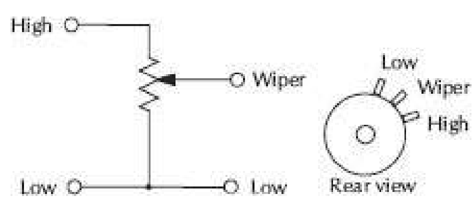


Figure 14-9. Three terminal potentiometer.

Wirewound pots offer *TCR*s of ±50ppm/°C and tolerances of ±5%. Resistive values are typically 10Ω to 100kΩ, with power ratings from 1W to 200W.

Carbon pots have *TCR*s of ± 400 ppm/°C to ±800ppm/°C and tolerances of ±20%. The resistive range spans 50Ω to 2MΩ, and power ratings are generally less than 0.5W.

Pots may be either linear or nonlinear, as shown in Fig. 14-10. The most common nonlinear pots are counterclockwise semilog and clockwise semilog. The counterclockwise semilog pot is also called an audio taper pot because when used as a volume control, it follows the human hearing equal loudness curve. If a linear pot is used as a simple volume control, only about the first 20% of the pot rotation would control the usable volume of the sound system. By using an audio taper pot as in Fig. 14-10 curve C2, the entire pot is used. Note there is only a 10%–20% change in resistance value between the common and wiper when the pot is 50% rotated.

Potentiometers are also produced with various taps that are often used in conjunction with *loudness* controls.

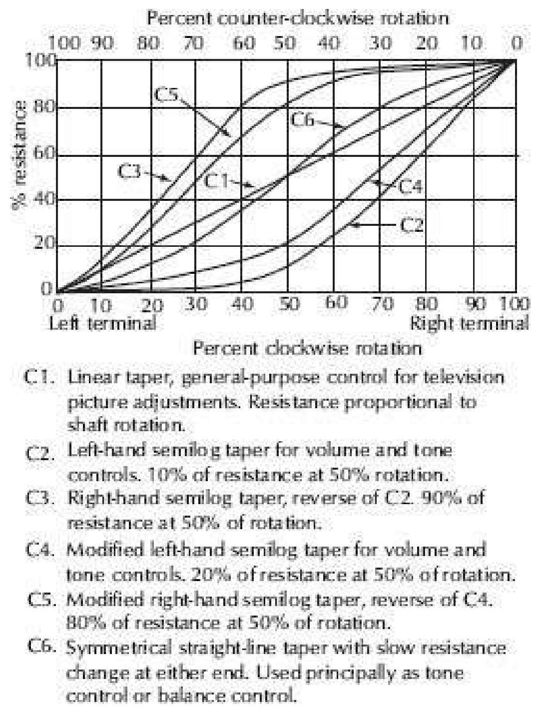


Figure 14-10. Tapers for six standard potentiometers in resistivity versus rotation.

Potentiometers also come in combinations of two or more units controlled by a single control shaft or controlled individually by concentric shafts. Switches with various contact configurations can also be assembled to single or ganged potentiometers and arranged for actuation during the first few degrees of shaft rotation.

A *wirewound potentiometer* is made by winding resistance wire around a thin insulated card, Fig. 14-11A. After winding, the card is formed into a circle and fitted around a form. The card may be tapered, Fig. 14-11B, to permit various rates of change of resistance as shown in Fig. 14-10. The wiper presses along the wire on the edge of the card.

Contact Resistance. Noisy potentiometers have been a problem that has plagued audio circuits for years. Although pots have become better in tolerance and construction, noise is still the culprit that forces pots to be replaced. Noise is usually caused by dirt or, in the case of wirewound potentiometers, oxidation. Many circuits have gone up in smoke because bias-adjusting resistors, which are wirewound for good *TCR*, oxidize and the contact resistance increases to a point where it is more than the value of the pot.

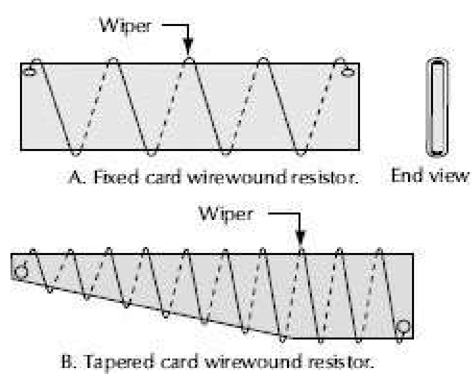


Figure 14-11. Construction of a wirewound resistor.

Sometimes the pot can be cleaned by spraying it with a contact cleaner or silicone and then vigorously rotating it. Usually, however, it is best to replace it because anything else is only temporary.

Any dc voltage present on the pot is also a source of noise. Such voltage is often produced by leaky coupling capacitors at the input connector or output circuit of the wiper, allowing dc voltage to appear at the wiper contact. If there is a resistance between the resistor and the wiper, the dc current flowing through the wiper contact to the output stage will create a voltage drop. Because the wiper is moving, the contact resistance constantly changes creating what looks like a varying ac voltage. Using Fig. 14-12, the value at V_{Load}, whether ac or dc, can be calculated with Eqs. 14-10 and 14-11. If the wiper resistance is 0—i.e., a perfect pot—the output voltage V^{Load} is

$$V_{Load} = V_1\left(\frac{R_y}{R_1 + R_y}\right) \quad (14\text{-}10)$$

where,

$$R_y = \frac{R_2 R_{Load}}{R_2 + R_{Load}}$$

If a pot wiper has a high resistance, R_w, the output voltage ${}^{V}Load$ is

$$V_{Load} = V_w\left(\frac{R_{Load}}{R_w + R_{Load}}\right) \quad (14\text{-}11)$$

where,

$$V_w = V_1\left(\frac{R_2(R_w + R_{Load})}{R_2 + R_w + R_{Load}}\right).$$

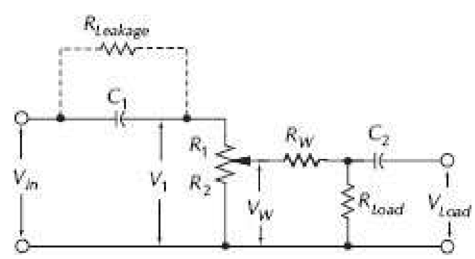


Figure 14-12. Effects of wiper noise on potentiometer output.

14.1.4 Measuring Resistance

Resistance can be measured with either a two wire or four wire method. When measuring high resistance, the two wire method is adequate. When measuring low resistance, or measuring a resistance a distance from the meter, four wire measurement is preferred.[1]

Digital multimeters (DMMs) typically employ the constant-current method to measure resistance, which sources a constant current (I_{SOUR}) to the device under test (DUT) and measures the voltage (V_{MEAS}). Resistance (R_{DUT}) of the DUT is then calculated and displayed using the known current and measured voltage $R_{DUT} = V_{MEAS}/I_{SOUR}$. Fig. 14-13 shows a simple diagram of the constant-current test.

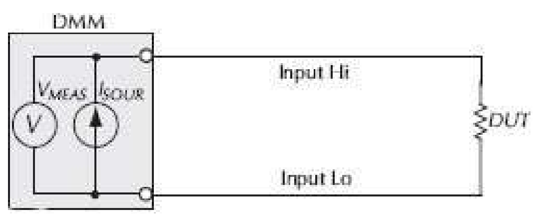


Figure 14-13. The constant-current method of resistance measurement, in a two-wire test configuration.

14.1.4.1 Two-Wire Resistance Measurements

Fig. 14-14 represents a two-wire resistance test configuration employing the constant current method.

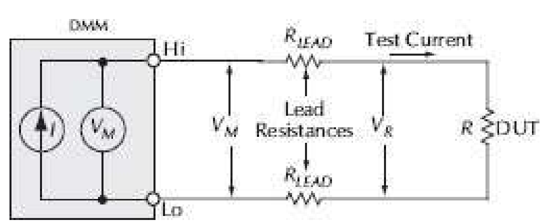


Figure 14-14. Two-wire resistance measurement schematic.

The main measurement issue with the two-wire method, as applied to low resistance measurements, is that the total lead resistance (R_{LEAD}) is added to the measurement. The test current I causes a small but significant voltage drop across the two lead resistances so the voltage V_M measured by the meter won't be exactly the same as the voltage V_R directly across the test resistance R so the measured resistance is

$$R_{MEAS} = V_M / I = R + (2 \times R_{LEAD}) \qquad (14\text{-}12)$$

Typical lead resistances range from 10mΩ to 1Ω, so it's very difficult to obtain accurate two-wire resistance measurements when the resistance under test is low. For example, using test leads with a 100mΩ combined resistance to perform a two-wire resistance measurement on a 500mΩ resistor will result in a 20% measurement error in addition to that of the instrument.

14.1.4.2 Four-Wire (Kelvin) Resistance Measurements

Due to the limitations of the two-wire method, a different approach is used for low resistance measurements that reduce the effect of test lead resistance. For measuring DUTs with very low resistances, the four-wire (Kelvin) connection shown in Fig. 14-15 is preferred because the voltage is measured at the DUT so the voltage drop in the test leads is eliminated.[1]

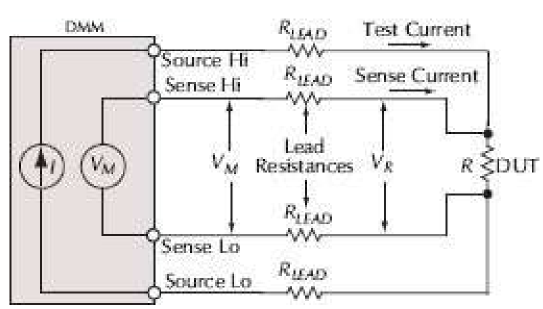


Figure 14-15. Four-wire resistance measurement configuration.

With this configuration, the test current I is forced through the test resistance R via one set of test leads, while the voltage V_M across the DUT is measured through a second set of leads (sense leads). The measured resistance is now

$$R_{MEAS} = V_M/I = V_R/I \qquad (14\text{-}13)$$

The voltage-sensing leads should be connected as close to the resistor under test as possible to avoid including part of the resistance of the test leads in the measurement. The voltage measured by the meter V_M is essentially the same as the voltage V_R across the resistance R, so the resistance value is much more accurate than when measured with the two-wire method.

14.2 Capacitors

Capacitors are used for both dc and ac applications. In dc circuits they are used to store and release energy such as filtering power supplies and for providing on demand, a single high voltage pulse of current.

In ac circuits capacitors are used to block dc, allowing only ac to pass, bypassing ac frequencies, or discriminating between higher and lower ac frequencies. In a circuit with a pure capacitor, the current will lead the voltage by 90°.

The value of a capacitor is normally written on the capacitor.

Where capacitors are connected in series with each other, the total capacitance is

$$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} \ldots + \frac{1}{C_n}} \tag{14-14}$$

and is always less than the value of the smallest capacitor.

When connected in parallel, the total capacitance is

$$C_T = C_1 + C_2 \ldots + C_n \tag{14-15}$$

and is always larger than the largest capacitor.

When a dc voltage is applied across a group of capacitors connected in series, the voltage drop across the combination is equal to the applied voltage. The drop across each individual capacitor is inversely proportional to its capacitance, and assuming each capacitor has an infinitely large effective shunt resistance, can be calculated by the equation

$$V_C = V_A\left(\frac{C_X}{C_T}\right) \tag{14-16}$$

where,

V_C is the voltage across the individual capacitor in the series (C_1, C_2, ... C_n) in V,

V_A is the applied voltage in V,

C_X is the capacitance of the individual capacitor under consideration in F,

C_T is the sum of all of the capacitors in series.

When used in an ac circuit, the capacitive reactance, or the impedance the capacitor injects into the circuit, is found with the equation

$$X_C = \frac{1}{2\pi f C} \qquad (14\text{-}17)$$

where,

X_C is the capacitive reactance in Ω,

f is the frequency in Hz,

C is the capacitance in F.

To determine the impedance of circuits with resistance, capacitance, and inductance, see section 14.4.

Capacitance is the concept of energy storage in an electric field. If a potential difference is found between two points, an electric field exists. The electric field is the result of the separation of unlike charges, therefore, the strength of the field will depend on the amounts of the charges and their separator. The amount of work necessary to move an additional charge from one point to the other will depend on the force required and therefore upon the amount of charge previously moved. In a capacitor, the charge is restricted to the area, shape, and spacing of the capacitor electrodes, sometimes known as *plates*, as well as the property of the material separating the plates.

When electrical current flows into a capacitor, a force is established between two parallel plates separated by a dielectric. This energy is stored and remains even after the input current flow ceases. Connecting a conductor across the capacitor provides a plate-to-plate path by which the charged capacitor can regain electron balance, that is, discharge its stored energy. This conductor can be a resistor, hard wire, or even air. The value of a parallel plate capacitor can be found with the equation

$$C = \frac{x\varepsilon[(N-1)A]}{d} \times 10^{-13} \qquad (14\text{-}18)$$

where,

C is the capacitance in F,

x is 0.0885 when *A* and *d* are in cm, and 0.225 when *A* and *d* are in in,

ε is the dielectric constant of the insulation,

N is the number of plates,

A is the area of the plates,

d is the spacing between the plates.

The work necessary to transport a unit charge from one plate to the other is

$$e = kg \qquad (14\text{-}19)$$

where,

e is the volts expressing energy per unit charge,

k is the proportionality factor between the work necessary to carry a unit charge between the two plates and the charge already transported and is equal to 1/C where *C* is the capacitance in F,

g is the coulombs of charge already transported.

The value of a capacitor can now be calculated from the equation

$$C = \frac{q}{e} \qquad (14\text{-}20)$$

where,

q is the charge in C,

e is found with Eq. 14-19.

The energy stored in a capacitor is found with the equation

$$W = \frac{CV^2}{2} \qquad (14\text{-}21)$$

where,

W is the energy in J,

C is the capacitance in F,

V is the applied voltage in V.

Dielectric Constant (K). The dielectric constant is the property of a given material that determines the amount of electrostatic energy that may be stored in that material per unit volume for a given voltage. The value of K expresses the ratio of a capacitor in a vacuum to one using a given dielectric. The K of air is 1 and is the reference unit employed for expressing K of other materials. If K of the capacitor is increased or decreased, the capacitance will increase or decrease respectively if other quantities and physical dimensions are kept constant. Table 14-3 is a listing of K for various materials.

Table 14-3. Comparison of Capacitor Dielectric Constants

Dielectric	K (Dielectric Constant)
Air or vacuum	1.0
Aluminum oxide	8.4
Bakelite	4.9
Ceramic	12.0–10,000
Glass	4.8–8.0
Mica	5.4–8.7
Mineral oil	2.2–2.3
Paper	2.0–6.0
Plastic	2.1–6.0

Plexiglass	2.8
Polyethylene	2.3
Polystyrene	2.6
Porcelain	5.1–5.9
Quartz	3.8–4.4
Silicone oil	2.7–2.8
Tantalum pentoxide	26.0

The dielectric constant of materials is generally affected by both temperature and frequency, except for quartz, Styrofoam, and Teflon, whose dielectric constants remain essentially constant. Small differences in the composition of a given material will also affect the dielectric constant.

Force. The equation for calculating the force of attraction between the two plates is

$$F = \frac{AV^2}{K(1504S)^2} \qquad (14\text{-}22)$$

where,

F is the attractive force in dyn,

A is the area of one plate in cm^2,

V is the potential energy difference in V,

K is the dielectric constant,

S is the separation between the plates in cm.

14.2.1 Time Constants

When a dc voltage is impressed across a capacitor, a time (*t*) is required to charge the capacitor to a voltage. This is determined with the equation

$t = RC$ (14-23)

where,

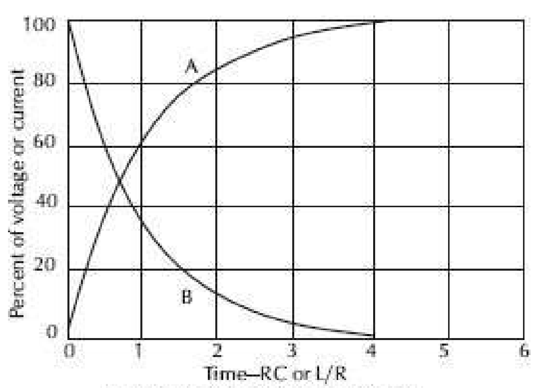
t is the time in s,

R is the resistance in Ω,

C is the capacitance in F.

In a circuit consisting of only resistance and capacitance, the time constant t is defined as the time it takes to charge the capacitor to 63.2% of the maximum voltage. During the next time constant, the capacitor is charged or the current builds up to 63.2% of the remaining difference of full value, or to 86.5% of the full value. Theoretically, the charge on a capacitor or the current through a coil can never actually reach 100% but is considered to be 100% after five time constants have passed. When the voltage is removed, the capacitor discharges and the current decays 63.2% per time constant to zero.

These two factors are shown graphically in Fig. 14-16. Curve A shows the voltage across a capacitor when charging. Curve B shows the capacitor voltage when discharging. It is the voltage across the resistor on charge or discharge.



A. Voltage across C when charging.
B. Voltage across C when discharging.

Figure 14-16. Universal time graph.

14.2.2 Network Transfer Function

Network transfer functions are the ratio of the output to input voltage (generally a complex number) for a given type of network containing resistive and reactive elements. The transfer functions for networks consisting of resistance and capacitance are given in Fig. 14-17. The expressions for the transfer functions of the networks are:

A is jω or $j2\pi f$,
B is RC,
C is R_1C_1,
D is R_2C_2,
n is a positive multiplier,
f is the frequency in Hz,
C is in F,
R is in Ω.

14.2.3 Characteristics of Capacitors

The operating characteristics of a capacitor determine what it was designed for and therefore where it is best used.

Capacitance (*C*). The *capacitance* of a capacitor is normally expressed in microfarads (μF or 10^{-6}F) or picofarads (pF or 10^{-12}F) with a stated accuracy or tolerance. Tolerance is expressed as plus or minus a certain percentage of the nominal or nameplate value. Another tolerance rating is *GMV* (guaranteed minimum value), sometimes referred to as *MRV* (minimum rated value). The capacitance will never be less than the marked value when used under specified operating conditions but the capacitance could be more than the named value.

Dielectric Absorption (*DA*). *Dielectric absorption* is a reluctance on the part of the dielectric to give up stored electrons when the capacitor is discharged. If a capacitor is discharged through a resistance, and the resistance is removed, the electrons that remained in the dielectric will reconvene on the electrode, causing a voltage to appear across the capacitor. This is also called *memory*.
When an ac signal, such as sound, with its high rate of attack is impressed across the capacitor, time is required for the capacitor to follow the signal because the free electrons in the dielectric move slowly. The result is compressed signal. The procedure for testing *DA* calls for a 5 minute capacitor charging time, a 5 second discharge, then a 1 minute open circuit, after which the recovery voltage is read. The percentage of *DA* is defined as the ratio of recovery to charging voltage times 100.

Dissipation Factor (*DF*). The *dissipation factor* is the ratio of the effective series resistance of a capacitor to its reactance at a specified frequency and is given in percent. It is also the reciprocal of Q. It is, therefore, a similar indication of power loss within the capacitor and, in general, should be as low as possible.

Equivalent Series Inductance (*ESL*). The *equivalent series inductance* can be useful or detrimental. It does reduce the high-frequency performance of the capacitor, however it can be used in conjunction with the capacitors capacitance to form a resonant circuit.

Equivalent Series Resistance (*ESR*). All capacitors have an *equivalent series resistance* expressed in ohms or milliohms. This loss comes from lead resistance, termination losses, and dissipation in the dielectric material.

Insulation Resistance. *Insulation resistance* is basically the resistance of the dielectric material, and determines the period of time a capacitor, once charged with a dc voltage, will hold its charge by a specified percentage. The insulation resistance is generally very high. In electrolytic capacitors, the leakage current should not exceed

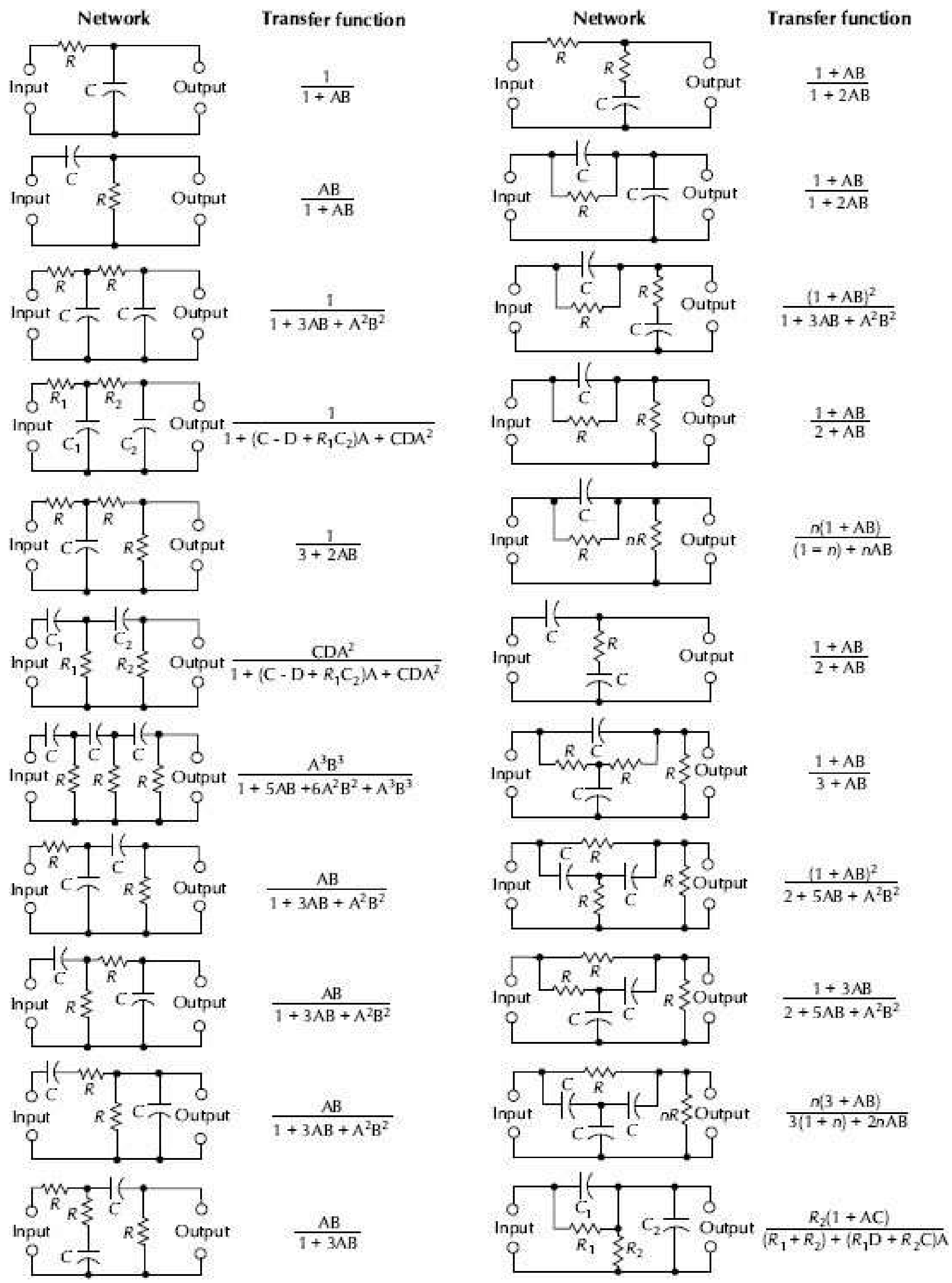


Figure 14-17. Resistance-capacitance network transfer functions.

$$I_L = 0.04C + 0.30 \qquad (14\text{-}24)$$

where,

I_L is the leakage current in μA,
C is the capacitance in μF.

Impedance (Z_c). Impedance is the total opposition offered to alternating or pulsating current measured in ohms. Impedance is the vector sum of the resistive and reactive components of a capacitor and expressed mathematically as

$$Z_c = \sqrt{(ESR)^2 + (X_L - X_c)^2} \qquad (14\text{-}25)$$

where,
ESR is the equivalent series resistance in Ω,
X_L is the impedance of the inductance in Ω,
X_C is the impedance of the capacitance in Ω.

Maximum Working Voltage. All capacitors have a *maximum working voltage* that should not be exceeded. The capacitors working voltage is a combination of the dc value plus the peak ac value that may be applied during operation. For instance, if a capacitor has $10V_{dc}$ applied to it and an ac voltage of $10V_{rms}$ or $17V_{pea}$k is applied, the capacitor will have to be capable of withstanding 27V.

Power Factor (*PF*). The *power factor* represents the fraction of input volt-amperes or power dissipated in the capacitor dielectric and is virtually independent of the capacitance, applied voltage, and frequency. *PF* is the preferred measurement in describing capacitive losses in ac circuits.

Quality Factor (*Q*). The *quality factor* of a capacitor is the ratio of the capacitors reactance to its resistance at a specified frequency.

Q is found by the equation

$$Q = \frac{1}{2\pi f C R} \quad (14\text{-}26)$$

where,

f is the frequency in Hz,

C is the value of capacitance in F,

R is the internal resistance in Ω.

14.2.4 Types of Capacitors

The uses made of capacitors become more varied and more specialized each year. They are used to filter, tune, couple, block dc, pass ac, shift phase, bypass, feed through, compensate, store energy, isolate, suppress noise, and start motors, among other things. While doing this, they frequently have to withstand adverse conditions such as shock, vibration, salt spray, extreme temperatures, high altitude, high humidity, and radiation. They must also be small, lightweight, and reliable.

Capacitors are grouped according to their dielectric material and mechanical configuration. Because they may be hardwired or mounted on circuit boards, capacitors come with leads on one end, two ends, or they may be mounted in a dual-in-line (DIP) or single in-line (SIP) package. Figs. 14-18 and 14-19 show the various types of capacitors, their characteristics, and their color codes. Table 14-4 gives the characteristics of ceramic, film, and electrolytic capacitors and Table 14-5 gives the pros and cons of each type.

Table 14-4. Capacitor Characteristics[3]

Characteristics	Ceramic			Film		Electrolytic	
	NPO (COG)	7R	Y5V	Polyester	Polypropylene	Aluminum	Tantalum
Operating Temperature	−55 to +125°C	−55 to +125°C	−30 to +85°C	−55 to +125°C	−55 to +105°C	−40 to +105°C	−55 to +125°C
Dielectric Constant	15–150	600–5200	7000–22000	3.1–3.3	2.1–2.3	7–10	24
DF	0.10%	2.50%	5%	0.35%	2%	8%	20%
ΔTC	±30 ppm 1°C	±15%	+22/−82%	±12%	±1%	±10%	±8%
ESR	Excellent	Good	Fair	Fair	Fair	Poor	Poor
ESL	Excellent	Excellent	Good	Fair	Fair	Poor	Poor
Frequency Response	Superior	Excellent	Excellent	Fair	Fair	Poor	Poor
Polar	No	No	No	No	No	Yes	Yes
Environmental Concerns	No	No	No	Yes	Yes	Yes	Yes

Table 14-5. Pros and Cons of Ceramic, Film, Aluminum Electrolytic, and Tantalum Electrolytic Capacitors

Capacitor Type		Derating		Pros	Cons
		Voltage	Temperature		
Ceramic	Non-polarized	None	None	• Non-Polarized • Small size • Transient resistant • Low cost	• Large voltage Coefficient • Aging • Limited cap range • Short failure mode
Film	Non-polarized	None	None	• Non-Polarized • Transient resistant • Stability: Voltage & temp	• Large size • Higher cost • Limited soldering heat
Aluminum Electrolytic*	Polarized	None	None	• High cap & high VDC • Surge VDC resistant • Self healing • Open failure mode • Low cost • Stability: Voltage	• Polarized • Limited lifetime • Large size
Tantalum Electrolytic	Polarized	Yes	Yes	• Long lifetime • Small sizes • Stability: Voltage & temp	• Polarized • Low VDC • Limited surge resistance • Short failure mode (Typ)

*Aluminum Electrolytic includes liquid electrolyte, hybrid construction and solid polymer types

14.2.4.1 Film Capacitors

Film capacitors consist of alternate layers of metal foil, and one or more layers of a flexible plastic insulating material (dielectric) in ribbon form rolled and encapsulated.

14.2.4.2 Paper Foil-Filled Capacitors

Paper foil-filled capacitors consist of alternate layers of aluminum foil and paper rolled together. The paper may be saturated with oil and the assembly mounted in an oil-filled, hermetically sealed metal case. They are often used as motor capacitors and are rated at 60Hz.

14.2.4.3 Mica Capacitors

Two types of *mica capacitors* are in use. In one type, alternate layers of metal foil and mica insulation, are stacked together and encapsulated. In the silvered-mica type, a silver electrode is screened on the mica insulators that are then assembled and encapsulated. Mica capacitors have small capacitance values and are usually used in high frequency circuits.

14.2.4.4 Ceramic Capacitors

Ceramic capacitors are the most popular capacitors for bypass and coupling applications because of their variety of sizes, shapes, and ratings.

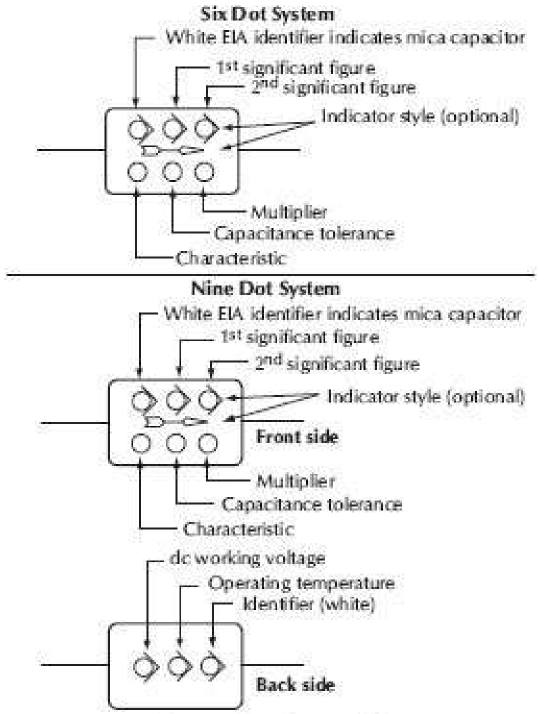
Ceramic capacitors also come with a variety of *K* values or dielectric constant. The higher the *K* value, the smaller the size of the capacitor. However, high *K*-value capacitors are less stable. High-*K* capacitors have a dielectric constant over 3000, are very small, and have values between 0.001μF to several microfarads.

When temperature stability is important, capacitors with a *K* in the 10 to 200 region are required. If a high *Q* capacitor is also required, the capacitor will be physically larger. Ceramic capacitors can be made with a zero capacitance/temperature change. These are called *negative-positive-zero* (NPO). They come in a capacitance range of 1.0pF to 0.033μF.

A temperature-compensated capacitor with a designation of N750 is used when temperature compensation is required. The 750 indicates that the capacitance will decrease at a rate of 750ppm/°C with a temperature rise or the capacitance value will decrease 1.5% for a 20°C (68°F) temperature increase. N750 capacitors come in values between 4.0pF and 680pF.

14.2.4.5 Electrolytic Capacitors

Electrolytic capacitors are still not perfect. Low temperatures reduce performance and can even freeze electrolytes, while high temperatures can dry them out and the electrolytes themselves can leak and corrode the equipment. Also, repeated surges over the rated working voltage, excessive ripple currents, and high operating temperature reduce performance and shorten capacitor life. Even with their faults, electrolytic capacitors account for one-third of the total dollars spent on capacitors because they provide high capacitance in small volume at a relatively low cost per microfarad-volt.



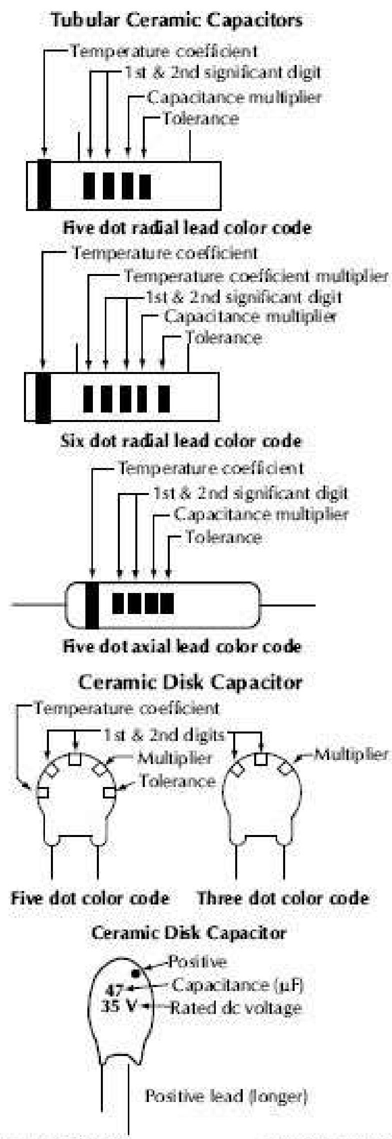
Mica Capacitor Characteristics

Characteristic	Temperature coefficient of capacitance (ppm/°C)	Maximum capacitance drift
B	Not specified	Not specified
C	±200	±0.5% +0.5 pF
D	±100	±0.3% +0.1 pF
E	−20 to +100	±0.3% +0.1 pF
F	0 to +70	±0.5% +0.1 pF

Color	Char.	Digits 1st	Digits 2nd	Multiplier	Tolerance	dc working voltage	Operating temperature range
Black		0	0	1	±20%	100	
Brown	B	1	1	10	±1%		−55°C to +85°C
Red	C	2	2	100	±2%	300	
Orange	D	3	3	1000			−55°C to +125°C
Yellow	E	4	4	10,000		500	
Green	F	5	5	100,000	±5%		
Blue		6	6	1,000,000			
Violet		7	7	10,000,000			
Gray		8	8				
White		9	9				
Gold		–	–	0.1	±0.5%	1000	
Silver		–	–	0.01	±10%		

Figure 14-18. Color codes for mica capacitors.

Basic to the construction of electrolytic capacitors is the electrochemical formation of an oxide film on a metal surface. Intimate contact is made with this oxide film by means of another electrically conductive material. The metal on which the oxide film is formed serves as the anode or positive terminal of the capacitor; the oxide film is the dielectric, and the cathode or negative terminal is either a conducting liquid or a gel. The most commonly used basic materials are aluminum and tantalum.



Capacitance in picofarads						Temperature coefficient		
Color	Digits		Multiplier	Tolerance			5 dot	6 dot
	1st	2nd		10 pF	>10 pF	pp°C	Sig. Fig.	Multiplier
Black	0	0	1	±2.0 pF	±20%	0	0.0	−1
Brown	1	1	10	±0.1 pF	±1%	−33		−10
Red	2	2	100		±2%	−75	1.0	−100
Orange	3	3	1000		±3%	−150	1.5	−1000
Yellow	4	4				−230	2.0	−10,000
Green	5	5		±0.5 pF	±5%	−300	3.3	+1
Blue	6	6				−470	4.7	+10
Violet	7	7				−750	7.5	+100
Gray	8	8	0.01	±0.25 pF		−150 to −1500		+1000
White	9	9	0.1	±1.0 pF	10%	+100 to −75		+10,000
Gold	—	—						
Silver	—	—						

Figure 14-19. Color codes for tubular and disk ceramic capacitors.

Aluminum Electrolytic Capacitors. *Aluminum electrolytic* *capacitors* use aluminum as the base material. The surface is often etched to increase the surface area as much as 100 times that of unetched foil, resulting in higher capacitance in the same volume.

The thickness of the oxide coating dielectric is determined by the voltage used to form it. The working voltage of the capacitor is somewhat less than this formation voltage. Thin films result in low voltage, high capacitance units; thicker films produce higher voltage, lower capacitance units for a given case size.

As a capacitor section is wound, a system of paper spacers is put in place to separate the foils. This prevents the possibility of direct shorts between anode and cathode foils that might result because of rough surfaces or jagged edges on either foil. The spacer material also absorbs the electrolyte with which the capacitor is impregnated, and thus assures uniform and intimate contact with all of the surface eccentricities of the etched anode foil throughout the life of the capacitor. The cathode foil serves only as an electrical connection to the electrolyte which is in fact the true cathode of the electrolytic capacitor.

The electrolyte commonly used in aluminum electrolytic capacitors is an ionogen that is dissolved in and reacts with glycol to form a pastelike mass of medium resistivity. This is normally supported in a carrier of high purity craft or hemp paper. In addition to the glycol electrolyte, low resistivity nonaqueous electrolytes are used to obtain a lower ESR and wider operating temperatures.

The foil-spacer-foil combination is wound into a cylinder, inserted into a suitable container, impregnated, and sealed.

- **Electrical Characteristics**. The equivalent circuit of an electrolytic capacitor is shown in Fig. 14-20. A and B are the capacitor terminals. The shunt resistance, R_s, in parallel with the effective capacitance, C, accounts for the dc leakage current through the capacitor. Heat is generated in the ESR if there is ripple current and heat is generated in the shunt resistance by the voltage. In an aluminum electrolytic capacitor, the ESR is due mainly to the spacer-electrolyte-oxide system. Generally it varies only slightly except at low temperatures where it increases greatly. L is the self-inductance of the capacitor caused by terminals, electrodes, and geometry.

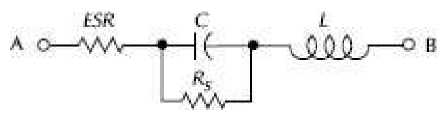


Figure 14-20. Simplified equivalent circuit of an electrolytic capacitor.

- **Impedance**. The *impedance* of a capacitor is frequency dependent, as shown in Fig. 14-21. Here, ESR is the equivalent series resistance, X_C is the capacitive reactance, X_L is the inductive reactance, and Z is the impedance. The initial downward slope is a result of the capacitive reactance. The trough (lowest impedance) portion of the curve is almost totally resistive, and the rising upper or higher frequency portion of the curve is due to the capacitor's self-inductance. If the ESR were plotted separately, it would show a small ESR decrease with frequency to about 5kHz to 10kHz, and then remain relatively constant throughout the remainder of the frequency range.

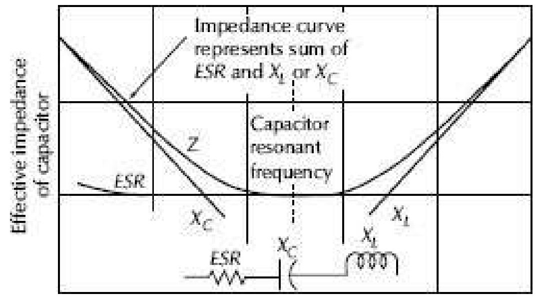


Figure 14-21. Impedance characteristics of a capacitor.

- **Leakage Current**. *Leakage current* in an electrolytic capacitor is the direct current that passes through a capacitor when a correctly polarized dc voltage is applied to its terminals. This current is proportional to temperature and becomes increasingly important when capacitors are used at elevated ambient temperatures. Imperfections in the oxide dielectric film cause high leakage currents. Leakage current decreases slowly after a voltage is applied and usually reaches *steady-state* conditions after 10 minutes.

  If a capacitor is connected with its polarity backward, the oxide film is forward biased and offers very little resistance to current flow, resulting in high current, which, if left unchecked, will cause overheating and self destruction of the capacitor.

  The total heat generated within a capacitor is the sum of the heat created by the I^2R losses in the *ESR* and that created by the $I_{Leakage} \times V_{applied}$.
- **ac Ripple Current**. The *ac ripple current* rating is one of the most important factors in filter applications, because excessive current produces a greater than permissible temperature rise, shortening capacitor life. The maximum permissible rms ripple current for any capacitor is limited by the temperature within the capacitor and the rate of heat dissipation from the capacitor. Lower *ESR* and longer cans or enclosures increase the ripple

current rating.

- **Reverse Voltage**. Aluminum electrolytic capacitors can withstand a *reverse voltage* of up to 1.5V without noticeable effect from its operating characteristics. Higher reverse voltages, when applied over extended periods, will lead to some loss of capacitance. Excess reverse voltages applied for short periods will cause some change in capacitance but may not lead to capacitor failure during the reverse voltage application or during subsequent operation in the normal polarity direction.

A major use of large value capacitors is for filtering in dc power supplies. After a capacitor is fully charged, when the rectifier conduction decreases, the capacitor discharges into the load until the next half cycle, Fig. 14-22. Then on the next cycle the capacitor recharges again to the peak voltage. The Δe shown in the illustration is equal to the total peak-to-peak ripple voltage. This is a complex wave which contains many harmonics of the fundamental ripple frequency and is the ripple that causes the noticeable heating of the capacitor.

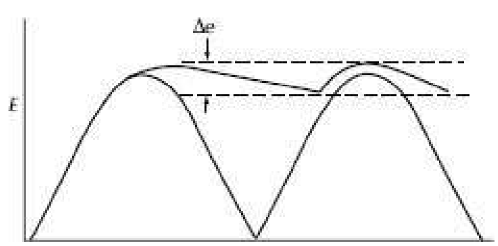


Figure 14-22. Capacitor charge and discharge on a full-wave rectifier output.

This can be mathematically determined or the ripple current through the capacitor can be measured by inserting a low impedance true rms ammeter in series with the capacitor. It is very important that the impedance of the meter be small compared with that of the capacitor, otherwise, a large measurement error will result.

- **Standard Life Tests**. *Standard life tests* at rated voltage and maximum rated temperatures are usually the criteria for determining the quality of an electrolytic capacitor. These two conditions rarely occur simultaneously in practice. Capacitor life expectancy is doubled for each 10°C (18°F) decrease in operating temperature, so a capacitor operating at room temperature will have a life expectancy 64 times that of the same capacitor operating at 85°C (185°F).
- **Surge Voltage**. The *surge voltage* specification of a capacitor determines its ability to withstand the high transient voltages that occur during the start up period of equipment. Standard tests specify a short on and long off period for an interval of 24 hours or more; the allowable surge voltage levels are usually 10% above the rated voltage of the capacitor. Fig. 14-23 shows how temperature, frequency, time, and applied voltage affect electrolytic capacitors.

Tantalum Capacitors. Tantalum electrolytics have become the preferred type where high reliability and long service life are paramount considerations.

Most metals form crystalline oxides that are nonprotecting, such as rust on iron or black oxide on copper. A few metals form dense, stable, tightly adhering, electrically insulating oxides. These are the so-called *valve* metals and include titanium, zirconium, niobium, tantalum, hafnium, and aluminum. Of these, the most valuable for the electronics industry are aluminum and tantalum.

The dielectric used in all tantalum electrolytic capacitors is tantalum pentoxide. Although wet foil capacitors use a porous paper separator between their foil plates, its function is merely to hold the electrolyte solution and to keep the foils from touching.

Rating for rating, tantalum capacitors tend to have as much as three times better capacitance/volume efficiency than aluminum electrolytic capacitors, because tantalum pentoxide has a dielectric constant of 26, some three times greater than that of aluminum oxide. This, in addition to the fact that extremely thin films can be deposited during manufacturing, makes the tantalum capacitor extremely efficient with respect to the number of microfarads available per unit volume.

The capacitance of any capacitor is determined by the surface area of the two conducting plates, the distance between the plates, and the dielectric constant of the insulating material between the plates. The distance between the plates in tantalum electrolytic capacitors is very small since it is only the thickness of the tantalum pentoxide film. The dielectric constant of the tantalum pentoxide is high, therefore, the capacitance of a tantalum capacitor is high.

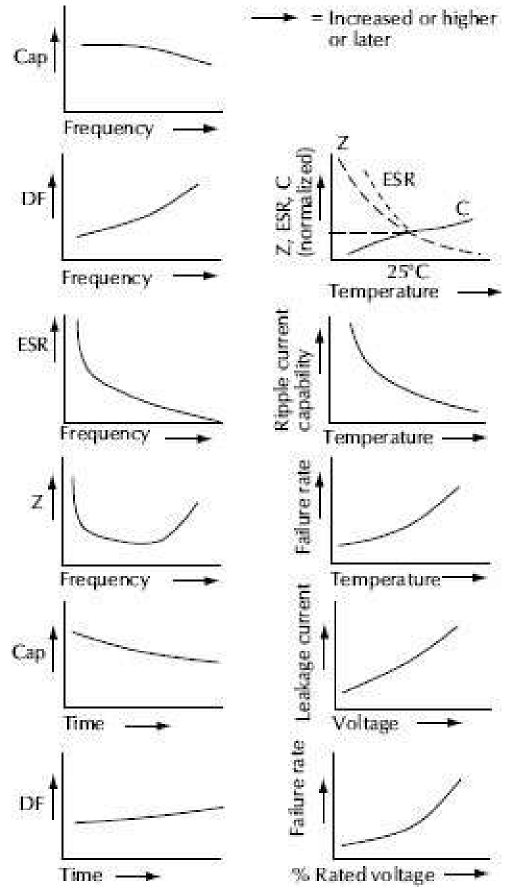


Figure 14-23. Variations in aluminum electrolytic characteristics caused by temperature, frequency, time, and applied voltage. Courtesy Sprague Electric Company.

Tantalum capacitors contain either liquid or solid electrolytes. The liquid electrolyte in wet-slug and foil capacitors, usually sulfuric acid, forms the cathode or negative plate. In solid-electrolyte capacitors a dry material, manganese dioxide, forms the cathode plate. The anode lead wire from the tantalum pellet consists of two pieces. A tantalum lead is embedded in, or welded to, the pellet, which is welded, in turn, to a nickel lead. In hermetically sealed types, the nickel lead is terminated to a tubular eyelet. An external lead of nickel or solder-coated nickel is soldered or welded to the eyelet. In encapsulated or plastic-encased designs, the nickel lead, which is welded to the basic tantalum lead, extends through the external epoxy resin coating or the epoxy end fill in the plastic outer shell.

Foil Tantalum Capacitors. *Foil tantalum capacitors* are made by rolling two strips of thin foil, separated by a paper saturated with electrolyte, into a convolute roll. The tantalum foil, which is to be the anode, is chemically etched to increase its effective surface area, providing more capacitance in a given volume. This is followed by anodizing in a chemical solution under direct voltage. This produces the dielectric tantalum pentoxide film on the foil surface.

Foil tantalum capacitors can be manufactured in dc working voltage values up to 300V. However, of the three types of tantalum electrolytic capacitors, the foil design has the lowest capacitance per unit volume. It is also the least often encountered since it is best suited for the higher voltages primarily found in older designs of equipment and requires more manufacturing operations than do the two other types. Consequently, it is more expensive and is used only where neither a solid electrolyte nor a wet-slug tantalum capacitor can be employed.

Foil tantalum capacitors operate from -55°C to +125°C (-67°F to +257°F) and are found primarily in industrial and military electronics equipment.

Wet-Electrolyte Sintered Anode Tantalum Capacitors. *Wet-electrolyte sintered anode tantalum capacitors* often called *wet-slug* tantalum capacitors, use a pellet of sintered tantalum powder to which a lead has been attached. This anode has an enormous surface area for its size because of its construction. Tantalum powder of suitable fineness, sometimes mixed with binding agents, is machine-pressed into pellets. The second step is a sintering operation in which binders, impurities, and contaminants are vaporized and the tantalum particles are sintered into a porous mass with a very large internal surface area. A tantalum lead wire is attached by welding the wire to the pellet. (In some cases, the lead is embedded during pressing of the pellet before sintering.) A film of tantalum pentoxide is electrochemically formed on the surface areas of the fused tantalum particles. The oxide is then grown to a thickness determined by the applied voltage.

Finally the pellet is inserted into a tantalum or silver container that contains an electrolyte solution. Most liquid electrolytes are gelled to prevent the free movement of the solution inside the container and to keep the electrolyte in intimate contact with the capacitor cathode. A suitable end seal arrangement prevents the loss of the electrolyte. Wet-slug tantalum capacitors are manufactured in a working voltage range up to 150Vdc.

Solid-Electrolyte Sintered Anode Tantalum Capacitors. *Solid-electrolyte sintered anode tantalum capacitors* differ from the wet versions in their electrolyte. Here, the electrolyte is manganese dioxide, which is formed on the tantalum pentoxide dielectric layer by impregnating the pellet with a solution of manganous nitrate. The pellets are then heated in an oven and the manganous nitrate is converted to manganese dioxide. The pellet is next coated with graphite followed by a layer of metallic silver, which provides a solderable surface between the pellet and its can. The pellets, with lead wire and header attached, are inserted into the can where the pellet is held in place by solder. The can cover is also soldered into place.

Another variation of the solid-electrolyte tantalum capacitor encases the element in plastic resins, such as epoxy materials. It offers excellent reliability and high stability for consumer and commercial electronics with the added feature of low cost. Still other designs of solid tantalum capacitors, as they are commonly known, use plastic film or sleeving as the encasing material and others use metal shells which are back filled with an epoxy resin. And, of course, there are small tubular and rectangular molded plastic encasements as well.

Which Tantalum Capacitor to Use. In choosing between the three basic types of tantalum capacitors, the circuit designer customarily uses foil tantalum capacitors only where high voltage constructions are required or where there is substantial reverse voltage applied to a capacitor during circuit operation.

Wet-electrolyte sintered anode capacitors, or wet-slug tantalum capacitors, are used where the lowest dc leakage is required. The conventional silver-can design will not tolerate any reverse voltages. However, in military or aerospace applications, tantalum cases are used instead of silver cases where utmost reliability is desired. The tantalum-cased wet-slug units will withstand reverse voltages up to 3V, will operate under higher ripple currents, and can be used at temperatures up to 200°C (392°F).

Solid-electrolyte designs are the least expensive for a given rating and are used in many applications where their very small size for a given unit of capacitance is important. They will typically withstand up to 15% of the rated dc working voltage in a reverse direction. They also have good low temperature performance characteristics and freedom from corrosive electrolytes.

Polymer Capacitors. A type known as “polymer capacitors” refers to the cathode plate of a polarized capacitor which offers some advantages over tantalum. The organic conductive polymer, poly-ethylenedioxythiophene (also known as PEDOT), was first used in capacitors in the late 1990s and is an alternative to the traditional manganese dioxide (MnO_2) used in tantalum capacitors.

The physical difference between a polymer and MnO_2 tantalum capacitor is the cathode material used. There are significant advantages to using the polymer such as much lower ESR, increased reliability, benign failure modes, reduced voltage de-rating, and lower costs.

The conductivity for PEDOT is 100 to 1,000S/cm, and MnO_2 is 1 to 10S/cm, which reduces ESR by orders of magnitudes. The polymer cathode exhibits very little change in conductivity (or resistivity) across temperature changes compared to traditional cathode materials. Neither tantalum-MnO_2 nor polymer-tantalum capacitors have a wear-out mechanism, as their constructions do not contain an electrolyte. They are a solid-state structure and could have a nominal lifetime on the order of hundreds or thousands of years. The most appropriate consideration for a polymer-tantalum is their "turn-on" performance when compared to MnO_2. A tantalum capacitor is most likely to fail when voltage is first applied. The experiences of tantalum catching fire on failure are actually related to the MnO_2 being used as the cathode and not the tantalum or its oxide. When polymer is used as the cathode, there isn't an opportunity for the capacitor to ignite.

There are a few trade-offs to consider when selecting a polymer-tantalum for an application. The conductive polymer has a limit to its operational temperatures. While some polymer capacitors can be rated to 125°C, most are limited to 85°C or 105°C applications.

While polymer based capacitors do have the ability to self-heal, their high conductivity means less localized heating. With less localized heating, there will be less self-healing or proofing from the cathode layer. This results in higher leakage currents. For example an MnO_2 design with 1nA of leakage current may have 1μA of leakage current when a polymer cathode is used.

At room temperature (25°C), the MnO_2 (standard) tantalum starts losing capacitance after about 10kHz while the polymer tantalum maintains its capacitance past 200kHz. This is with both capacitors rated for the same voltage, same capacitance, and same case size. However, if multiple standard tantalums were in use, they may be able to be replaced with a single polymer.

14.2.4.6 Suppression Capacitors

Suppression capacitors are used to reduce interference that comes in or out through the power line. They are effective because they are frequency dependent in that they become a short circuit at radio frequencies, without affecting low frequencies. Suppression capacitors are identified as X capacitors and Y capacitors. Fig. 14-24 shows two examples of radio interference suppression. Fig. 14-24A is for protection class I which would include drills and hair dryers. Fig. 14-24B is for protection class II where no protective conductor is connected to the metal case G.

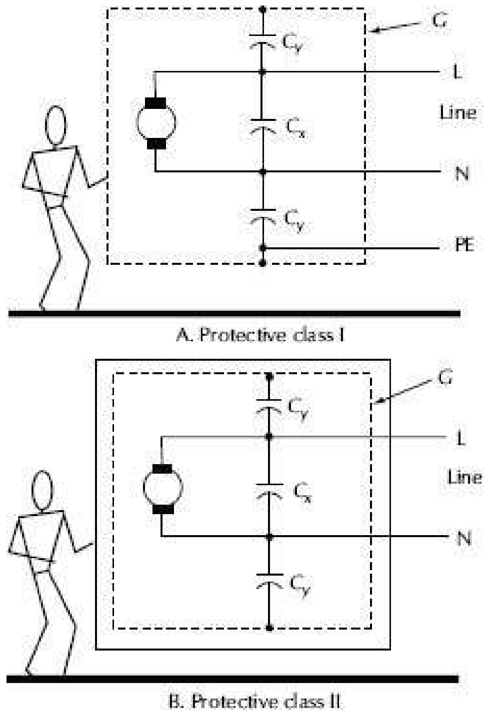


Figure 14-24. Radio frequency suppression with X and Y capacitors. Courtesy Vishay Roederstein.

X Capacitors. X capacitors are used across the mains to reduce symmetrical interference where a failure in the capacitor—i.e., the capacitor shorts out—will not cause injury, shock or death.

Y Capacitors. Y capacitors are used between a live conductor and a cabinet or case to reduce asymmetrical interference. Y capacitors have high electrical and mechanical specifications so they are much less likely to fail.

XY Capacitors. When used together they are called XY capacitors.

14.2.4.7 Supercapacitors

Supercapacitors, Ultracapacitors, more technically known as electrochemical double-layer capacitors, are one more step beyond the electrolytic capacitors. The charge-separation distance in ultracapacitors has been reduced to literally the dimensions of the ions within the electrolyte. In supercapacitors, the charges are not separated by millimeters or micrometers (microns) but by a few nanometers. The charge-separation distance has in each instance dropped by three orders of magnitude, from 10^{-3}m to 10^{-6}m to 10^{-9}m.

14.2.4.7.1 How a Supercapacitor Works

Instead of storing energy using a solid dielectric like conventional capacitors, supercapacitors involve two layers and are often referred to as EDLCs (electrochemical double-layer capacitors). In an EDLC, a physical mechanism generates the electric double layer that performs the function of a dielectric. The charge-discharge cycle is created through an ion absorption layer at the surface of the positive and negative activated carbon electrodes. The distance of the static separation of charge in an EDLC double-layer is extremely small – on the order of 0.3 to 0.8 nm. Fig. 14-25 shows the ion activity for the charge (left) and discharge (right) cycles.

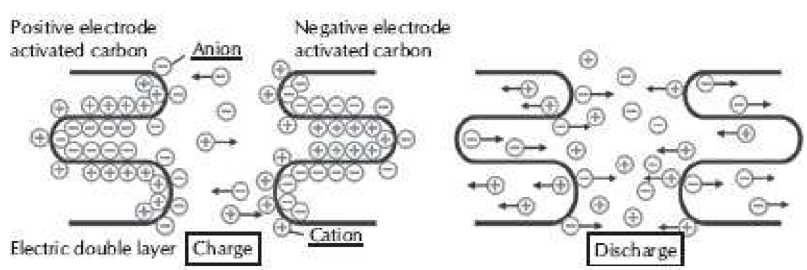


Figure 14-25. EDLCs typically store electrical charges in activated carbon electrodes. Courtesy Panasonic.

EDLCs use ion migration inside a thin membrane of activated carbon to store electrical charges. Applying a voltage across the capacitor's two electrodes causes the ions in the electrolyte to migrate in an attempt to reverse the charge on the electrodes (the charging part of the cycle). Positively charged ions move to the negative electrode and negatively charged ions move the positive electrode, forming two charged layers inside the electrolyte: one positive and one negative. Removing the voltage causes the ions to move in the opposite direction, creating the discharge half of the cycle.

Once the ultracapacitor is charged and energy stored, a load can use this energy. The amount of energy stored is very large compared to a standard capacitor because of the enormous surface area created by the porous carbon electrodes and the small charge separation of 10 angstroms created by the dielectric separator. However, it stores a much smaller amount of energy than does a battery. Since the rates of charge and discharge are determined solely by its physical properties, the ultracapacitor can release energy much faster (with more power) than a battery that relies on slow chemical reactions.

Many applications can benefit from ultracapacitors, whether they require short power pulses or low-power support of critical memory systems. Using an ultra-capacitor in conjunction with a battery combines the power performance of the battery with the greater energy storage capability of the ultracapacitor. It can extend the life of a battery, save on replacement and maintenance costs, and enable a battery to be downsized. At the same time, it can increase available energy by providing high peak power whenever necessary. The combination of ultracapacitors and batteries requires additional dc/dc power electronics, which increases the cost of the circuit.

Supercapacitors merged with batteries (hybrid battery) will become the new superbattery. Just about everything that is now powered by batteries will be improved by this much better energy supply. They can be made in most any size, from postage stamp to hybrid car battery pack. Their light weight and low cost make them attractive for most portable electronics and phones, as well as for aircraft and automobiles.

- **Advantages of a Supercapacitor**

1. Virtually unlimited life cycle—cycles millions of times—10 to 12 year life.
2. Can be charged in seconds.
3. Cannot be overcharged.
4. High self-discharge—the self-discharge rate is considerably higher than that of an electrochemical battery. No maintenance (replacement).
5. Standardized cases.
6. Low ESR.
7. High power density.
8. Circuit board mountable.
9. Capable of very high rates of charge and discharge.
10. High cycle efficiency (95% or more).

- **Disadvantages of a Supercapacitor:**

1. Supercapacitors and ultracapacitors are relatively expensive in terms of cost per watt.
2. Linear discharge voltage prevents use of the full energy spectrum.
3. Low energy density—typically holds one-fifth to one-tenth the energy of an electrochemical battery.
4. Cells have low voltages; therefore, serial connections are needed to obtain higher voltages, which require voltage balancing if more than

three capacitors are connected in series.

5. Requires sophisticated electronic control and switching equipment.

- **Advantages of a Battery**

1. High energy density.
2. Wide operating temperature range.
3. Long discharge time.

- **Disadvantages of a Battery**

1. High cost, relative to lifespan.
2. Subject to shipping regulations.
3. Cycles (<1,000).

The latest supercapacitors have capacitive densities so high that they can be used for applications previously reserved for batteries only. Supercaps will not displace batteries in most applications, as they are not as volumetrically efficient and are more expensive. However supercaps have advantages over batteries, making them a preferred choice where large amount of energy storage and delivered in repeated bursts are required.

Supercapacitors can be repeatedly charged and discharged without losing performance as opposed to batteries that can normally only be charged 400 or 500 times. Batteries and supercapacitors are often used in conjunction with each other because over time, rechargeable batteries will lose their ability to fully charge from their daily recharging so the addition of a supercapacitor to supplement the battery can greatly extend its useful life. The parallel-wired batteries supply the bulk energy, while the supercapacitors supply short pulses or energy bursts (which can degrade batteries).

Battery types that pair well with supercaps include:

- Lithium thionyl chloride.
- Lithium Manganese Dioxide.
- Lithium Iodine.
- Zinc air.
- Zinc/silver oxide.
- Poly (carbon Monofloride-lithium).

Because of their ability to quickly charge and deliver surges or energy bursts, supercaps are a good fit for data streams or devices that generate pulses. Remotely monitored sensors that rely on energy harvesting and security strobe lighting are two such examples. In such installations, the supercap may be used solo or with a battery.

Supercapacitors have found many other uses including a wide variety of energy storage and energy harvesting applications. In UPS systems, power conditioners, and power inverters, two supercapacitors may replace a bank of conventional capacitors, reducing size and weight. Used as memory backup in personal computers and computer-based systems, they do not need to be replaced over the lifespan of the product.

In battery-support applications, supercapacitors supply power to the system in short energy bursts, while the batteries recharge the supercapacitors between the bursts. The life of the battery can now be extended by a factor of 3 or more. The cost of the supercap is offset by the savings found by not having to replace batteries.

14.2.4.7.2 Connecting Supercapacitors Together

When supercapacitors need to be connected together in modules to meet capacitance and/or voltage requirements, the capacitors will need to be balanced. Unlike batteries, capacitors can vary in value significantly, and still be within acceptable tolerance. With the variations in capacitance values, the voltage across an individual capacitor can exceed the voltage capabilities of the device, causing excessive heat to be generated within the capacitor and premature failure.

Two common balancing methods are used: passive and active. Passive balancing is an inexpensive solution involving resisters wired in parallel with the capacitors to equalize the voltage across the capacitors. Active balancing requires the use of ICs connected across the capacitors. The chip will monitor the network and applying changes as required. A number of semiconductor manufacturers offer chips for this application. After the capacitor balancing is resolved, there are additional requirements in the development of a supercapacitor module including physical requirements, internal connections, environmental requirements, and termination requirements. It is best to contact the manufacturer on these requirements.

14.2.4.7.3 Calculating Backup Time

To calculate the desired backup time the supercapacitor will provide if the power goes off, the starting and ending voltage on the capacitor, the current draw from the capacitor, and the capacitor size must be known.

Assuming that the load draws a constant current while running from V_{BACKUP}, then the worst-case backup time in hours would use the equation:

$$Backup\ time = \frac{\frac{C(V_{BACKUPSTART} - V_{BACKUPMIN})}{I_{BACKUPMAX}}}{3600} \qquad (14\text{-}27)$$

where,

C is the capacitor value in F,

$V_{BACKUPSTART}$ is the initial voltage in V. The voltage applied to V_{CC}, less the voltage drop from the diodes, if any, used in the charging circuit,

$V_{BACKUPMIN}$ is the ending voltage in V,

$I_{BACKUPMAX}$ is the maximum V_{BACKUP} current in A.

For example, to determine how long the backup time will be under the following conditions:

- 0.2F capacitor.
- $V_{BACKUPSTART}$ is 3.3V.
- $V_{BACKUPMIN}$ is 1.3V.
- *IBACKUPMAX* is 1000nA, then

$$Backup\ time = \frac{\frac{0.2(3.3 - 1.3)}{10^{-6}}}{300} = 111.1\ \text{h}$$

14.3 Inductors

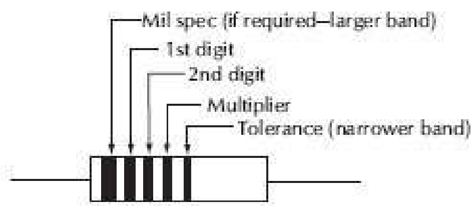
Inductance is used for the storage of electrical energy in a magnetic field, called *magnetic energy*. Magnetic energy is stored as long as current keeps flowing through the inductor. The current of a sine wave lags the voltage by 90° in a perfect inductor. Fig. 14-26 shows the color code for small inductors.

14.3.1 Types of Inductors

Inductors are constructed in a variety of ways, depending on their use.

14.3.1.1 Air Core Inductors

Air core inductors are either ceramic core or phenolic core.



Color	Inductance digits 1st	2nd	Multiplier	Tolerance
Black	0	0	1	
Brown	1	1	10	
Red	2	2	100	
Orange	3	3	1000	
Yellow	4	4	10,000	
Green	5	5		
Blue	6	6		
Violet	7	7		
Gray	8	8		
White	9	9		
Gold	-	-		±5%
Silver	-	-	-	±10%
No band	-	-	-	±20%

Figure 14-26. Color code for small inductors (in μH).

14.3.1.2 Axial Inductor

An *axial inductor* is constructed on a core with concentric leads on opposite ends, Fig. 14-27A. The core material may be phenolic, ferrite, or powdered iron.

14.3.1.3 Bobbin Core Inductor

Bobbin core inductors have the shape of a bobbin and may come with or without leads. They may be either axial or radial, Fig. 14-27B.

14.3.1.4 Ceramic Core

Ceramic core inductors are often used in high frequency applications where low inductance, low core losses, and high Q values are required. Ceramic has no magnetic properties so there is no increase in permeability due to the core material.

Ceramic has a low thermal coefficient of expansion allowing high inductance stability over a high operating temperature range.

14.3.1.5 Epoxy-Coated Inductor

Epoxy-coated inductors usually have a smooth surface and edges. The coating provides insulation.

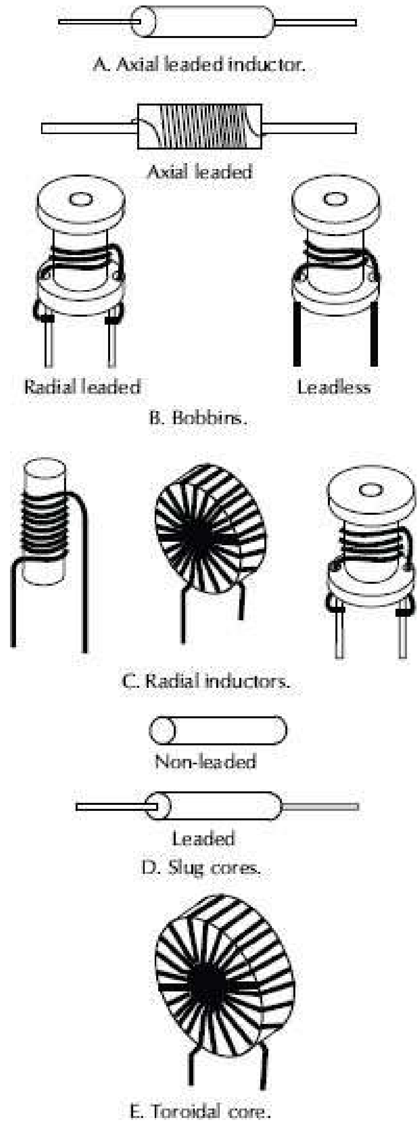


Figure 14-27. Various inductor core types.

14.3.1.6 Molded Inductor

A *molded inductor* has its case formed via a molding process, creating a smooth, well-defined body with sharp edges.

14.3.1.7 Ferrite Core

Ferrite cores can be easily magnetized. The core consists of a mixture of oxide of iron and other elements such as manganese and zinc (MnZn) or nickel and zinc (NiZn). The general composition is $xxFe_2O_4$ where xx is one of the other elements.

14.3.1.8 Laminated Cores

Laminated cores are made by stacking insulated laminations on top of each other. Some laminations have the grains oriented to minimize core losses, giving higher permeability. Laminated cores are more common in transformers.

14.3.1.9 MPP Core

MPP, or moly permalloy powder, is a magnetic material with an inherent distributed air gap, allowing it to store higher levels of magnetic flux compared to other materials. This allows more dc to flow through the inductor before the core saturates.

The core consists of 80% nickel, 4% molybdenum, and the remaining percentage is iron.

14.3.1.10 Multilayer Inductor

A *multilayer inductor* consists of layers of coil between layers of core material. The coil is usually bare metal and is sometimes referred to as *nonwirewound*.

14.3.1.11 Phenolic Core

Phenolic cores are often called *air cores* and are often used in high frequency applications where low inductance values, low core losses, and high Q values are required.

Phenolic has no magnetic properties so there is no increase in permeability due to the core material. Phenolic cores provide high strength, high fflammability ratings, and high temperature characteristics.

14.3.1.12 Powdered Iron Core

Powdered iron is a magnetic material with an inherent distributed air gap that allows the core to have high levels of magnetic flux. This allows a high level of dc to flow through the core before saturation.

Powdered iron cores are close to 100% iron whose particles are insulated and mixed with a binder such as epoxy or phenolic. They are pressed into a final shape and cured by baking.

14.3.1.13 Radial Inductor

A *radial inductor* is constructed on a core with leads on the same side, Fig. 14-27C, allowing for easy mounting on circuit boards, etc.

14.3.1.14 Shielded Inductor

A *shielded inductor* has its core designed to contain the majority of the magnetic field. Some are self shielding such as toroids, e-cores, and pot cores. Bobbin and slug cores require a magnetic sleeve for shielding.

14.3.1.15 Slug Core

Slug cores have the shape of a cylindrical rod and come with or without leads, Fig. 14-27D. They have higher flux density characteristics than other core shapes as most of the magnetic energy is stored in the air around the core.

14.3.1.16 Tape Wound Core

Tape wound cores are made by rolling insulated and precisely controlled thickness strips of alloy iron into a toroidal shape. The finished cores have an outside coating for protection. Tape wound cores are capable of storing high amounts of energy and contain a high permeability.

14.3.1.17 Toroidal Inductor

Toroidals are constructed by placing the winding on a donut-shaped core, Fig. 14-27E. Toroidal cores may be ferrite, powdered iron, tape wound, or alloy and high flux. Toroidals are self shielding and have efficient energy transfer, high coupling between windings, and early saturation.

14.3.2 Impedance Characteristics

Impedance. The impedance or inductive reactance (X_L) of an inductor to an ac signal is found with the equation

$$X_L = 2\pi f L \qquad (14\text{-}28)$$

where,

f is the frequency in Hz,

L is the inductance in H.

The inductance of a coil is only slightly affected by the type of wire used for its construction. The Q of the coil will be governed by the ohmic resistance of the wire. Coils wound with silver or gold wire have the highest Q for a given design.

To increase the inductance, inductors can be connected in series. The total inductance will always be greater than the largest inductor

$$L_T = L_1 + L_2 \cdots + L_n \tag{14-29}$$

To reduce the total inductance, place the inductors in parallel. The total inductance will always be less than the value of the lowest inductor

$$L_T = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} \ldots + \frac{1}{L_n}} \tag{14-30}$$

To determine the impedance of circuits with resistance, capacitance, and inductance, see Section 14.4.

Mutual Inductance. *Mutual inductance* is the property that exists between two conductors that are carrying current when the magnetic lines of force from one conductor link with the magnetic lines of force of the other. The mutual inductance of two coils with fields interacting can be determined by

$$M = \frac{L_A - L_B}{4} \tag{14-31}$$

where,

M is the mutual inductance of L_A and L_B in H,

L_A is the total inductance of coils L_1 and L_2 with fields aiding in H,

L_B is the total inductance of coils L_1 and L_2 with fields opposing in H.

The coupled inductance can be determined by the following equations.

In parallel with fields aiding

$$L_T = \frac{1}{\frac{1}{L_1+M} + \frac{M}{L_2+M}} \qquad (14\text{-}32)$$

In parallel with fields opposing

$$L_T = \frac{1}{\frac{1}{L_1-M} - \frac{M}{L_2-M}} \qquad (14\text{-}33)$$

In series with fields aiding

$$L_T = L_1 + L_2 + 2M \qquad (14\text{-}34)$$

In series with fields opposing

$$L_T = L_1 + L_2 - 2M \qquad (14\text{-}35)$$

where,
L_T is the total inductance in H,
L_1 and L_2 are the inductances of the individual coils in H,
M is the mutual inductance in H.

When two coils are inductively coupled to give transformer action, the coupling coefficient is determined by

$$K = \frac{M}{\sqrt{L_1 \times L_2}} \qquad (14\text{-}36)$$

where,
K is the coupling coefficient,
M is the mutual inductance in H,
L_1 and L_2 are the inductances of the two coils in H.

An inductor in a circuit has a reactance of $j2\pi fL\Omega$. Mutual inductance in a circuit also has a reactance equal to $j2\pi fM$Q. The operator j denotes reactance. The energy stored in an inductor can be determined by

$$W = \frac{LI^2}{2} \tag{14-37}$$

where,

W is the energy in J (W·s),

L is the inductance in H,

I is the current in A.

Coil Inductance. The following is the relationship of the turns in a coil to its inductance:

- The inductance is proportional to the square of the turns.
- The inductance increases as the permeability of the core material is increased.
- The inductance increases as the cross-sectional area of the core material is increased.
- The inductance increases as the length of the winding is increased.
- A shorted turn decreases the inductance. In an audio transformer, the frequency characteristic will be affected, and the insertion loss increased.
- Inserting an iron core in a coil increases the inductance; hence, its inductive reactance is increased.
- Introducing an air gap in an iron core coil reduces the inductance.

The maximum voltage induced in a conductor moving in a magnetic field is proportional to the number of magnetic lines of force cut by the conductor moving in the field. A conductor moving parallel to the lines of force cuts no lines of force so no current is generated in the conductor. A conductor moving at right angles to the lines of force will cut the maximum number of lines per inch per second; therefore, the voltage will be at the maximum.

A conductor moving at any angle to the lines of force cuts a number of lines of force proportional to the sine of the angles.

$$V = \beta L v \sin\theta \times 10^{-8} \tag{14-38}$$

V is the voltage produced,

β is the flux density,

L is the length of the conductor in cm,

v is the velocity in cm/s of the conductor moving at an angle θ.

The direction of the induced electromotive force (emf) is in the direction in which the axis of a right-hand screw, when turned with the velocity vector, moves through the smallest angle toward the flux density vector. This is called the *right-hand rule*.

The magnetomotive force produced by a coil is derived by

$$\begin{aligned} ampere\ turns &= T\left(\frac{V}{R}\right) \\ &= TI \end{aligned} \tag{14-39}$$

where,

T is the number of turns,

V is the voltage in V,

R is the resistance of the wire in Ω,

I is the current in A.

The inductance of single-layer, spiral, and multilayer coils can be calculated by using either Wheeler's or Nagaoka's equations. The accuracy of the calculation will vary between 1% and 5%. The inductance of a single-layer coil, Fig. 14-28A, may be found using Wheeler's equation

$$L = \frac{B^2N^2}{9B + 10A} \qquad (14\text{-}40)$$

For the multilayer coil, Fig. 14-28B,

$$L = \frac{0.8B^2N^2}{6B + 9A + 10C} \qquad (14\text{-}41)$$

For the spiral coil, Fig. 14-28C,

$$L = \frac{B^2N^2}{8B + 11C} \qquad (14\text{-}42)$$

where,

B is the radius of the winding,

N is the number of turns in the coil,

A is the length of the winding,

C is the thickness of the winding,

L is in µH.

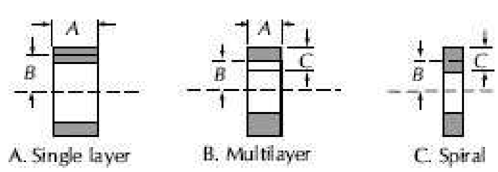


Figure 14-28. Single- and multilayer inductors.

Q. Q is the ratio of the inductive reactance to the internal resistance of the coil. The principal factors that affect Q are frequency, inductance, dc resistance, inductive reactance, and the type of winding. Other factors are the core losses, the distributed capacity, and the permeability of the core material. The Q for a coil where R and L are in series is

$$Q = \frac{2\pi f L}{R} \qquad (14\text{-}43)$$

where,

f is the frequency in Hz,

L is the inductance in H,

R is the resistance in Ω.

The Q of the coil can be measured as follows. Using the circuit of Fig. 14-29, Q of a coil may be easily measured for frequencies up to 1MHz. Since the voltage across an inductance at resonance equals $Q \times V$, where V is the voltage developed by the oscillator, it is necessary only to measure the output voltage from the oscillator and the voltage across the inductance.

The voltage from the oscillator is introduced across a low value of resistance R, about 1% of the anticipated radio frequency resistance of the LC combination, to assure that the measurement will not be in error by more than 1%. For average measurements, resistor R will be on the order of 0.10Ω. If the oscillator cannot be operated into an impedance of 0.10Ω, a matching transformer may be employed. It is desirable to make C as large as convenient to minimize the ratio of the impedance looking from the voltmeter to the impedance of the test circuit. The voltage across R is made small, on the order of 0.10V. The *LC* circuit is then adjusted to resonate and the resultant voltage measured. The value of *Q* may then be equated

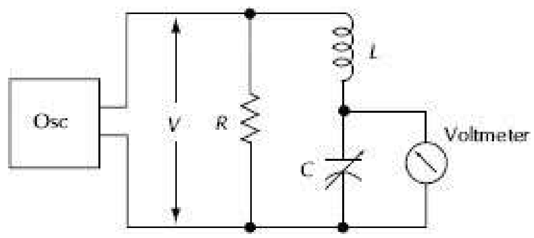


Figure 14-29. Circuit for measuring the *Q* of a coil.

$$Q = \frac{\text{Resonant voltage across } C}{\text{Voltage across } R} \qquad (14\text{-}44)$$

The *Q* of a coil may be approximated by the equation

$$Q = \frac{2\pi f L}{R} = \frac{X_L}{R} \qquad (14\text{-}45)$$

where,

f is the frequency in Hz,

L is the inductance in H,

R is the dc resistance in Ω as measured by an ohmmeter,

X_L is the inductive reactance of the coil.

Time Constant. When a dc voltage is applied to an *RL* circuit, a certain amount of time is required to change the voltage. In a circuit containing inductance and resistance, the time constant is defined as the time it takes for the current to reach 62.3% of its maximum value. The time constant can be determined with the equation

$$T = \frac{L}{R} \qquad (14\text{-}46)$$

where,

T is the time in s,

L is the inductance in H,

R is the resistance in Ω.

See section 14.2.1 for a further discussion of time constants. The effect of an inductor is the same as for a capacitor and resistor. Also, curve A in Fig. 14-16 shows the current through an inductor on buildup and curve B shows the current decay when the voltage is removed.

Right-Hand Rule. The *right-hand rule* is a method devised for determining the direction of a magnetic field around a conductor carrying a direct current. The conductor is grasped in the right hand with the thumb extended along the conductor. The thumb points in the direction of the current. If the fingers are partly closed, the fingertips will point in the direction of the magnetic field. Maxwell's rule states, "If the direction of travel of a right-handed corkscrew represents the direction of the current in a straight conductor, the direction of rotation of the corkscrew will represent the direction of the magnetic lines of force."

14.3.3 Ferrite Beads

The original ferrite beads were small round ferrites with a hole through the middle where a wire passed through. Today they come as the original style plus as multiple apertures and surface mount configurations.

The ferrite bead can be considered a frequency-dependent resistor whose equivalent circuit is a resistor in series with an inductor. As the frequency increases, the inductive reactance increases and then decreases, and the complex impedance of the ferrite material increases the overall impedance of the bead, Fig. 14-30.

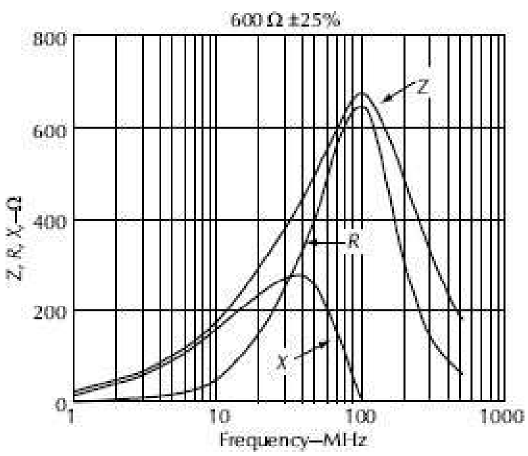


Figure 14-30. Impedance of ferrite beads. Courtesy of Vishay Dale.

At frequencies below 10MHz, the impedance is less than 10Ω. As the frequency increases, the impedance increases to about 100Ω and becomes mostly resistive at 100MHz.

Once the impedance is resistive, resonance does not occur as it would using an LC network. Ferrite beads do not attenuate low frequencies or dc so are useful for reducing EMI/EMC in audio circuits.

14.3.4 Skin Effect

Skin effect is the tendency of ac to flow near the surface of a conductor rather than flowing through the conductor's entire cross sectional area. This increases the resistance of the conductor because the magnetic field caused by the current creates eddy currents near the center of the conductor. The eddy currents oppose the normal flow of current near the center, forcing the main current flow out toward the surface as the frequency of the ac current increases.

To reduce this problem, a wire made up of separately insulated strands woven and/or bunched together is used. Commonly called Litz wire, the current is equally divided between all of the individual strands which equalizes the flux linkage and reactance of the individual strands, reducing the ac losses compared to solid wire.

14.3.5 Shielded Inductor

Some inductor designs are self-shielding. Examples are toroid, pot core, and E-core inductors. Slug cores and bobbins may require shielding, depending on the application. It is impossible to completely shield an inductor.

14.3.6 Impedance

The total impedance created by resistors, capacitors, and inductors in circuits can be determined with the following equations.

Parallel circuits

$$Z = \frac{RX}{\sqrt{R^2 + X^2}} \tag{14-47}$$

Series circuits

$$Z = \sqrt{R^2 + X^2} \tag{14-48}$$

Resistance and inductance in series

$$Z = \sqrt{R^2 + X_L^2} \tag{14-49}$$

$$\theta = \operatorname{atan}\frac{X_L}{R} \tag{14-50}$$

Resistance and capacitance in series

$$Z = \sqrt{R^2 + X_C^2} \tag{14-51}$$

$$\theta = \operatorname{atan}\frac{X_C}{R} \tag{14-52}$$

Inductance and capacitance in series when X_L is larger than X_C

$$Z = X_L - X_C \tag{14-53}$$

Inductance and capacitance in series when X_C is larger than X_L

$$Z = X_C - X_L \tag{14-54}$$

Resistance, inductance, and capacitance in series

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{14-55}$$

$$\theta = \operatorname{atan}\frac{X_L - X_C}{R} \tag{14-56}$$

Resistance and inductance in parallel

$$Z = \frac{RX_L}{\sqrt{R^2 + X_L^2}} \tag{14-57}$$

Capacitance and resistance in parallel

$$Z = \frac{RX_C}{\sqrt{R^2 + X_C^2}} \tag{14-58}$$

Capacitance and inductance in parallel when X_L is larger than X_C

$$Z = \frac{X_L \times X_C}{X_L - X_C} \tag{14-59}$$

Capacitance and inductance in parallel when X_C is larger than X_L

$$Z = \frac{X_C \times X_L}{X_C - X_L} \tag{14-60}$$

Inductance, capacitance, and resistance in parallel

$$Z = \frac{RX_LX_C}{\sqrt{X_L^2X_C^2 + R^2(X_L - X_C)^2}} \tag{14-61}$$

$$\theta = \operatorname{atan}\frac{R(X_L - X_C)}{X_LX_C} \tag{14-62}$$

Inductance and series resistance in parallel with resistance

$$Z = R_2\sqrt{\frac{R_1^2 + X_L^2}{(R_1 + R_2)^2 + X_L^2}} \tag{14-63}$$

$$\theta = \operatorname{atan}\frac{R_2X_L}{R_1^2 + X_L^2 + R_1R_2} \tag{14-64}$$

Inductance and series resistance in parallel with capacitance

$$Z = X_C\sqrt{\frac{R^2+X_L^2}{R^2+(X_L-X_C)^2}} \qquad (14\text{-}65)$$

$$\theta = \operatorname{atan}\frac{X_L(X_C-X_L)-R^2}{RX_C} \qquad (14\text{-}66)$$

Capacitance and series resistance in parallel with inductance and series resistance

$$Z = \sqrt{\frac{(R_1^2+X_L^2)(R_2^2+X_C^2)}{(R_1+R_2)^2+(X_L-X_C)^2}} \qquad (14\text{-}67)$$

$$Z = \operatorname{atan}\frac{X_L(R_2^2+X_C^2)-X_C(R_1^2+X_L^2)}{R_1(R_2^2+X_C^2)+R_2(R_1^2+X_L^2)} \qquad (14\text{-}68)$$

where,

Z is the impedance in Ω,

R is the resistance in Ω,

L is the inductance in H,

X_L is the inductive reactance in Ω,

X_C is the capacitive reactance in Ω.

θ is the phase angle in degrees by which the current leads the voltage in a capacitive circuit or lags the voltage in an inductive circuit. 0° indicates an in-phase condition.

14.4 Resonant Frequency

When an inductor and capacitor are connected in series or parallel, they form a resonant circuit. The resonant frequency can be determined from the equation

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi C X_C} = \frac{X_L}{2\pi L} \tag{14-69}$$

where,

L is the inductance in H,

C is the capacitance in F,

X_L and X_C are the impedance in Ω.

The resonant frequency can also be determined through the use of a reactance chart developed by the Bell Telephone Laboratories, Fig. 14-31. This chart can be used for solving problems of inductance, capacitance, frequency, and impedance. If two of the values are known, the third and fourth values may be found. As an example, what is the value of capacitance and inductance required to resonate at a frequency of 1000Hz in a circuit having an impedance of 500Ω? Entering the chart on the 1000Hz vertical line and following it to the 500Ω line (impedance is shown along the left-hand margin), the value of inductance is indicated by the diagonal line running upward as 0.08H (80mH), and the capacitance indicated by the diagonal line running downward at the right-hand margin is 0.3μF.

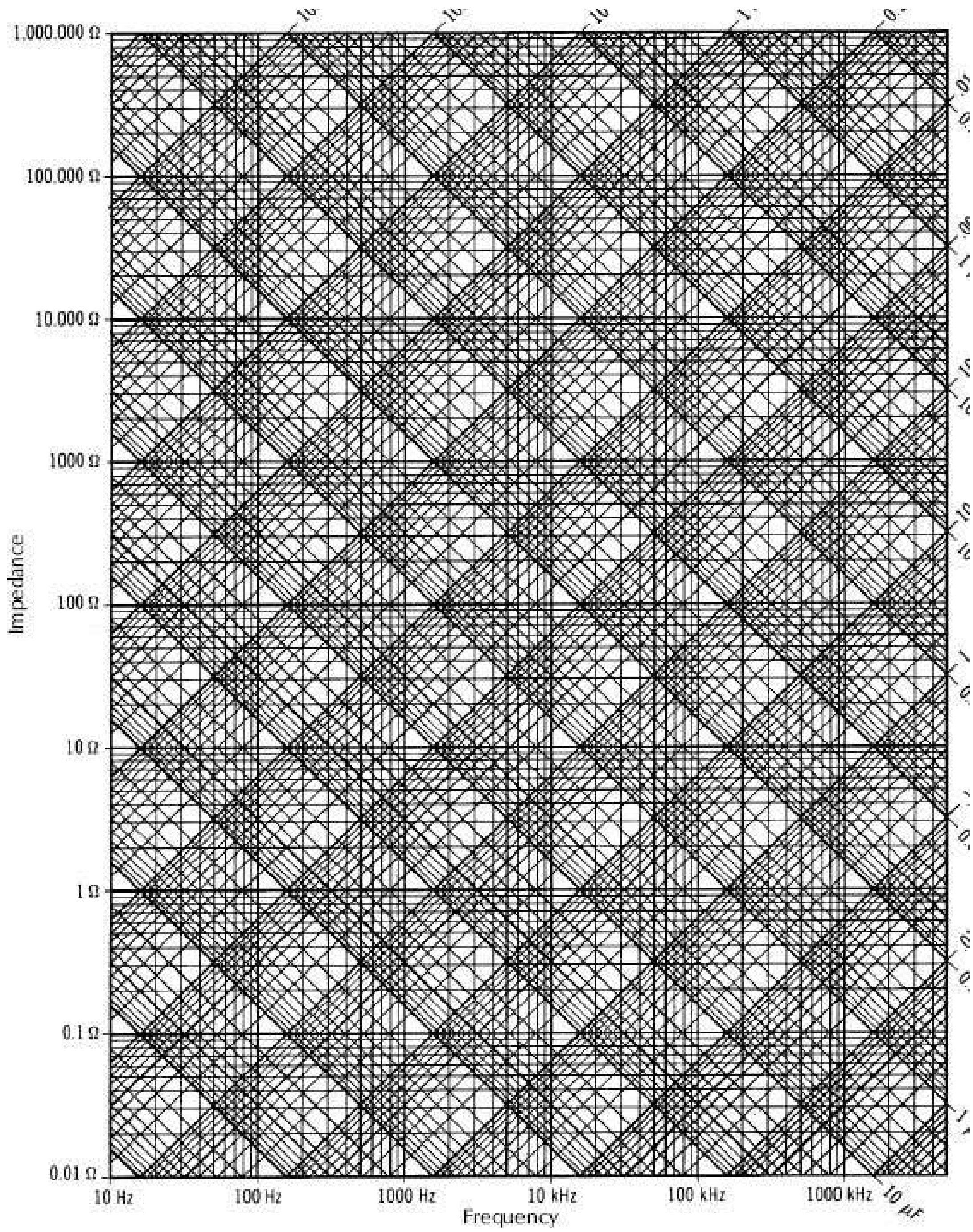


Figure 14-31. Reactance chart. Courtesy AT&T Bell Laboratories.

References

1. Roger Allan, "Temperature Sensors Are Hot…In Circuit Design," *Electronic Design*, May 2008

2. Jerry Janesch, “Two Wire vs. Four Wire Resistance Measurements,” *EDN*, April 2013.

3. Michael Roach, “Choices in Switch-Mode Power Supply Capacitors,” *Electronic Products*, April 2013.

Bibliography

Inductor and Magnetic Product Terminology, Document No. 34053, March 2002, Vishay Dale.

ILB, ILBB Ferrite Beads, Document No. 34091, May 1999, Vishay Dale.

Ohmite Manual of Engineering Information, Bulletin 1100, Ohmite Mfg. Co.

Ohmite Resistor Selection Design Data, Ohmite Mfg. Co.

Radio Interference Suppression Capacitors, Document No. 26529, September 2002, Vishay Roederstein.

Chapter 15

Audio Transformers

by Bill Whitlock

15.1 Audio Transformer Basics

Since the birth of audio electronics, the audio transformer has played an important role. When compared to modern miniaturized electronics, a transformer seems large, heavy, and expensive but it continues to be the most effective solution in many audio applications. The usefulness of a transformer lies in the fact that electrical energy can be transferred from one circuit to another without direct connection, e.g., isolation from ground loops, and in the process the energy can be readily changed from one voltage level to another, e.g., impedance matching. Although a transformer is not a complex device, considerable explanation is required to properly understand how it operates. This chapter is intended to help the audio system engineer properly select and apply transformers. In the interest of simplicity, only basic concepts of their design and manufacture will be discussed.

15.1.1 Basic Principles and Terminology

15.1.1.1 Magnetic Fields and Induction

As shown in Fig. 15-1, a *magnetic field* is created around any conductor (wire) in which current flows. The strength of the field is directly proportional to current. These invisible magnetic *lines of force*, collectively called *flux*, are set up at right angles to the wire and have a direction, or magnetic polarity, that depends on the direction of current flow. Note that although the flux around the upper and lower wires have different directions, the lines inside the loop aid because they point in the same direction. If an alternating current flows in the loop, the instantaneous intensity and polarity of the flux will vary at the same frequency and in direct proportion to Fig. 15-2, as expanding, contracting, and reversing in polarity with each cycle of the ac current. The law of induction states that a voltage will be *induced* in a conductor exposed to changing flux and that the induced voltage will be proportional to the rate of the flux change. This voltage has an instantaneous polarity which opposes the original current flow in the wire, creating an apparent resistance called *inductive reactance*. Inductive reactance is calculated according to the formula

$$X_L = 2\pi f L \qquad (15\text{-}1)$$

where,

X_L is inductive reactance in Ω

f is the frequency in Hz,

L is inductance in H.

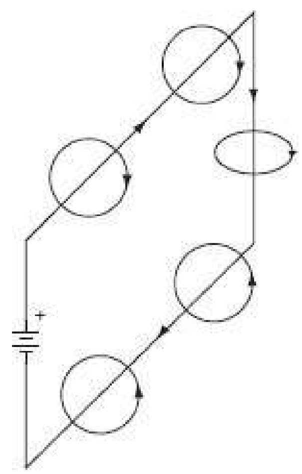


Figure 15-1. Magnetic field surrounding conductor.

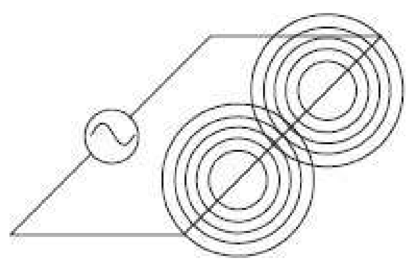


Figure 15-2. ac magnetic field.

An *inductor* generally consists of many turns or loops of wire called a *coil*, as shown in Fig. 15-3, which links and concentrates magnetic flux lines, increasing the *flux density*. The inductance of any given coil is determined by factors such as the number of turns, the physical dimensions and nature of the winding, and the properties of materials in the path of the magnetic flux.

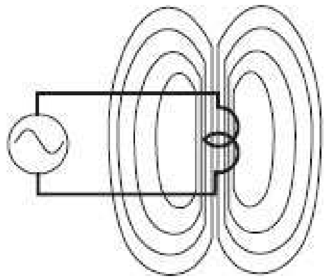


Figure 15-3. Coil concentrates flux.

According to the law of induction, a voltage will be induced in any conductor (wire) that cuts flux lines. Therefore, if we place two coils near each other as shown in Fig. 15-4, an ac *current* in one coil will induce an ac voltage in the second coil. This is the essential principle of energy transfer in a *transformer*. Because they require a changing magnetic field to operate, transformers will not work at dc. In an ideal transformer, the magnetic coupling between the two coils is total and complete, i.e., all the flux lines generated by one coil cut across all the turns of the other. The *coupling coefficient* is said to be unity or 1.00.

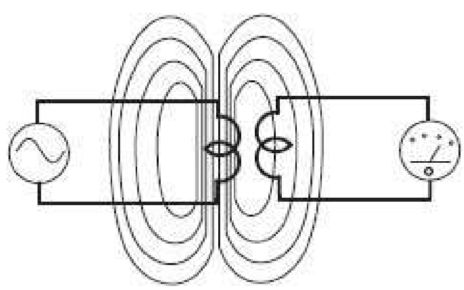


Figure 15-4. Inductive coupling.

15.1.1.2 Windings and Turns Ratio

The coil or winding that is driven by an electrical source is called the *primary* and the other is called the *secondary*. The ratio of the number of turns on the primary to the number of turns on the secondary is called the *turns ratio*. Since essentially the same voltage is induced in each turn of each winding, the primary to secondary *voltage ratio* is the same as the turns ratio. For example, with 100 turns on the primary and 50 turns on the secondary, the turns ratio is 2:1. Therefore, if 20V were applied to the primary, 10V would appear at the secondary. Since it reduces voltage, this transformer would be called a *step-down* transformer. Conversely, a transformer with a turns ratio of 1:2 would be called a *step-up* transformer since its secondary voltage would be twice that of the primary. Since a transformer cannot create power, the power output from the secondary of an ideal transformer can only equal (and in a real transformer can only be less than) the power input to the primary. Consider an ideal 1:2 step-up transformer. When 10V is applied to its primary, 20V appears at its secondary. Since no current is drawn by the primary (this is an ideal transformer—see "15.1.1.3, Excitation Current," its impedance appears to be infinite or an open circuit.

However, when a 20Ω load is connected to the secondary, a current of 1A flows making output power equal 20W. To do this, a current of 2A must be drawn by the primary, making input power equal 20W. Since the primary is now drawing 2A with 10V applied, its impedance appears to be 5Ω. In other words, the 20Ω load impedance on the secondary has been *reflected* to the primary as 5Ω. In this example, a transformer with a 1:2 turns ratio exhibited an impedance ratio of 1:4. Transformers always reflect impedances from one winding to another by the square of their turns ratio or, expressed as an equation

$$V = IR \qquad (14\text{-}2)$$

where,

Z_p is primary impedance,
Z_s is secondary impedance,
N_p/N_s is turns ratio, which is the same as the voltage ratio.

When a transformer converts voltage, it also converts impedance and vice versa.

The direction in which coils are wound—i.e., clockwise or counterclockwise—and/or the connections to the start or finish of each winding determines the instantaneous *polarity* of the ac voltages. All windings that are wound in the same direction will have the same polarity between start and finish ends. Therefore, relative to the primary, polarity can be inverted by either (1) winding the primary and secondary in opposite directions, or (2) reversing the start and finish connections to either winding. In schematic symbols for transformers, dots are generally used to indicate which ends of windings have the same polarity. Observing polarity is essential when making series or parallel connections to transformers with multiple windings. *Taps* are connections made at any intermediate point in a winding. For example, if 50 turns are wound, an electrical connection brought out, and another 50 turns completes the winding, the 100 turn winding is said to be *centertapped*.

15.1.1.3 Excitation Current

While an ideal transformer has infinite primary inductance, a real transformer does not. Therefore, as shown in Fig. 15-5, when there is no load on the secondary and an ac voltage is applied to the primary, an *excitation current* will flow in the primary, creating magnetic excitation flux around the winding. In theory, the current is due only to the inductive reactance of the primary winding. In accordance with Ohm's Law and the equation for inductive reactance,

$$I_E = \frac{E_p}{2\pi f L_p} \qquad (15\text{-}3)$$

where,

I_E is excitation current in A,

E_P is primary voltage in V,

f is frequency in Hz,

L_P is primary inductance in H.

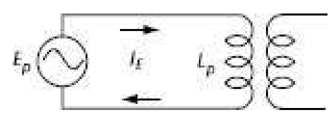


Figure 15-5. Excitation current.

Obviously, if primary inductance were infinite, excitation current would be zero. As shown in Fig. 15-6, when a load is connected, current will flow in the secondary winding. Because secondary current flows in the opposite direction, it creates magnetic flux which opposes the excitation flux. This causes the impedance of the primary winding to drop, resulting in additional current being drawn from the driving source. Equilibrium is reached when the additional flux is just sufficient to completely cancel that created by the secondary. The result, which may surprise some, is that flux density in a transformer is *not* increased by load current. This also illustrates how load current on the secondary is reflected to the primary.

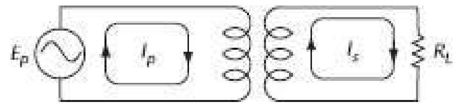


Figure 15-6. Cancellation of flux generated by load current.

Fig. 15-7 illustrates the relationships between voltage, excitation current, and flux in a transformer as *frequency* is changed. The horizontal scale is time. The primary voltage E_p is held constant as the frequency is changed (tripled and then tripled again). For example, the left waveform could represent one cycle at 100Hz, the middle 300Hz, and the right 900Hz. Because of the primary inductance, excitation current I_p will decrease linearly with frequency, i.e., halving for every doubling in frequency or decreasing at 6dB per octave. The magnitude of the magnetic flux will likewise *decrease* exactly the same way. Note that the inductance causes a 90° phase lag between voltage and current as well. Since the slew rate of a constant amplitude sine wave *increases* linearly with frequency, i.e., doubling for every doubling in frequency or increasing at 6dB per octave, the resultant flux rate of change remains *constant*. Note that the slope of the I_p and flux waveforms stays constant as frequency is changed. Since, according to the law of induction, the voltage induced in the secondary is proportional to this slope or rate of change, output voltage also remains uniform, or flat versus frequency.

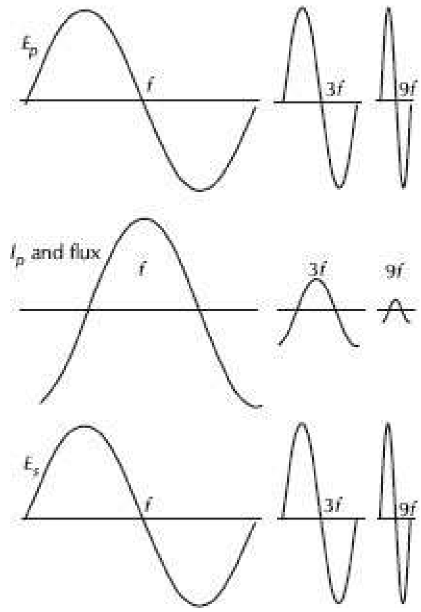


Figure 15-7. Excitation current and flux vary inversely with frequency.

15.1.2 Realities of Practical Transformers

Thus far, we have not considered the unavoidable parasitic elements which exist in any practical transformer. Even the design of a relatively simple 60Hz power transformer must take parasitics into account. The design of an audio transformer operating over a 20Hz to 20kHz frequency range is much more difficult because these parasitics often interact in complex ways. For example, materials and techniques that improve low-frequency performance are often detrimental to high-frequency performance and vice versa. Good transformer designs must consider both the surrounding electronic circuitry and the performance ramifications of internal design tradeoffs.

A schematic representation of the major low frequency parasitic elements in a generalized transformer is shown in Fig. 15-8. The IDEAL TRANSFORMER represents a perfect transformer having a turns ratio of 1: N and no parasitic elements of any kind. The actual transformer is connected at the PRI terminals to the driving voltage source, through its source impedance R_G, and at the SEC terminals to the load R_L.

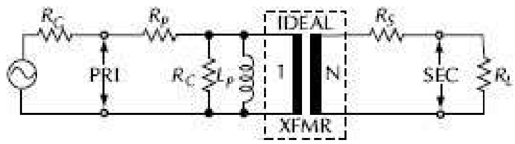


Figure 15-8. Transformer low-frequency parasitic elements.

One of the main goals in the design of any transformer is to reduce the excitation current in the primary winding to negligible levels so as not to become a significant load on the driving source. For a given source voltage and frequency, primary excitation current can be reduced only by increasing inductance L_P. In the context of normal electronic circuit impedances, very large values of inductance are required for satisfactory operation at the lowest audio frequencies. Of course, inductance can be raised by using a very large number of coil turns but, for reasons discussed later, there are practical limits due to other considerations. Another way to increase inductance by a factor of 10,000 or more is to wind the coil around a highly magnetic material, generally referred to as the *core*.

15.1.2.1 Core Materials and Construction

Magnetic circuits are quite similar to electric circuits. As shown in Fig. 15-11, magnetic flux always takes a closed path from one magnetic pole to the other and, like an electric current, always favors the paths of highest conductivity or least resistance. The equivalent of applied voltage in magnetic circuits is *magnetizing* *force*, symbolized H. It is directly proportional to ampere-turns (coil current I times its number of turns N) and inversely proportional to the flux path length ℓ in the magnetic circuit. The equivalent of electric current flow is *flux density*, symbolized B. It represents the number of magnetic flux lines per square unit of area. A graphic plot of the relationship between field intensity and flux density is shown in Fig. 15-9 and is referred to as the "B-H loop" or "hysteresis loop" for a given material. In the United States, the most commonly used units for magnetizing force and flux density are the Oersted and Gauss, respectively, which are CGS (centimeter, gram, second) units. In Europe, the SI (Système Internationale) units amperes per meter and tesla, respectively, are more common. The slope of the B-H loop indicates how an incremental increase in applied magnetizing force changes the resulting flux density. This slope is effectively a measure of conductivity in the magnetic circuit and is called *permeability*, symbolized μ. Any material inside a coil, which can also serve as a form to support it, is called a *core*. By definition, the permeability of a vacuum, or air, is 1.00 and common nonmagnetic materials such as aluminum, brass, copper, paper, glass, and plastic also have a permeability of 1 for practical purposes. The permeability of some common ferromagnetic materials is about 300 for ordinary steel, about 5000 for 4% silicon transformer steel, and up to about 100,000 for some nickel-iron-molybdenum alloys. Because such materials concentrate magnetic flux, they greatly increase the inductance of a coil. Audio transformers must utilize both high-permeability cores and the largest practical number of coil turns to create high primary inductance. Coil inductance increases as the square of the number of turns and in direct proportion to the permeability of the core and can be approximated using the equation

$$L = \frac{3.2N^2\mu A}{10^8 l} \qquad (15\text{-}4)$$

where,

L is the inductance in H,

N is the number of coil turns,

μ is the permeability of core,

A is the cross-section area of core in in^2,

l is the mean flux path length in in.

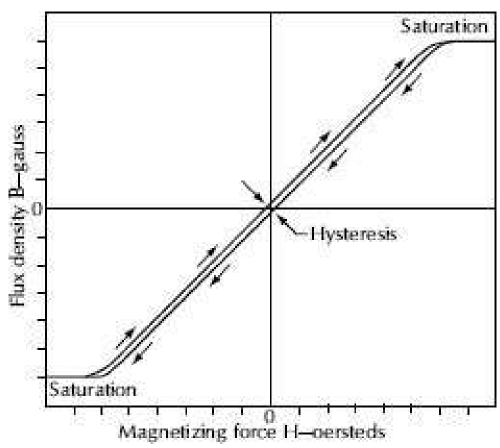


Figure 15-9. B-H loop for magnetic core material.

The permeability of magnetic materials varies with flux density. As shown in Fig. 15-9, when magnetic field intensity becomes high, the material can *saturate*, essentially losing its ability to conduct any additional flux. As a material saturates, its permeability decreases until, at complete saturation, its permeability becomes that of air or 1. In audio transformer applications, magnetic saturation causes low-frequency harmonic distortion to increase steadily for low-frequency signals as they increase in level beyond a threshold. In general, materials with a higher permeability tend to saturate at a lower flux density. In general, permeability also varies inversely with frequency.

Magnetic *hysteresis* can be thought of as a magnetic memory effect. When a magnetizing force saturates material that has high-hysteresis, it remains strongly magnetized even after the force is removed. High-hysteresis materials have wide or square B-H loops and are used to make magnetic memory devices and permanent magnets. However, if we magnetically saturate zero-hysteresis material, it will have no residual magnetism (flux density) when the magnetizing force is removed. But virtually all high-permeability core materials have *some* hysteresis, retaining a small memory of their previous magnetic state. Hysteresis can be greatly reduced by using certain metal alloys that have been annealed or heat-treated using special processes. In audio transformers, the nonlinearity due to magnetic hysteresis causes increased harmonic distortion for low-frequency signals at relatively low signal levels. Resistor R_C in Fig. 15-8 is a nonlinear resistance that, in the equivalent circuit model, represents the combined effects of magnetic saturation, magnetic hysteresis, and eddy-current losses.

The magnetic *operating point*, or zero signal point, for most transformers is the center of the B-H loop shown in ig. 15-9, where the net magnetizing force is zero. Small ac signals cause a small portion of the loop to be traversed in the direction of the arrows. Large ac signals traverse portions farther from the operating point and may approach the saturation end points. For this normal operating point at the center, signal distortions (discussed in detail later) caused by the curvature of the loop are symmetrical, i.e., they affect the positive and negative signal excursions equally.

Symmetrical distortions produce odd-order harmonics such as third and fifth. If dc current flows in a winding, the operating point will shift to a point on the loop away from the center. This causes the distortion of a superimposed ac signal to become nonsymmetrical. Nonsymmetrical distortions produce even-order harmonics such as second and fourth. When a small dc current flows in a winding, under say 1% of the saturation value, the effect is to add even-order harmonics to the normal odd-order content of the hysteresis distortion, which affects mostly low level signals. The same effects occur when the core becomes weakly magnetized, as could happen via the brief accidental application of dc to a winding, for example. However, the narrow B-H loop indicates that only a weak residual field would remain even if a magnetizing force strong enough to saturate the core were applied and then removed.

When a larger dc current flows in a winding, the symmetry of saturation distortion is also affected in a similar way. For example, enough dc current might flow in a winding to move the operating point to 50% of the core saturation value. Only half as much ac signal could then be handled before the core would saturate and, when it did, it would occur only for one direction of the signal swing. This would produce strong second-harmonic distortion. To avoid such saturation effects, *air gaps* are sometimes intentionally built into the magnetic circuit. This can be done, for example, by placing a thin paper spacer between the center leg of the E and I cores of Fig. 15-10. The magnetic permeability of such a gap is so low—even though it may be only a few thousandths of an inch—compared to the core material, that it effectively controls the flux density in the entire magnetic circuit. Although it drastically reduces the inductance of the coil, gapping is done to prevent flux density from reaching levels that would otherwise saturate the core, especially when substantial dc is present in a winding.

Figure 15-10. Core laminations are stacked and interleaved around bobbin that holds windings.

Because high-permeability materials are usually electrical conductors as well, small voltages are also induced in the cross-section of the core material itself, giving rise to *eddy currents*. Eddy currents are greatly reduced when the core consists of a stack of thin sheets called *laminations*, as shown in Fig. 15-10. Because the laminations are effectively insulated from each other, eddy currents generally become insignificant. The E and I shaped laminations shown form the widely used shell or double-window, core construction. Its parallel magnetic paths are illustrated in Fig. 15-11. When cores are made of laminations, care must be taken that they are flat and straight to avoid tiny air gaps between them that could significantly reduce inductance.

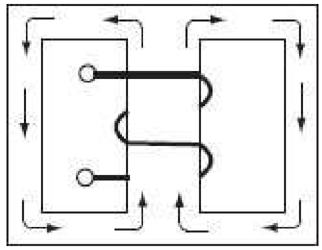


Figure 15-11. Magnetic circuits in shell core.

A *toroidal core* is made by rolling a long thin strip of core material into a coiled ring shape that looks something like a donut. It is insulated with a conformal coating or tape and windings are wound around the core through the center hole using special machines. With a toroidal core, there are no unintended air gaps that can degrade magnetic properties. Audio transformers don't often use toroidal cores because, especially in high bandwidth designs where multiple sections or Faraday shields are necessary, physical construction becomes very complex. Other core configurations include the *ring core*, sometimes called *semitoroidal*. It is similar to core of Fig. 15-11 but without the center section and windings are placed on the sides. Sometimes a solid—not laminations—metal version of a ring core is cut into two pieces having polished mating faces. These two *C-cores* are then held together with clamps after the windings are installed.

15.1.2.2 Winding Resistances and Auto-Transformers

If zero-resistance wire existed, some truly amazing transformers could be built. In a 60Hz power transformer, for example, we could wind a primary with tiny wire on a tiny core to create enough inductance to make excitation current reasonable. Then we could wind a secondary with equally tiny wire. Because the wire has no resistance and the flux density in the core doesn't change with load current, this postage-stamp-sized transformer could handle unlimited kilowatts of power—and it wouldn't even get warm! But, at least until practical superconducting wire is available, real wire has resistance. As primary and secondary currents flow in the winding resistances, the resulting voltage drops cause signal loss in audio transformers and significant heating in power transformers. This resistance can be reduced by using larger—lower gauge—wire or fewer turns, but the required number of turns and the tolerable power loss (or resulting heat) all conspire to force transformers to become physically larger and heavier as their rated power increases. Sometimes silver wire is suggested to replace copper, but since its resistance is only about 6% less, its effect is minimal and certainly not cost-effective. However, there is an alternative configuration of transformer windings, called an *auto-transformer*, which can reduce the size and cost in certain applications. Because an auto-transformer electrically connects primary and secondary windings, it can't be used where electrical isolation is required! In addition, the size and cost advantage is maximum when the required turns ratio is very close to 1:1 and diminishes at higher ratios, becoming minimal in practical designs at about 3:1 or 1:3.

For example, in a hypothetical transformer to convert 100V to 140V, the primary could have 100 turns and the secondary 140 turns of wire. This transformer, with its 1:1.4 turns ratio, is represented in the upper diagram of Fig. 15-12. If 1A of secondary (load) current I_S flows, transformer output power is 140W and 1.4A of primary current I_P will flow since input and output power must be equal in the ideal case. In a practical transformer, the wire size for each winding would be chosen to limit voltage losses and/or heating.

An auto-transformer essentially puts the windings in series so that the secondary voltage adds to (boosting) or subtracts from (bucking) the primary input voltage. A step-up auto-transformer is shown in the middle diagram of Fig. 15-12. Note that the dots indicate ends of the windings with the same instantaneous polarity. A 40V secondary (the upper winding) series connected, as shown with the 100V primary, would result in an output of 140V. Now, if 1A of secondary load current I_S flows, transformer output power is only 40W and only 0.4A of primary current I_P will flow. Although the total power delivered to the load is still 140W, 100W have come directly from the driving source and only 40W have been transformed and added by the auto-transformer. In the auto-transformer, 100 turns of smaller wire can be used for the primary and only 40 turns of heavier wire is needed for the secondary. Compare this to the total of 240 turns of heavier wire required in the transformer.

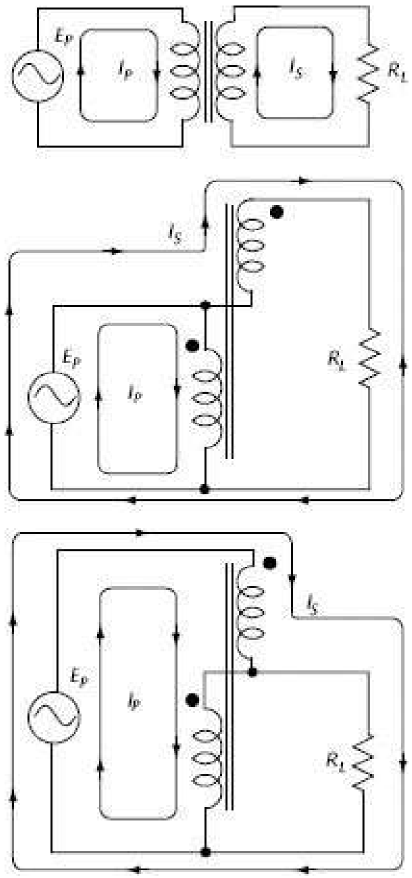


Figure 15-12. Auto-transformers employ a buck/boost principle.

A step-down auto-transformer is shown in the bottom diagram of Fig. 15-12. Operation is similar except that the secondary is connected so that its instantaneous polarity subtracts from or bucks the input voltage. For example, we could step down U.S. 120Vac power to Japanese 100Vac power by configuring a 100V to 20V step-down transformer as an auto-transformer. Thus, a 100W load can be driven using only a 20W rated transformer.

The windings of low level audio transformers may consist of hundreds or even many thousands of turns of wire, sometimes as small as #46 gauge, whose 0.0015 inch diameter is comparable to a human hair. As a result, each winding may have a dc resistance as high as several thousand ohms. Transformer primary and secondary winding resistances are represented by R_P and R_S, respectively, in Fig. 15-8.

15.1.2.3 Leakage Inductance and Winding Techniques

In an ideal transformer, since all flux generated by the primary is linked to the secondary, a short circuit on the secondary would be reflected to the primary as a short circuit. However, in real transformers, the unlinked flux causes a residual or *leakage inductance* that can be measured at either winding. Therefore, the secondary would appear to have residual inductance if the primary were shorted and vice-versa. The leakage inductance is shown as L_L in the model of Fig. 15-13. Note that leakage inductance is reflected from one winding to another as the square of turns ratio, just as impedances are.

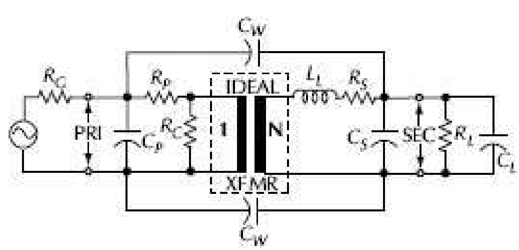


Figure 15-13. Transformer high frequency parasitic elements.

The degree of flux coupling between primary and secondary windings depends on the physical spacing between them and how they are placed with respect to each other. The lowest leakage inductance is achieved by winding the coils on a common axis and as close as possible to each other. The ultimate form of this technique is called *multi-filar* winding where multiple wires are wound simultaneously as if they were a single strand. For example, if two windings, i.e. primary and secondary, are wound as one, the transformer is said to be *bi-filar* wound. Note in the cross-section view of Fig. 15-14 how the primary and secondary windings are side-by-side throughout the entire winding. Another technique to reduce leakage inductance is to use *layering*, a technique in which portions or *sections* of the primary and/or secondary are wound in sequence over each other to interleave them. For example, Fig. 15-15 shows the cross-section of a three-layer transformer where half the primary is wound, then the secondary, followed by the other half of the primary. This results in considerably less leakage inductance than just a secondary over primary two-layer design. Leakage inductance decreases rapidly as the number of layers is increased.

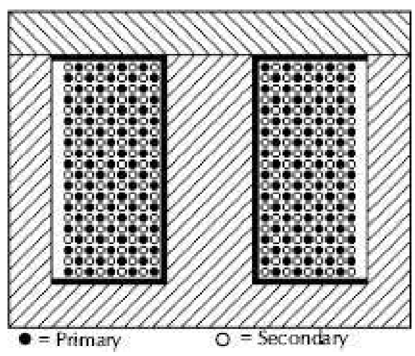


Figure 15-14. Bi-filar windings.

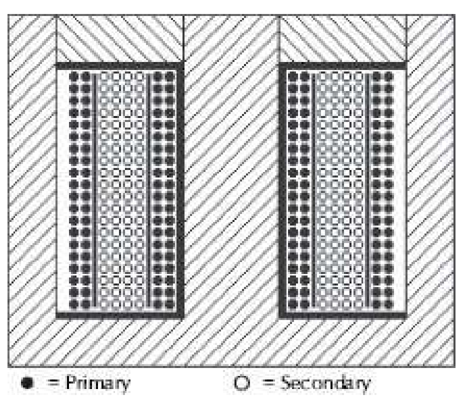


Figure 15-15. Layered windings.

15.1.2.4 Winding Capacitances and Faraday Shields

To allow the maximum number of turns in a given space, the insulation on the wire used to wind transformers is very thin. Called *magnet wire*, it is most commonly insulated by a thin film of polyurethane enamel. A transformer winding is made, in general, by spinning the bobbin shown in Fig. 15-10 on a machine similar to a lathe and guiding the wire to form a layer one wire thick across the length of the bobbin. The wire is guided to traverse back and forth across the bobbin to form a coil of many layers as shown in Fig. 15-15, where the bobbin cross-section is the solid line on three sides of the winding. This simple side-to-side, back-and-forth winding results in considerable layer-to-layer capacitance within a winding or winding section. More complex techniques such as universal winding are sometimes used to substantially reduce winding capacitances. These capacitances within the windings are represented by C_P and C_S in the circuit model of Fig. 15-13. Additional capacitances will exist between the primary and secondary windings and are represented by capacitors C_W in the model. Sometimes layers of insulating tape are added to increase the spacing, therefore reducing capacitance, between primary and secondary windings. In the bi-filar windings of Fig. 15-14, since the wires of primary and secondary windings are side by side throughout, the inter-winding capacitances C_W can be quite high.

In some applications, interwinding capacitances are very undesirable. They are completely eliminated by the use of a *Faraday shield* between the windings. Sometimes called an *electrostatic shield*, it generally takes the form of a thin sheet of copper foil placed between the windings. Obviously, transformers that utilize multiple layers to reduce leakage inductance will require

Faraday shields between all adjacent layers. In Fig. 15-15 the dark lines between the winding layers are the Faraday shields. Normally, all the shields surrounding a winding are tied together and treated as a single electrical connection. When connected to circuit ground, as shown in Fig. 15-16, a Faraday shield intercepts the capacitive current that would otherwise flow between transformer windings.

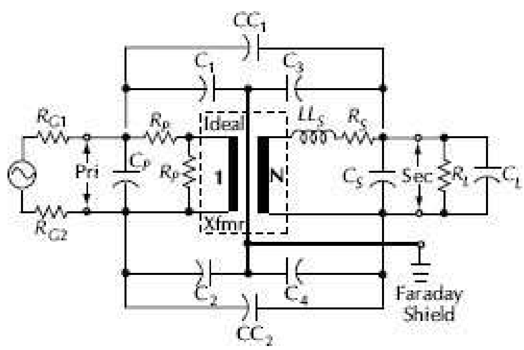


Figure 15-16. High frequency equivalent circuit of a transformer with Faraday shield and driven by a balanced source.

Faraday shields are nearly always used in transformers designed to eliminate ground noise. In these applications, the transformer is intended to respond only to the voltage difference or signal across its primary and have no response to the noise that exists equally (or common-mode) at the terminals of its primary. A Faraday shield is used to prevent capacitive coupling, via C_W in Fig. 15-13, of this noise to the secondary. For any winding connected to a balanced line, the matching of capacitances to ground is critical to the rejection of common-mode noise, or CMRR, as discussed in Chapter 36, *Grounding and Interfacing*. In Fig. 15-16, if the primary is driven by a balanced line, C_1 and C_2 must be very accurately matched to achieve high CMRR. In most applications, such as microphone or line input transformers, the secondary is operated unbalanced, i.e., one side is grounded. This relaxes the matching requirements for capacitances C_3 and C_4. Although capacitances CC_1 and CC_2 are generally quite small (a few pF) they have the effect of diminishing CMRR at high audio frequencies and limiting rejection of RF interference.

15.1.2.5 Magnetic Shielding

A *magnetic shield* has a completely different purpose. Devices such as power transformers, electric motors, and television or computer monitor cathode-ray tubes generate powerful ac magnetic fields. If such a field takes a path through the core of an audio transformer, it can induce an undesired voltage in its windings, most often heard as hum. If the offending source and the victim transformer have fixed locations, orientation of one or both can sometimes nullify the pickup. In Fig. 15-11 note that an external field that flows vertically through the core will cause a flux gradient across the length of the coil, inducing a voltage in it, but a field that flows horizontally through the core will not. Such magnetic pickup is usually worse in input transformers (discussed later) because they generally have more turns. It should also be noted that higher permeability core materials are more immune to external fields. Therefore, an unshielded output transformer with a high nickel core will be more immune than one with a steel core.

Another way to prevent such pickup is to surround the core with a closed (no air gap) magnetic path. This *magnetic shield* most often takes the form of a can or box with tight-fitting lid and is made of high permeability material. While the permeability of ordinary steel, such as that in electrical conduit, is only about 300, special-purpose nickel alloys can have permeability as high as 100,000. Commercial products include Mumetal®, Permalloy®, HyMu® and Co-Netic®.[1,2] Since the shield completely surrounds the transformer, the offending external field will now flow through it instead of the transformer core. Generally speaking, care must be taken not to mechanically stress these metals because doing so will significantly decrease their permeability. For this reason, most magnetic shield materials must be re-annealed after they are fabricated.

The effectiveness of magnetic shielding is generally rated in dB. The transformer is placed in an external magnetic field of known strength, generally at 60Hz. Its output without and with the shield is then compared. For example, a housing of 0.125 in thick cast iron reduces pickup by about 12dB, while a 0.030in thick Mumetal can reduces it by about 30dB. Where low-level transformers operate near strong magnetic fields, several progressively smaller shield cans can be nested around the transformer. Two or three Mumetal cans can provide 60dB and 90dB of shielding, respectively. In very strong fields, because high permeability materials might saturate, an iron or steel outer can is sometimes used.

Toroidal power transformers can have a weaker radiated magnetic field than other types. Using them can be an advantage if audio transformers must be located nearby. However, a toroidal transformer must be otherwise well designed to produce a low external field. For example, every winding must completely cover the full periphery of the core. The attachment points of the transformer lead wires are frequently a problem in this regard. To gain size and cost advantages, most commercial power transformers of any kind are designed to operate on the verge of magnetic saturation of the core. When saturation occurs in any transformer, magnetic field radiation dramatically increases. Power transformers designed to operate at low flux density will prevent this. A standard commercial power transformer, when operated at reduced primary voltage, will have a very low external field—comparable to that of a standard toroidal design.

15.1.3 General Application Considerations

For any given application, a number of parameters must be considered when selecting or designing an appropriate audio transformer. We will discuss how the performance of a transformer can be profoundly affected by its interaction with surrounding circuitry.

15.1.3.1 Maximum Signal Level, Distortion, and Source Impedance

Because these parameters are inextricably interdependent, they must be discussed as a group. Although transformer operating level is often specified in terms of power such as dBm or watts, what directly affects distortion is the equivalent driving voltage. Distortion is caused by excitation current in the primary winding which is proportional to primary voltage, not power. Referring to Fig. 15-8, recall that nonlinear resistance R_C represents the distortion-producing mechanisms of the core material. Consider that, if both R_G–the driving source impedance, and R_P–*the* internal winding resistance, were zero the voltage source—by definition zero impedance—would effectively short out R_C, resulting in zero distortion! But in a real transformer design there is a fixed relationship between signal level, distortion, and source impedance. Since distortion is also a function of magnetic flux density, which increases as frequency *decreases*, a maximum operating level specification must also specify a frequency. The specified maximum operating level, maximum allowable distortion at a specified low frequency, and maximum allowable source impedance will usually dictate the type of core material that must be used and its physical size. And, of course, cost plays a role, too.

The most commonly used audio transformer core materials are *M6 steel* (a steel alloy containing 6% silicon) and 49% *nickel* or 84% *nickel* (alloys containing 49% or 84% nickel plus iron and molybdenum). Nickel alloys are substantially more expensive than steel. Fig. 15-17 shows how the choice of core material affects low-frequency distortion as signal level changes using the same windings and core size. The increased distortion at low levels is due to magnetic hysteresis and at high levels is due to magnetic saturation. Fig. 15-18 shows how distortion decreases rapidly with increasing frequency. Because of differences in their hysteresis distortion, the falloff is most rapid for the 84% nickel and least rapid for the steel. Fig. 15-19 shows how distortion is strongly affected by the impedance of the driving source. The plots begin at 40Ω because that is the resistance of the primary winding. Therefore, maximum operating levels predicated on higher frequencies, higher distortion, and lower source impedance will always be higher than those predicated on lower frequencies, lower distortion, and lower source impedance.

As background, it should be said that THD, or total harmonic distortion, is a remarkably inadequate way to describe the perceived awfulness of distortion. Distortion consisting of low-order harmonics, 2nd or 3rd for example, is dramatically less audible than that consisting of high order harmonics, 7th or 13th for example. Consider that, at very low frequencies, even the finest loudspeakers routinely exhibit harmonic distortion in the range of several percent at normal listening levels. Simple distortion tests whose results correlate well with the human auditory experience simply don't exist. Clearly, such perceptions are far too complex to quantify with a single figure.

One type of distortion that is particularly audible is intermodulation or IM distortion. Test signals generally combine a large low-frequency signal with a smaller high-frequency signal and measure how much the amplitude of the high frequency is modulated by the lower frequency. Such intermodulation creates tones at new, nonharmonic frequencies. The classic SMPTE (Society of Motion Picture and Television Engineers) IM distortion test mixes 60Hz and 7kHz signals in a 4:1 amplitude ratio. For virtually all electronic amplifier circuits, there is an approximate relationship between harmonic distortion and SMPTE IM distortion. For example, if an amplifier measured 0.1% THD at 60Hz at a given operating level, its SMPTE IM distortion would measure about three or four times that, or 0.3% to 0.4% at an equivalent operating level. This correlation is due to the fact that electronic non-linearities generally distort audio signals without regard to frequency. Actually, because of negative feedback and limited gain bandwidth, most electronic distortions become worse as frequency increases.

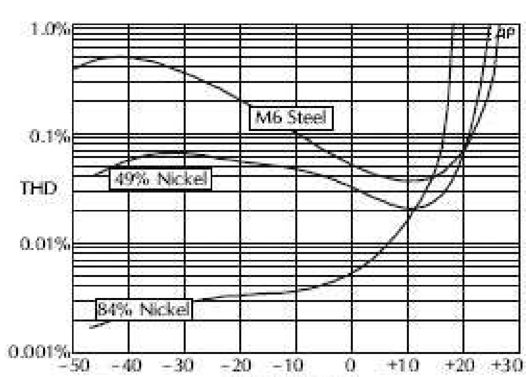


Figure 15-17. Measured THD at 20Hz and 40Ω source versus signal level for three types of core materials.

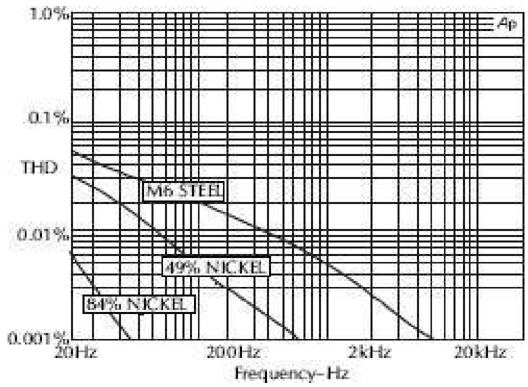


Figure 15-18. Measured THD at 0dBu and 40Ω source versus frequency for the cores of Fig. 15-17.

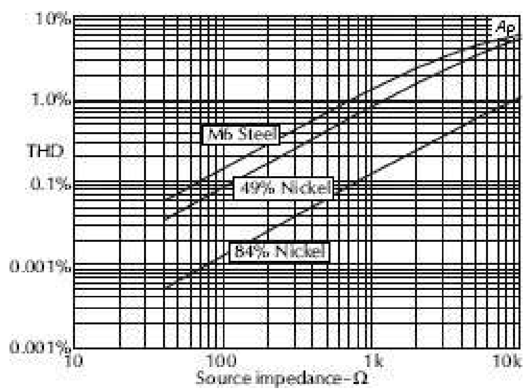


Figure 15-19. Measured THD at 0dBu and 20Hz versus source impedance for the cores of Figs. 15-17 and 15-18.

Distortion in audio transformers is different in a way that makes it sound unusually benign. It is caused by the smooth symmetrical curvature of the magnetic transfer characteristic or B-H loop of the core material shown in Fig. 15-9. The nonlinearity is related to flux density that, for a constant voltage input, is inversely proportional to frequency. The resulting harmonic distortion products are nearly pure third harmonic. In Fig. 15-18, note that distortion for 84%

nickel cores roughly quarters for every doubling of frequency, dropping to less than 0.001% above about 50Hz. Unlike that in amplifiers, the distortion mechanism in a transformer is frequency selective. This makes its IM distortion much less than might be expected. For example, the Jensen JT-10KB-D line input transformer has a THD of about 0.03% for a +26dBu input at 60Hz. But, at an equivalent level, its SMPTE IM distortion is only about 0.01% or about a tenth of what it would be for an amplifier having the same THD.

15.1.3.2 Frequency Response

The simplified equivalent circuit of Fig. 15-20 shows the high-pass RL filter formed by the circuit resistances and transformer primary inductance L_P. The effective source impedance is the parallel equivalent of $R_G + R_P$ and $R_S + R_L$. When the inductive reactance of L_P equals the effective source impedance, low-frequency response will fall to 3dB below its mid-band value. For example, consider a transformer having an L_P of 10 henrys and winding resistances R_P and R_S of 50Ω each. The generator impedance R_G is 600Ω and the load R_L is 10kΩ. The effective source impedance is then 600Ω + 50Ω in parallel with 10kΩ + 50Ω, which computes to about 610Ω. A 10H inductor will have 610Ω of reactance at about 10Hz, making response 3dB down at that frequency. If the generator impedance R_G were made 50Ω instead, response would be –3dB at 1.6Hz. Lower source impedance will always extend low-frequency bandwidth. Since the filter is single pole, response falls at 6dB per octave. As discussed earlier, the permeability of most core material steadily increases as frequency is lowered and typically reaches its maximum somewhere under 1Hz. This results in an actual roll-off rate *less* than 6dB per octave and a corresponding improvement in phase distortion—deviation from linear phase. Although a transformer cannot have response to 0Hz or dc, it can have much less phase distortion than a coupling capacitor chosen for the same cutoff frequency. Or, as a salesperson might say, “It’s not a defect, it’s a feature.”

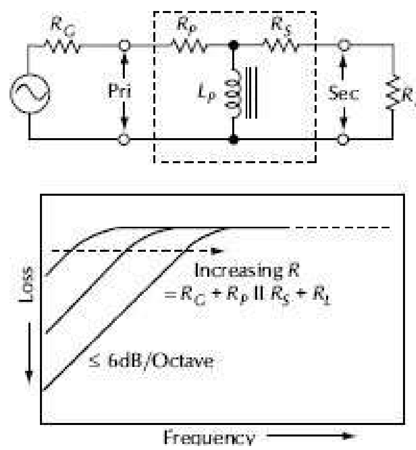


Figure 15-20. Simplified low frequency transformer equivalent circuit.

The simplified equivalent schematic of Fig. 15-21 shows the parasitic elements that limit and control high-frequency response.

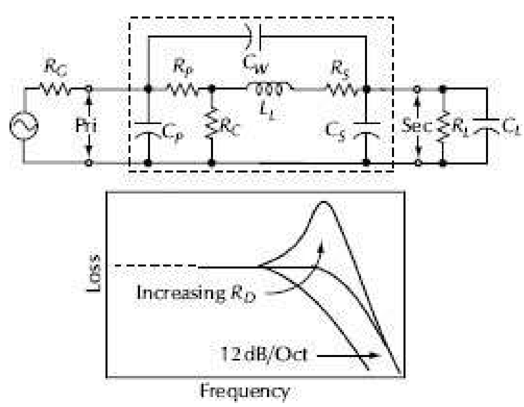


Figure 15-21. Simplified high-frequency transformer equivalent circuit.

Except in bi-filar wound types discussed below, leakage inductance L_L and load capacitance are the major limiting factors. This is especially applicable to layer-wound designs with in Faraday shields because of the increase in leakage inductance. Note that a low-pass filter is formed by series leakage inductance L_L with shunt winding capacitance C_S plus external load capacitance C_L. Since this filter has two reactive elements, it is a two-pole filter subject to response variations caused by *damping*. Resistive elements in a filter provide damping, dissipating energy when the inductive and capacitive elements resonate. As shown in the figure, if damping resistance R_D is too high, response will rise before it falls and if damping resistance is too low, response falls too early. Optimum damping results in the widest bandwidth with no response peak. It should be noted that placing capacitive loads C_L on transformers with high leakage inductance not only lowers their bandwidth but also changes the resistance required for optimum damping. For most transformers, R_L controls damping. In the time domain, under-damping manifests itself as ringing on square-waves as shown in Fig. 15-22. When loaded by its specified load resistance, the same transformer responds as shown in Fig. 15-23. In some transformers, source impedance also provides significant damping.

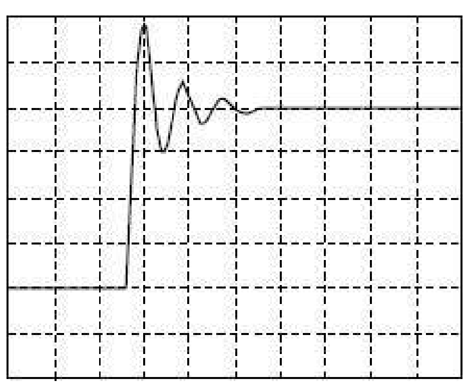


Figure 15-22. Undamped response.

In bi-filar wound transformers, leakage inductance L_L is very low but interwinding capacitance C_W and winding capacitances C_P and C_S are quite high. Leakage inductance must be kept very small in applications such as line drivers because large cable capacitances C_L would otherwise be disastrous to high-frequency response. Such transformers are generally referred to as output transformers. Also note that a low-pass filter is formed by series R_G and shunt C_P plus C_S. Therefore, driving sources may limit high-frequency response if their source impedance R_G is too high. In normal 1:1 bi-filar output transformer designs, C_W actually works to capacitively couple very high frequencies between windings. Depending on the application, this can be either a defect or a feature.

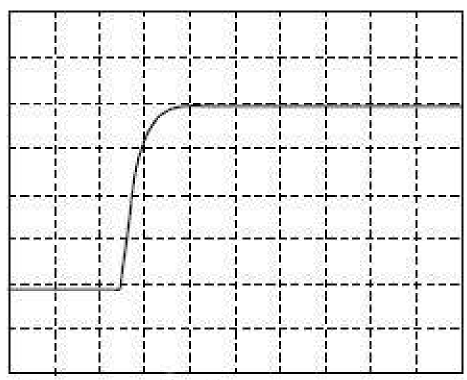


Figure 15-23. Proper damping.

15.1.3.3 Insertion Loss

The power output from a transformer will always be slightly less than power input to it. As current flows in its windings, their dc resistance causes additional voltage drops and power loss as heat. Broadly defined, *insertion loss* or *gain* is that caused by inserting a device into the signal path. But, because even an ideal lossless transformer can increase or decrease signal level by virtue of its turns ratio, the term *insertion loss* is usually defined as the difference in output signal level between the real transformer and an ideal one with the same turns ratio.

The circuit models, Thevenin equivalent circuits, and equations for both ideal and real transformers are shown in Fig. 15-24. For example, consider an ideal 1:1 turns ratio transformer and $R_G = R_L = 600\Omega$. Since N_s/N_p is 1, the equivalent circuit becomes simply E_i in series with R_G or 600Ω. When R_L is connected, a simple voltage divider is formed, making $E_O = 0.5E_i$ or a 6.02dB loss. For a real transformer having $R_P = R_S = 50\Omega$, the equivalent circuit becomes E_i in series with $R_G + R_P + R_S$ or 700Ω. Now, the output $E_O = 0.462E_i$ or a 6.72dB loss. Therefore, the insertion loss of the transformer is 0.70dB.

Calculations are similar for transformers with turns ratios other than 1:1, except that voltage is multiplied by the turns ratio and reflected impedances are multiplied by the turns ratio squared as shown in the equations. For example, consider a 2:1 turns ratio transformer, $R_G = 600\Omega$, and $R_L = 150\Omega$. The ideal transformer output appears as $0.5E_i$ in series with $R_G/4$ or 150Ω. When R_L is connected, a simple voltage divider is formed making $E_O = 0.25E_i$ or a 12.04dB loss. For a real transformer having $R_P = 50\Omega$ and $R_S = 25\Omega$, the equivalent circuit becomes $0.5\ E_i$ in series with $(R_G + R_P)/4 + R_S$ or 187.5Ω. Now, the output $E_O = 0.222E_i$ or a 13.07dB loss. Therefore, the insertion loss of this transformer is 1.03dB.

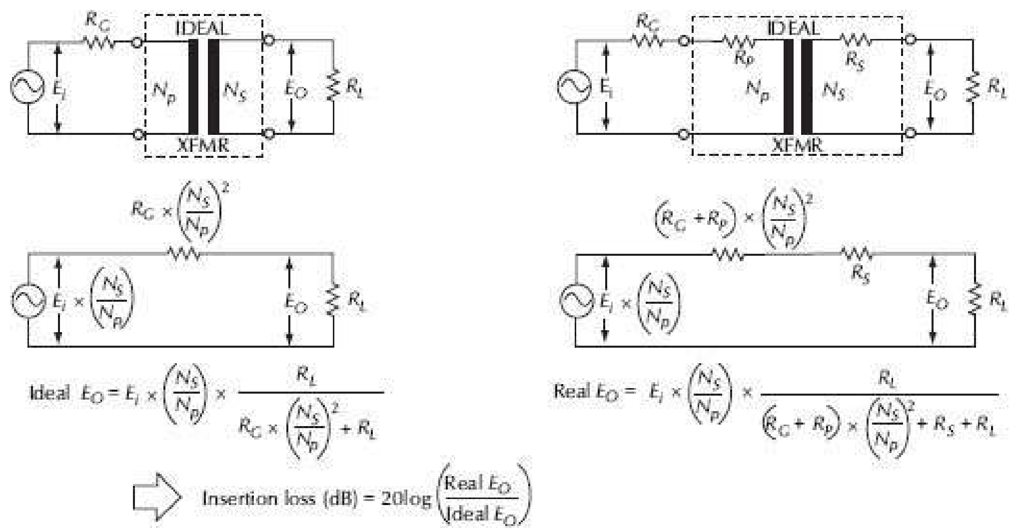


Figure 15-24. Insertion loss compares the outputs of real and ideal transformers.

15.1.3.4 Sources with Zero or Negative Impedance

One effect of using negative feedback around a high gain amplifier is to reduce its output impedance. Output impedance is reduced by the feedback factor, which is open-loop gain in dB minus closed-loop gain in dB. A typical op-amp with an open-loop gain of 80dB, set for closed-loop gain of 20dB, the feedback factor is 80dB – 20dB = 60dB or 1000, will have its open-loop output impedance of 50Ω reduced by the feedback factor (1000) to about 0.05Ω. Within the limits of linear operation—i.e., no current limiting or voltage clipping—the feedback around the amplifier effectively forces the output to remain constant regardless of loading. For all practical purposes the op-amp output can be considered a true voltage source.

As seen in Fig. 15-19, the distortion performance of any transformer is significantly improved when the driving source impedance is less than the dc resistance of the primary. However, little is gained for source impedances below about 10% of the winding dc resistance. For example, consider a typical line output transformer with a primary dc resistance of 40Ω. A driving source impedance well under 4Ω will result in lowest distortion. The line drivers shown in Fig. 15-28 and Fig. 15-29 use a paralleled inductor and resistor to isolate or decouple the amplifier from the destabilizing effects of load (cable) capacitance at very high frequencies. Because the isolator's impedance is well under an ohm at all audio frequencies, it is much preferred to the relatively large series, or build-out, resistor often used for the purpose.

It's possible for an amplifier to generate *negative* output resistance to cancel the winding resistance of the output transformer. As shown in Fig. 15-8, transformer distortion can be thought of as the non-linear behavior or R_C. If the effective sum of R_G and R_P become zero, R_C is effectively shorted and transformer distortion becomes extremely low. Audio Precision uses such a patented circuit in their System 1 audio generator to reduce transformer-related distortion to extremely low levels.

15.1.3.5 Bi-Directional Reflection of Impedances

The impedances associated with audio transformers may seem confusing. Much of the confusion probably stems from the fact that *transformers can simultaneously reflect two different impedances —one in each direction*. One is the impedance of the driving source, as seen from the secondary, and the other is the impedance of the load, as seen from the primary. Transformers simply reflect impedances, modified by the square of their turns ratio, from one winding to another. However, because of their internal parasitic elements discussed previously, transformers tend to produce optimum results when used within a specified range of external impedances.

There is essentially no intrinsic impedance associated with the transformer itself. With no load on its secondary, the primary of a transformer is just an inductor and its impedance will vary linearly with frequency. For example, a 5H primary winding would have an input impedance of about 3kΩ at 100Hz, 30kΩ at 1kHz, and 300kΩ at 10kHz. In a proper transformer design, this self-impedance, as well as those of other internal parasitics, should have negligible effects on normal circuit operation. The following applications will illustrate the point.

A 1:1 *output* transformer application is shown in Fig. 15-25. It has a winding inductance of about 25H and negligible leakage inductance. The open circuit impedance, at 1kHz, of either winding is about 150kΩ. Since the dc resistance is about 40Ω per winding, if the primary is short circuited, the secondary impedance will drop to 80Ω. If we place the transformer between a zero-impedance amplifier (more on that later) and a load, the amplifier will see the load through the transformer and the load will see the amplifier through the transformer. In our example, the amplifier would look like 80Ω to the output line/load and the 600Ω line/load would look like 680Ω to the amplifier. If the load were 20kΩ, it would look like slightly less than 20kΩ because the open circuit transformer impedance, 150kΩ at 1kHz, is effectively in parallel with it. For most applications, these effects are trivial.

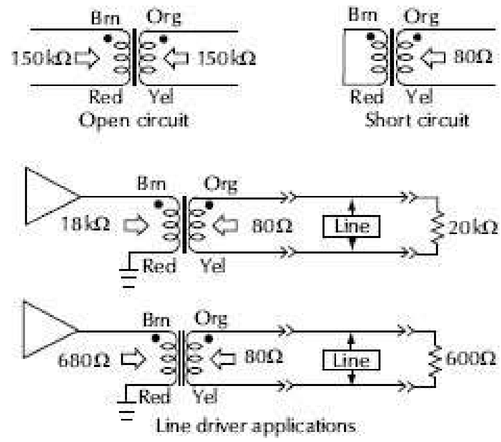


Figure 15-25. Impedance reflection in a 1:1 transformer.

A 4:1 *input* transformer example is shown in Fig. 15-26. It has a primary inductance of about 300H and negligible winding capacitance. The open circuit impedance, at 1kHz, of the primary is about 2MΩ. Because this transformer has a 4:1 turns ratio and, therefore a 16:1 impedance ratio, the secondary open circuit impedance is about 125kΩ. The dc resistances are about 2.5kΩ for the primary and 92Ω for the secondary. Since this is an input transformer, it must be used with the specified secondary load resistance of 2.43kΩ for proper damping (flat frequency response). This load on the secondary will be transformed by the turns ratio to look like about 42kΩ at the primary. To minimize the noise contribution of the amplifier stage, we need to know what the transformer secondary looks like, impedancewise, to the amplifier. If we assume that the primary is driven from the line in our previous output transformer example with its 80Ω source impedance, we can calculate that the secondary will look like about 225Ω to the amplifier input. Actually, any source impedance less than 1kΩ would have little effect on the impedance seen at the secondary.

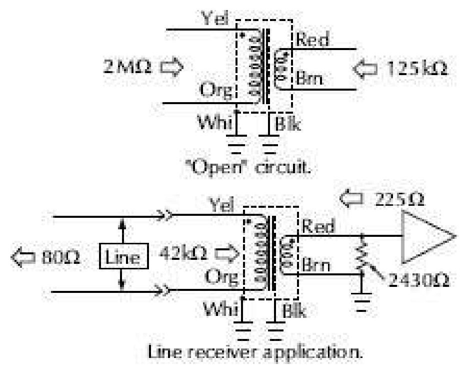


Figure 15-26. Impedance reflection in a 4:1 transformer.

Transformers are not intelligent—they can't magically couple signals in one direction only. Magnetic coupling is truly bi-directional. For example, Fig. 15-27 shows a three-winding 1:1:1 transformer connected to drive two 600Ω loads. The driver sees the loads in parallel or, neglecting winding resistances, 300Ω. Likewise, a short on either output will be reflected to the driver as a short. Of course, turns ratios and winding resistances must be taken into account to calculate actual driver loading. For the same reason, stereo L and R outputs that drive two windings on the same transformer are effectively driving each other, possibly causing distortion or damage.

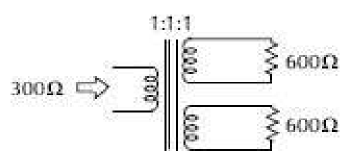
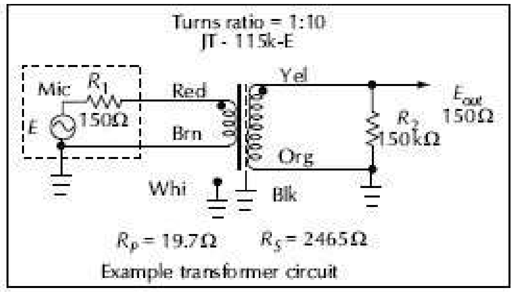
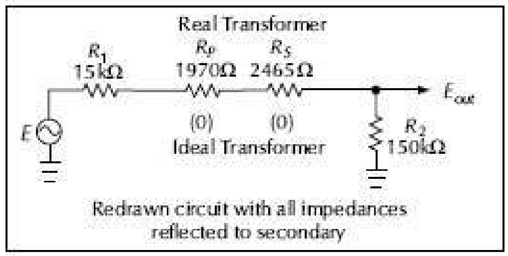


Figure 15-27. Multiple loads are effectively paralleled.

15.1.3.6 Transformer Noise Figure

Although the step-up turns ratio of a transformer may provide noise-free voltage gain, some 20dB for a 1:10 turns ratio, it's important to understand that improvements in signal-to-noise ratio (*SNR*) are not solely due to this gain. Because most amplifying devices generate current noise as well as voltage noise at their inputs, their noise performance will suffer when turns ratio is not the optimum for that particular amplifier (see Chapter 25, *Preamplifiers and Mixers* section 25.1.2.3 *Noise*). *Noise figure* measures, in dB, how much the output signal-to-noise ratio of a system is degraded by a given system component. All resistances, including the winding resistances of transformers, generate thermal noise. Therefore, the noise figure of a transformer indicates the increase in thermal noise or hiss when it replaces an ideal noiseless transformer having the same turns ratio, i.e., voltage gain. The noise figure of a transformer is calculated as shown in Fig. 15-28.





Primary side impedances get multiplied by the square of the turns ratio.

$R_1 = 150\Omega \times 10^2 = 15\text{k}\Omega$

$R_P = 19.7\Omega \times 10^2\Omega = 1970\Omega$

R_S = Secondary DCR = 2465Ω

R_2 = 150kΩ (load)

The transformer noise figure is calculated by comparing a real transformer with its winding resistances to an ideal transformer with no winding resistances.
First, transform all impedances to the secondary as shown to the left.

There are two components to the calculation.

1. The additional noise due to the increased output impedance.

$$\text{REAL } Z_{out} \quad \frac{150\text{k}\Omega \times (15\text{k}\Omega + 1970 + 2465)}{150\text{k} + 15\text{k} + 1970 + 2465} = 17.205\text{k}\Omega$$

$$\text{IDEAL } Z_{out} \quad \frac{1150\text{k}\Omega + 15\text{k}\Omega}{50\text{k}\Omega \times 15\text{k}\ \Omega} = 13.636\text{k}\Omega$$

$$NF = 20\text{Log}\sqrt{\frac{17{,}206\ (REAL)}{13{,}636\ (IDEAL)}} = 1.01\text{ dB}$$

2. The decrease in signal level at the output due to the increased series losses.

$$IDEAL\ E_{out} \quad \frac{150\text{k}\Omega}{150\text{k}\Omega + 15\text{k}\Omega} = 0.909$$

$$REAL_{out} \quad \frac{150\text{k}\Omega}{150\text{k}\Omega + 15\text{k}\Omega + 1970 + 2465} = 0.885$$

$$NF = 20\text{Log}\frac{0.909\ (IDEAL)}{0.885\ (REAL)} = 0.232\text{dB}$$

3. *Total NF* = 1.01dB + 0.232dB = 1.23 dB

Figure 15-28. Finding the noise figure of a transformer.

15.1.3.7 Basic Classification by Application

Many aspects of transformer performance, such as level handling, distortion, and bandwidth, depend critically on the impedance of the driving source and, in most cases, the resistance and capacitance of the load. These impedances play such an important role that they essentially classify audio transformers into two basic types. Most simply stated, *output* transformers are used when load impedances are low, as in line drivers, while *input* transformers are used when load impedances are high, as in line receivers. The load for a line driver is not just the high-impedance equipment input it drives—it also includes the cable capacitance, whose impedance can become quite low at 20kHz. The conflicting technical requirements for output and input types make their design and physical construction very different. Of course, some audio transformer applications need features of both input and output transformers and are not so easily classified.

Output transformers must have very low leakage inductance in order to maintain high-frequency bandwidth with capacitive loads. Because of this, they rarely use Faraday shields and are most often multi-filar wound. For low insertion loss, they use relatively few turns of large wire to decrease winding resistances. Since they use fewer turns and operate at relatively high signal levels, output transformers seldom use magnetic shielding. On the other hand, input transformers directly drive the usually high-resistance, low-capacitance input of amplifier circuitry. Many input transformers operate at relatively low signal levels, frequently have a Faraday shield, and are usually enclosed in at least one magnetic shield.

15.2 Audio Transformers for Specific Applications

Broadly speaking, audio transformers are used because they have two very useful properties. First, they can benefit circuit performance by transforming circuit impedances, to optimize amplifier noise performance, for example. Second, because there is no direct electrical connection between its primary and secondary windings, a transformer provides electrical or galvanic *isolation* between two circuits. As discussed in Chapter 36, *Grounding and Interfacing*, isolation in signal circuits is a powerful technique to prevent or cure noise problems caused by normal ground voltage differences in audio systems. To be truly useful, a transformer should take full advantage of one or both of these properties but not compromise audio performance in terms of bandwidth, distortion, or noise.

15.2.1 Equipment-Level Applications

15.2.1.1 Microphone Input

A microphone input transformer is driven by the nominal 150Ω, or 200Ω in Europe, source impedance of professional microphones. One of its most important functions is to transform this impedance to a generally higher one more suited to optimum noise performance. As discussed in Chapter 25, *Preamplifiers and Mixers*, this optimum impedance may range from 500Ω to over 15kΩ, depending on the amplifier. For this reason, microphone input transformers are made with turns ratios ranging from 1:2 to 1:10 or higher. The circuit of Fig. 15-29 uses a 1:5 turns ratio transformer, causing the microphone to appear as a 3.7kΩ source to the IC amplifier, which optimizes its noise. The input impedance of the transformer is about 1.5kΩ. It is important that this impedance remain reasonably flat with frequency to avoid altering the microphone response at frequency extremes, see Chapter 25, *Preamplifiers and Mixers*, Fig. 25-6.

In all balanced signal connections, common-mode noise can exist due to ground voltage differences or magnetic or electrostatic fields acting on the interconnecting cable. It is called common mode noise because it appears equally on the two signal lines, at least in theory. Perhaps the most important function of a balanced input is to reject (not respond to) this common-mode noise. A figure comparing the ratio of its differential or normal signal response to its common-mode response is called *common mode rejection ratio* or CMRR. An input transformer must have two attributes to achieve high CMRR. First, the capacitances of its two inputs to ground must be very well matched and as low as possible. Second, it must have minimal capacitance between its primary and secondary windings. This is usually accomplished by precision winding of the primary to evenly distribute capacitances and the incorporation of a Faraday shield between primary and secondary. Because the common-mode input impedances of a transformer consist only of capacitances of about 50 pF, transformer CMRR is maintained in real-world systems where the source impedances of devices driving the balanced line and the capacitances of the cable itself are not matched with great precision.[3]

Because tolerable common-mode voltage is limited only by winding insulation, transformers are well suited for phantom power applications. The standard arrangement using precision resistors is shown in Fig. 15-29. Resistors of lesser precision may degrade CMRR. Feeding phantom power through a center tap on the primary requires that both the number of turns and the dc resistance on either side of the tap be precisely matched to avoid small dc offset voltages across the primary. In most practical transformer designs, normal tolerances on winding radius and wire resistance make this a less precise method than the resistor pair. Virtually all microphone input transformers will require loading on the secondary to control high-frequency response. For the circuit in the figure, network R_1, R_2, and C_1 shape the high-frequency response to a Bessel roll-off curve. Because they operate at very low signal levels, most microphone input transformers also include magnetic shielding.

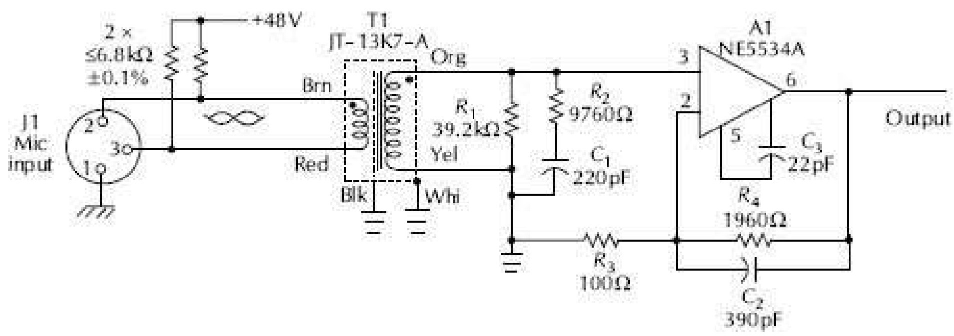


Figure 15-29. Microphone preamplifier with 40dB overall gain.

15.2.1.2 Line Input

A line input transformer is driven by a balanced line and, most often, drives a ground-referenced (unbalanced) amplifier stage. As discussed in Chapter 36, *Grounding and Interfacing*, modern voltage-matched interconnections require that line inputs have impedances of 10kΩ or more, traditionally called *bridging* inputs. In the circuit of Fig. 15-30, a 4:1 step-down transformer is used which has an input impedance of about 40kΩ.

High CMRR is achieved in line input transformers using the same techniques as for microphone input transformers. Again, because its common-mode input impedances consist of small capacitances, a good input transformer will exhibit high CMRR even when signal sources are real-world equipment with less-than-perfect output impedance balance. The dirty little secret of most electronically balanced input stages, especially simple differential amplifiers, is that they are very susceptible to tiny impedance imbalances in driving sources. However, they usually have impressive CMRR figures when the driving source is a laboratory generator. The pitfalls of measurement techniques will be discussed in section 15.3.1.

As with any transformer having a Faraday shield, line input transformers have significant leakage inductance and their secondary load effectively controls their high-frequency response characteristics. The load resistance or network recommended by the manufacturer should be used to achieve specified bandwidth and transient response. Input transformers are intended to immediately precede an amplifier stage with minimal input capacitance. Additional capacitive loading of the secondary should be avoided because of its adverse effect on frequency and phase response. For example, capacitive loads in excess of about 100pF—about 3ft of standard shielded cable—can degrade performance of a standard 1:1 input transformer.

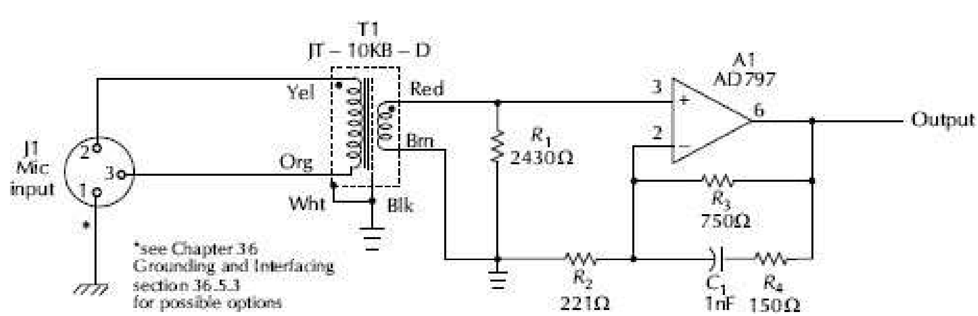


Figure 15-30. Low-noise unity-gain balanced line input stage.

15.2.1.3 Moving-Coil Phono Input

Moving-coil phonograph pickups are extremely low-impedance, very low-output devices. Some of them have source impedances as low as 3Ω, making it nearly impossible to achieve optimum noise performance in an amplifier. The transformer shown in Fig. 15-31 has a three-section primary that can be series-connected as a 1:4 step-up for 25Ω to 40Ω devices and parallel-connected as a 1:12 step-up for 3Ω to 5Ω devices. In either case, the amplifier sees a 600Ω source impedance that optimizes low-noise operation. The transformer is packaged in double magnetic shield cans and has a Faraday shield. The loading network R_1, R_2, and C_1 tailor the high-frequency response to a Bessel curve.

15.2.1.4 Line Output

A line-level output transformer is driven by an amplifier and typically loaded by several thousand pF of cable capacitance plus the 20kΩ input impedance of a balanced bridging line receiver. At high frequencies, most driver output current is actually used driving the cable capacitance. Sometimes, terminated 150Ω or 600Ω lines must be driven, requiring even more driver output current. Therefore, a line output transformer must have a low output impedance that stays low at high frequencies. This requires both low resistance windings and very low leakage inductance, since they are effectively in series between amplifier and load. To maintain impedance balance of the output line, both driving impedances and inter-winding capacitances must be well matched at each end of the windings. A typical bi-filar-wound design has winding resistances of 40Ω each, leakage inductance of a few micro-henries, and a total inter-winding capacitance of about 20nF matched to within 2%

across the windings.

The high-performance circuit of Fig. 15-32 uses op-amp A_1 and current booster A_2 in a feedback loop setting overall gain at 12dB. A_3 provides the high gain for a dc servo feedback loop used to keep dc offset at the output of A_2 under 100μV. This prevents any significant dc flow in the primary of transformer T_1. X_1 provides capacitive load isolation for the amplifier and X_2 serves as a tracking impedance to maintain high-frequency impedance balance of the output. High-conductance diodes D_1 and D_2 clamp inductive kick to protect A_2 in case an unloaded output is driven into hard clipping.

The circuit of Fig. 15-33 is well suited to the lower signal levels generally used in consumer systems. Because its output floats, it can drive either balanced or unbalanced outputs, but not at the same time. Floating the unbalanced output avoids ground loop problems that are inherent to unbalanced interconnections.

In both previous circuits, because the primary drive of T_1 is single-ended, the voltages at the secondary will *not* be symmetrical, especially at high frequencies. THIS IS NOT A PROBLEM. Contrary to widespread myth and as explained in Chapter 36, *Grounding and Interfacing*, signal symmetry has absolutely nothing to do with noise rejection in a balanced interface! Signal symmetry in this, or any other floating output, will depend on the magnitude and matching of cable and load impedances to ground. If there is a requirement for signal symmetry, the transformer should be driven by dual, polarity-inverted drivers.

The circuit of Fig. 15-34 uses a cathode follower circuit that replaces the usual resistor load in the cathode with an active current sink. The circuit operates at quiescent plate currents of about 10mA and presents a driving source impedance of about 60Ω to the transformer, which is less than 10% of its primary dc resistance. C_2 is used to prevent dc flow in the primary. Since the transformer has a 4:1 turns ratio, or 16:1 impedance ratio, a 600Ω output load is reflected back to the driver circuit as about 10kΩ. Since the signal swings on the primary are four times as large as those on the secondary, high-frequency capacitive coupling is prevented by a Faraday shield. The secondary windings may be parallel connected to drive a 150Ω load. Because of the Faraday shield, output winding capacitances are low and the output signal symmetry will be determined largely by the balance of line and load impedances.

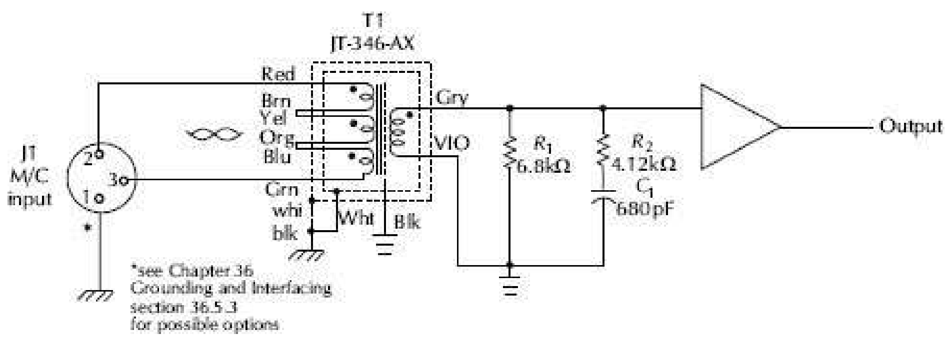


Figure 15-31. Preamp for 25Ω moving-coil pickups.

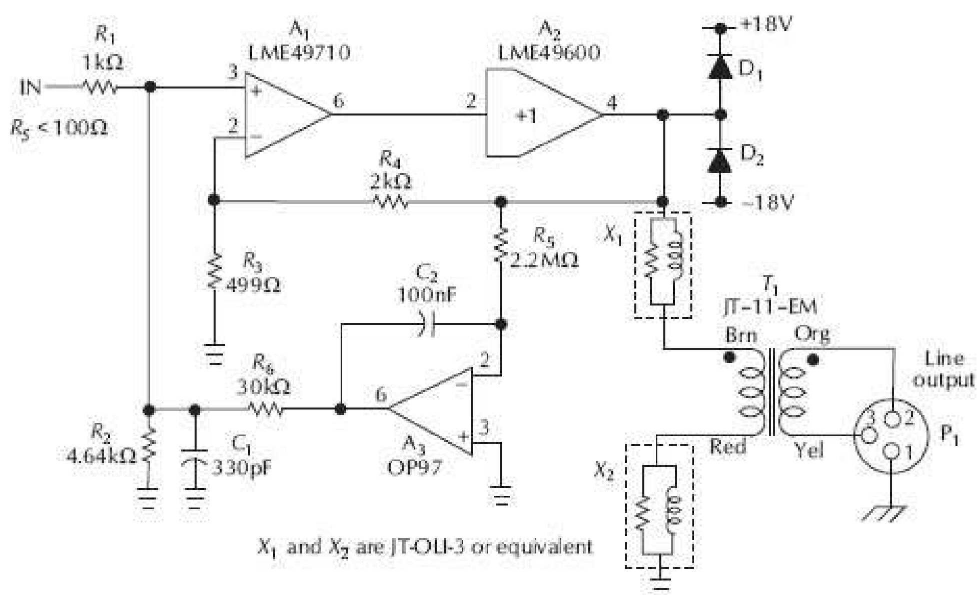


Figure 15-32. Typical line output application circuit.

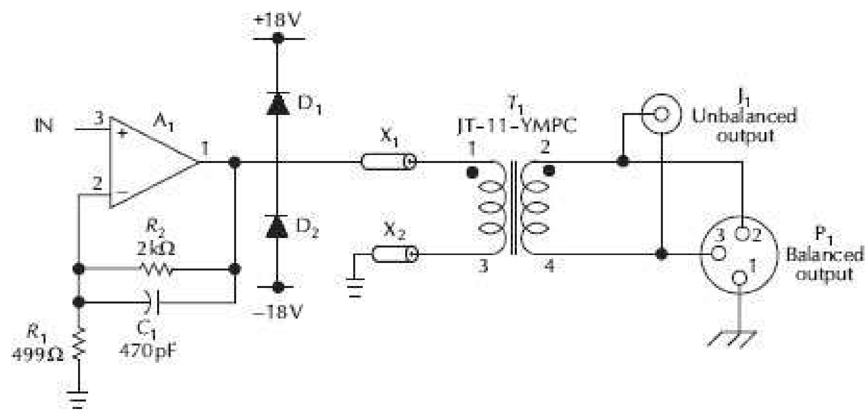


Figure 15-33. Universal isolated output application.

15.2.1.5 Inter-Stage and Power Output

Inter-stage coupling transformers are seldom seen in contemporary equipment but were once quite popular in vacuum-tube amplifier designs. They typically use turns ratios in the 1:1 to 1:3 range and, as shown in Fig. 15-35, may use a center-tapped secondary producing phase-inverted signals to drive a push-pull output stage. Because both plate and grid circuits are relatively high impedance, windings are sometimes section-wound to reduce capacitances. Resistive loading of the secondary is usually necessary both to provide damping and to present a uniform load impedance to the driving stage. Although uncommon, inter-stage transformers for solid-state circuitry are most often bi-filar wound units similar to line output designs.

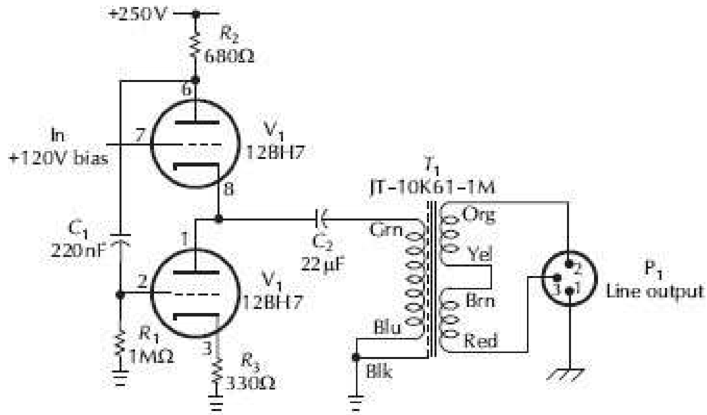


Figure 15-34. Double cathode-follower line driver.

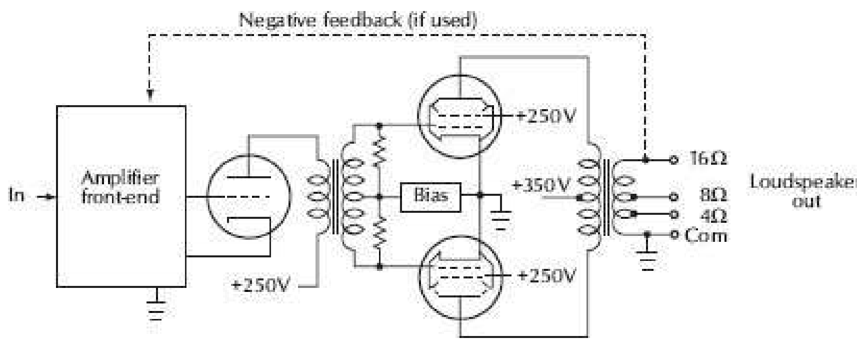


Figure 15-35. Push-pull vacuum-tube power amplifier.

The classic push-pull *power output* stage, with many variations over the years, has been used in hi-fi gear, PA systems, and guitar amplifiers. The turns ratio of the output transformer is generally chosen for a reflected load at the tubes of several thousand ohms plate-to-plate. A typical 30:1 turns ratio may require many interleaved sections to achieve bandwidth extending well beyond 20kHz.

If the quiescent plate currents and the number of turns in each half of the primary winding are matched, magnetic flux in the core will cancel at dc. Since any current-balancing in vacuum-tubes is temporary at best, these transformers nearly always use steel cores to better tolerate this unbalanced dc in their windings. The relatively high driving impedance of the tube plates results in considerable transformer related distortion. To reduce distortion, feedback around the transformer is often employed. To achieve stability (freedom from oscillation), very wide bandwidth (actually low phase shift) is required of the transformer when a feedback loop is closed around it. As a result, some of these output transformer designs are very sophisticated. Some legendary wisdom suggests as a rough guide that a good-fidelity output transformer should have a core weight and volume of at least 0.34lbs and 1.4in^3 respectively per watt of rated power.[4]

A single-ended power amplifier is created by removing the lower tube and the lower half of the transformer primary from the circuit of Fig. 15-35. Now plate current will create a strong dc field in the core. As discussed in section 15.1.2.1, the core will likely require an air gap to avoid magnetic saturation. The air gap reduces inductance, limiting low-frequency response, and increases even-order distortion products. Such a single-ended pentode power amplifier was ubiquitous in classic AM 5-tube table radios of the fifties and sixties.

15.2.1.6 Microphone Output

There are two basic types of output transformers used in microphones, step-up and step-down. In a ribbon microphone, the ribbon element may have an impedance of well under 1Ω, requiring a step-up transformer with a turns ratio of 1:12 or more to raise its output level and make its nominal output impedance around 150Ω. Typical dynamic elements have impedances from 10Ω to 30Ω, which require step-up turns ratios from 1:2 to 1:4. These step-up designs are similar to line output transformers in that they have no

Faraday or magnetic shields, but are smaller because they operate at lower signal levels.

A condenser microphone has integral circuitry to buffer and/or amplify the signal from its extremely high-impedance transducer. Since this low-power circuitry operates from the phantom supply, it may be unable to directly drive the 1.5kΩ input impedance of a typical microphone preamp. The output transformer shown in Fig. 15-36, which has an 8:1 step-down ratio, will increase the impedance seen by Q_1 to about 100kΩ. Because of its high turns ratio, a Faraday shield is used to prevent capacitive coupling of the primary signal to the output.

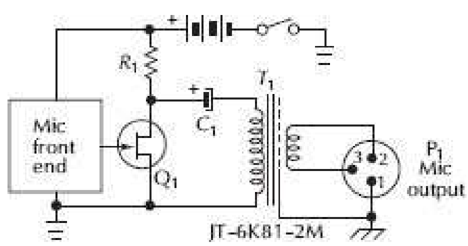


Figure 15-36. Condenser microphone output transformer.

15.2.2 System-Level Applications

15.2.2.1 Microphone Isolation or Splitter

The primary of a transformer with a 1:1 turns ratio can bridge the output of a 150Ω to 200Ω microphone feeding one pre-amp and the secondary of the transformer can feed a duplicate of the microphone signal to another pre-amp. Of course, a simple Y cable could do this but there are potential problems. There are often large and noisy voltages between the grounds of two pre-amplifiers. The isolation provided by the transformer prevents the noise from coupling to the balanced signal line. To reduce capacitive noise coupling, Faraday shields are included in better designs and double Faraday shields in the best. As discussed in section 15.1.3.5, the input impedances of all the pre-amps, as well as all the cable capacitances, will be seen in parallel by the microphone. This places a practical upper limit on how many ways the signal can be split. Transformers are commercially available in 2, 3, and 4-winding versions. A 3-way splitter box schematic is shown in Fig. 15-37. Since the microphone is *directly* connected only to the direct output, it is the only one that can pass phantom power to the microphone. To each preamp, each isolated output looks like a normal floating (ungrounded) microphone. The ground lift switches are normally left open to prevent possible high ground current flow in the cable shields. See reference 7 for a detailed discussion and a real-world microphone splitter design example

15.2.2.2 Microphone Impedance Conversion

Some legacy dynamic microphones are high-impedance, about 50kΩ, and have two-conductor cable and connector (unbalanced). When such a microphone must be connected to a standard balanced low-impedance microphone pre-amp, a transformer with a turns ratio of about 15:1 is necessary. Similar transformers can be used to adapt a low-impedance microphone to the unbalanced high-impedance input of a legacy pre-amplifier. Commercial products are available which enclose such a transformer in the XLR adapter barrel.

15.2.2.3 Line to Microphone Input or Direct Box

Because its high-impedance, unbalanced input accepts line-level signals and its output drives the low-level, low-impedance balanced microphone input of a mixing console, the device shown in Fig. 15-38 is called a direct box. It is most often driven by an electric guitar, synthesizer, or other stage instrument. Because it uses a transformer, it provides ground isolation as well. In this typical circuit, since the transformer has a 12:1 turns ratio, the impedance ratio is 144:1. When the microphone input has a typical 1.5kΩ input impedance, the input impedance of the direct box is about 200kΩ. The transformer shown has a separate Faraday shield for each winding to minimize capacitively coupled ground noise.

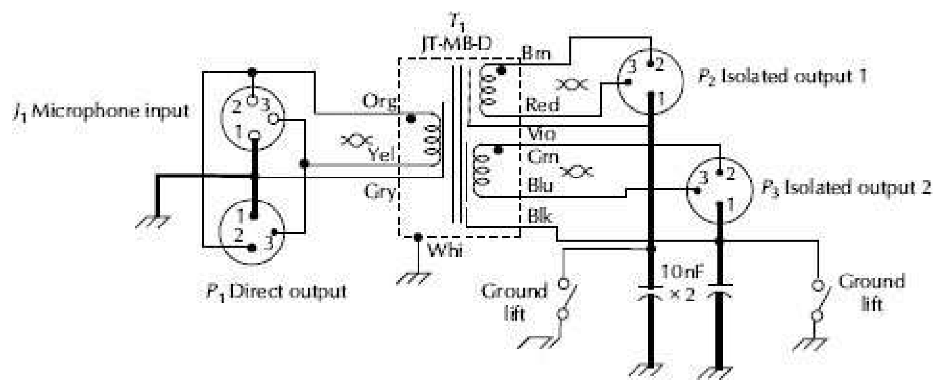


Figure 15-37. A 3-way microphone splitter box.

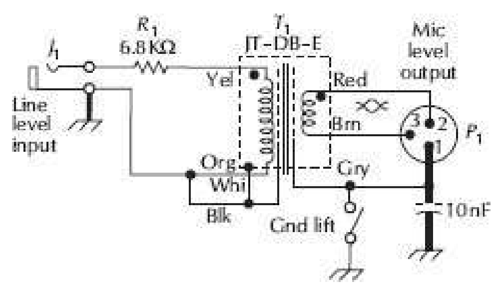


Figure 15-38. A transformer-isolated direct box.

15.2.2.4 Line Isolation or Hum Eliminators

There are a remarkable number of black boxes on the market intended to solve ground loop problems. This includes quite a number of transformer-based boxes. With rare exception, those boxes contain *output* transformers. Tests were performed to compare noise rejection of the original interface to one with an added output transformer and to one with an added input transformer. The tests accurately simulated typical real-world equipment, see the definitions at the end of this section.

Fig. 15-39 shows results of CMRR tests on a ***balanced*** interface using the IEC 60268-3 test procedure discussed in section 15.3.1.2. This test recognizes that the impedances of real-world balanced outputs are not matched with the precision of laboratory equipment. While the output transformer reduces 60Hz hum by over 20dB, it has little effect on buzz artifacts over about 1kHz. The input transformer increases rejection to over 120dB at 60Hz and to almost 90dB at 3kHz, where the human ear is most sensitive to faint sounds.

Fig. 15-40 shows results of ground noise rejection tests on an ***unbalanced*** interface. By definition, there is 0dB of inherent rejection in an unbalanced interface, see Chapter 36, *Grounding and Interfacing*. While the output transformer reduces 60Hz hum by about 70dB, it reduces buzz artifacts around 3kHz by only 35dB. The input transformer increases rejection to over 100dB at 60Hz and to over 65dB at 3kHz.

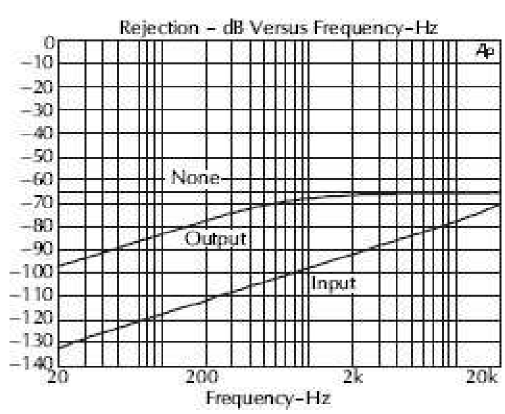


Figure 15-39. Balanced output to balanced input.

Fig. 15-41 shows results of CMRR tests when an ***unbalanced*** output drives a ***balanced*** input. A two-wire connection of this interface will result in zero rejection, see Chapter 36, *Grounding and Interfacing*. Assuming a three-wire connection, the –30dB plot shows how CMRR of typical electronically-balanced input stages is degraded by the 600Ω source imbalance. Again, the output transformer improves 60Hz hum by over 20dB, it has little effect on buzz artifacts over about 1kHz. The input transformer increases rejection to almost 100dB at 60Hz and to about 65dB at 3kHz.

Fig. 15-42 shows results of ground noise rejection tests when a ***balanced*** output drives an ***unbalanced*** input. Because our balanced output does not float, the direct connection becomes an unbalanced interface having, by definition, 0dB of rejection. While the output transformer reduces 60Hz hum by about 50dB, it reduces buzz artifacts around 3kHz by less than 20dB. The input transformer increases rejection to almost 100dB at 60Hz and to about 65dB at 3kHz. In this application it is usually desirable to attenuate the signal by about 12dB, from +4dBu or 1.228V to –10dBV or 0.316V, as well as provide ground isolation. This can be conveniently done by using a 4:1 step-down input transformer such as the one in Fig. 15-29, which will produce rejection comparable to that shown here.

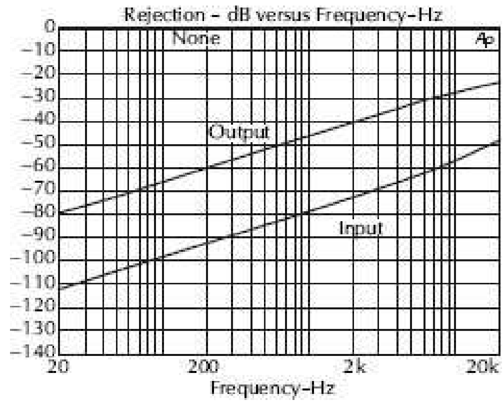


Figure 15-40. Unbalanced output to unbalanced input.

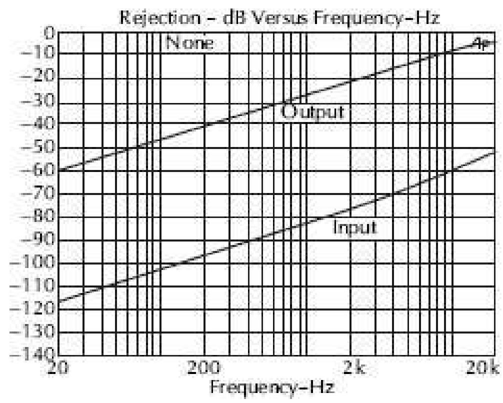


Figure 15-41. Unbalanced output to balanced input.

One might fairly ask "Why not use a 1:4 step-up transformer when an ***unbalanced*** output drives a ***balanced*** input to get 12dB of signal gain?" Because of the circuit impedances involved, the answer is because it doesn't work very well. Recall that a 1:4 turns ratio has an impedance ratio of 1:16. This means that the input impedance of the pro balanced input we drive will be reflected back to the consumer output at one-sixteenth that. Since the source impedance—usually unspecified, but not the same as *load* impedance—of a consumer outputs is commonly 1kΩ or more, the reflected loading losses are high. A 1:4 step-up transformer would have its own insertion losses, which we will rather optimistically assume at 1dB. The table below shows actual gain using this transformer with some typical equipment output and input impedances (*Z* is impedance). Since IEC standards require only that consumer outputs have a source impedance of 2.2k£ or less, matters could be even worse than shown in Table 15-1, which stops at 1kΩ

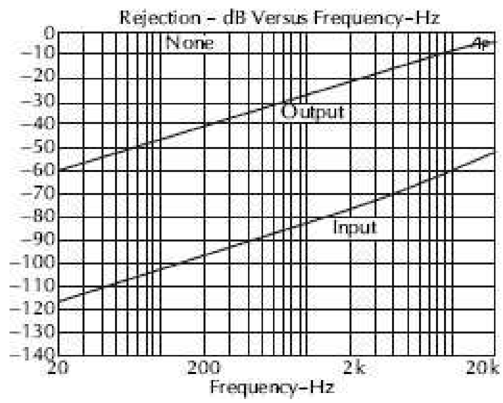


Figure 15-42. Balanced output to unbalanced input.

Table 15-1. Gain Derived from a 1:4 Step-up Transformer in Typical Circuits

Consumer Output Z	Pro Balanced Input Z		
	10kΩ (625Ω)	**20kΩ (1.25kΩ)**	**40kΩ (2.5kΩ)**
200Ω	8.6dB	9.7dB	10.3dB
500Ω	5.9dB	8.1dB	9.4dB
1kΩ	2.7dB	5.9dB	8.1dB

Not only will gain usually be much less than 12dB, the load reflected back to the consumer output, shown in parentheses in the table, is far too low and will likely cause high distortion, loss of headroom, and poor low-frequency response. Often the only specification of a consumer output is 10kΩ minimum load IEC standards require consumer inputs to have input impedances of

22kΩ or more. It is futile to increase the turns ratio of the transformer in an attempt to overcome the gain problem—it only makes the reflected loading problems worse. In most situations, a 1:1 transformer can be used because the pro equipment can easily provide the required gain. Of course, a 1:1 *input* transformer will provide far superior noise immunity from ground loops as well.

The point here is that the noise rejection provided by an input transformer with a Faraday-shield is <u>far</u> superior to that provided by an output type. But the input transformer <u>must</u> be used at the receiver or destination end of an interface cable. In general, input transformers should drive no more than three feet of typical shielded cable—the capacitance of longer cables will erode their high-frequency bandwidth. Although output type transformers without a Faraday shield are not as good at reducing noise, their advantage is that they can be placed anywhere along an interface cable, at the driver end, at a patch-bay, or at the destination end, and work equally well (or poorly, compared to an input transformer). In all the test cases discussed in this section, results of using *both* an output and an input transformer produced results *identical* to those using only an input transformer. For example, an unbalanced output does not need to be balanced by a transformer before transmission through a cable (this is a corollary of the balance versus symmetry myth), it needs only an input transformer at the receiver. There is rarely a need to ever use both types on the same line.

Definitions (in context of comparison tests only):

Balanced Output. ”A normal, non-floating source having a differential output impedance of 600Ω and common-mode output impedances of 300Ω, matched to within ±0.1%.

Balanced Input. A typical electronically-balanced stage—an instrumentation circuit using 3 op-amps—having a differential input impedance of 40kΩ and common-mode input impedances of 20kΩ, trimmed for a CMRR over 90dB when directly driven by the above Balanced Output.

Unbalanced Output. A ground-referenced output having an output impedance of 600Ω. This is representative of typical consumer equipment.

Unbalanced Input. A ground-referenced input having an input impedance of 50kΩ. This is representative of typical consumer equipment.

No Transformer. A direct wired connection.

Output Transformer. A Jensen JT-11-EMCF—a popular 1:1 line output transformer.

Input Transformer. A Jensen JT-11P-1—the most popular 1:1 line input transformer.

15.2.2.5 Loudspeaker Distribution or Constant Voltage

When a number of low-impedance loudspeakers are located far from a power amplifier, there are no good methods to interconnect them in a way that properly loads the amplifier. The problem is compounded by the fact that power losses due to the resistance of the inter-connecting wiring can be substantial. The wire gauge required is largely determined by the current it must carry and its length. Borrowing a technique from power utility companies, boosting the distribution voltage reduces the current for a given amount of power and allows smaller wire to be used in the distribution system. *Step-down* matching transformers, most often having taps to select power level and/or loudspeaker impedance, are used at each location. This scheme not only reduces the cost of wiring but allows system designers the freedom to choose how power is allocated among the speakers. These so-called constant-voltage loudspeaker distribution systems are widely used in public address, paging, and background music systems. Although the most popular is 70V, others include 25V, 100V, and 140V. Because the higher voltage systems offer the lowest distribution losses for a given wire size, they are more common in very large systems. It should also be noted that only the 25V system is considered low-voltage by most regulatory agencies and the wiring in higher voltage systems may need to conform to power wiring practices.

It is important to understand that these nominal voltages exist on the distribution line only when the driving amplifier is operating at full rated power. Many specialty power amplifiers have outputs rated to drive these lines directly but ordinary power amplifiers rated to drive speakers can also drive such lines, according to Table 15-2.

Table 15-2. Amplifier Power Required at Various Impedances Versus Output Voltage

Amplifier Rated Output, Watts			**Output Voltage**
at 8Ω	at 4Ω	at 2Ω	
1250	2500	5000	100
625	1250	2500	70.7
312	625	1250	50

156	312	625	35.3
78	156	312	25

For example, an amplifier rated to deliver 1,250W of continuous average power into an 8Ω load will drive a 70V distribution line directly as long as the sum of the power delivered to all the loudspeakers doesn't exceed 1,250W. Although widely used, the term *rms watts* is technically ambiguous.[5] In many cases, the benefits of constant-voltage distribution are desired, but the total power required is much less. In that case a *step-up* transformer can be used to increase the output voltage of an amplifier with less output. This is often called matching it to the line because such a transformer is actually transforming the equivalent line impedance down to the rated load impedance for the amplifier. Most of these step-up transformers will have a low turns ratio. For example, a 1:1.4 turns ratio would increase the 50V output to 70V for an amplifier rated at 300W into 8Ω. In such low-ratio applications, the *auto-transformer* discussed in section 15.1.2.2 has cost and size advantages. Fig. 15-43 is a schematic of an auto-transformer with taps for turns ratios of 1:1.4 or 1:2 which could be used to drive a 70V line from amplifiers rated for either 300 or 150W respectively at 8Ω. Several power amplifier manufacturers offer such transformers as options or accessories.

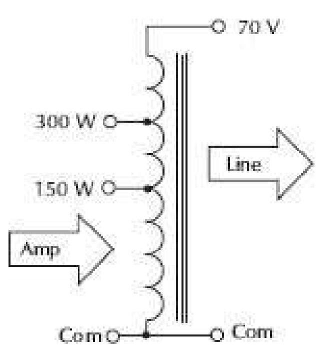


Figure 15-43. Step-up auto-transformer

Table 15-3. Transformer Step-Down Turns Ratio Required to Produce the Desired Loudspeaker Power

Speaker Power in Watts			Loudspeaker	Transformer Step-Down Turns Ratio Required			
16Ω	8Ω	4Ω	Volts	100V	70V	35V	25V
32	64	128	22.63	4.42	3.12	1.56	1.10
16	32	64	16.00	6.25	4.42	2.21	1.56
8	16	32	11.31	8.84	6.25	3.12	2.21
4	8	16	8.00	12.50	8.84	4.42	3.12
2	4	8	5.66	17.70	12.50	6.25	4.42
1	2	4	4.00	25.00	17.70	8.84	6.25
0.5	1	2	2.83	35.30	25.00	12.50	8.84
0.25	0.5	1	2.00	50.00	35.30	17.70	12.50
0.125	0.25	0.5	1.41	71.00	50.00	25.00	17.70

A line to voice-coil transformer is usually necessary to step-down the line voltage and produce the desired loudspeaker power, Table 15-3.

These step-down transformers can be designed several ways. Fig. 15-44 shows a design where the line voltage is selected at the primary side and the power level is selected at the secondary while Fig. 15-45 shows a design where power level is selected on the primary side and loudspeaker impedance is selected at the secondary.

As may be seen from the repeating patterns in the table above, there are many combinations of line voltage, loudspeaker impedance, and power level that result in the same required turns ratio in the matching transformer.

Since the constant-voltage line has a very low source impedance, and the transformer is loaded by a low-impedance loudspeaker, transformer high-frequency response is usually not a design issue. As in any transformer, low-frequency response is determined by primary inductance and total source impedance, which is dominated by the primary winding resistance since the driving source impedance is very low. Winding resistances of both primary and secondary contribute to insertion loss. In efforts to reduce size and cost, the fewest turns of the smallest wire possible are often used, which raises insertion loss and degrades low-frequency response. Generally, an insertion loss of 1dB or less is considered good and 2dB is marginally acceptable for these applications.

It is very important to understand that, while the low-level frequency response of a transformer may be rated as −1dB at 40Hz, its rated power does NOT apply at that frequency. Rated power, or maximum signal level is discussed in section 15.1.3.1. In general, level handling is increased by more primary turns and more core material and it takes more of both to handle more power at lower frequencies. This ultimately results in physically larger, heavier, and more expensive transformers. When any transformer is driven at its rated level at a lower frequency than its design will support, magnetic core saturation is the result. The sudden drop in permeability of the core effectively reduces primary inductance to zero. The transformer primary now appears to have only the dc resistance of its winding, which may be only a fraction of an ohms. In the best scenario, some ugly-sounding distortion will occur and the line amplifier will simply current limit. In the worst scenario, the amplifier will not survive the inductive energy fed back as the transformer comes out of saturation. This can be especially dangerous if large numbers of transformers saturate simultaneously.

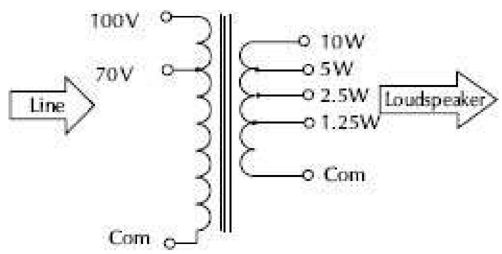


Figure 15-44. Transformer with secondary taps for power selection.

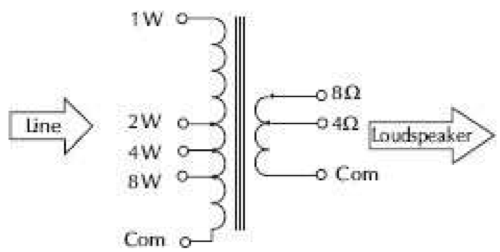


Figure 15-45. Transformer with primary taps for power selection.

In 1953, the power ratings of loudspeaker matching transformers were based on 2% distortion at 100Hz.[6] Traditionally, the normal application of these transformers has been speech systems and this power rating standard assumes very little energy will exist under 100Hz. The same reference recommends that transformers used in systems with emphasized bass should have ratings higher than this 100Hz nominal power rating and those used to handle organ music should have ratings of at least four times nominal. Since the power ratings for these transformers is rarely qualified by an honest specification stating the applicable frequency, it seems prudent to assume that ***the historical 100Hz power rating applies to most commercial transformers***.

If a background music system, for example, requires good bass response, it is wise to use over-rated transformers. Reducing the voltage on the primary side of the transformer will extends its low-frequency power handling. It's possible, using the table above, to use different taps to achieve the same ratio while driving less than nominal voltage into the transformer primary. For example, a 70V line could be connected to the 100V input of the transformer in Fig. 15-33 and, for example, the 10W secondary tap used to actually deliver 5W. In any constant-voltage system, saturation problems can be reduced by appropriate high-pass filtering. Simply attenuate low-frequency signals before they can reach the transformers. In voice-only systems, problems that arise from breath pops, dropped microphones, or signal switching transients can be effectively eliminated by a 100Hz high-pass filter ahead of the power amplifier. In music systems, attenuating frequencies too low for the speakers to reproduce can be similarly helpful.

15.2.2.6 Telephone Isolation or Repeat Coil

In telephone systems it was sometimes necessary to isolate a circuit which was grounded at both ends. This metallic circuit problem was corrected with a repeat coil to improve longitudinal balance. Translating from telephone lingo, this balanced line had poor common-mode noise rejection which was corrected with a 1:1 audio isolation transformer. The Western Electric 111C repeat coil was widely used by radio networks and others for high-quality audio transmission over 600Ω phone lines. It has split primary and secondary windings and a Faraday shield. Its frequency response was 30Hz to 15kHz and it had less than 0.5dB insertion loss. Split windings allow them to be parallel connected for 150Ω use.

Fig. 15-46 shows a modern version of this transformer as a general purpose isolator for low-impedance circuits, such as in a recording studio patch-bay. Optional components can be useful in some applications. For example, network R_1 and C_1 will flatten the input impedance over frequency, R_2 will trim the input impedance to exactly 600Ω, and R_3 can be used to properly load the transformer when the external load is high-impedance or bridging.

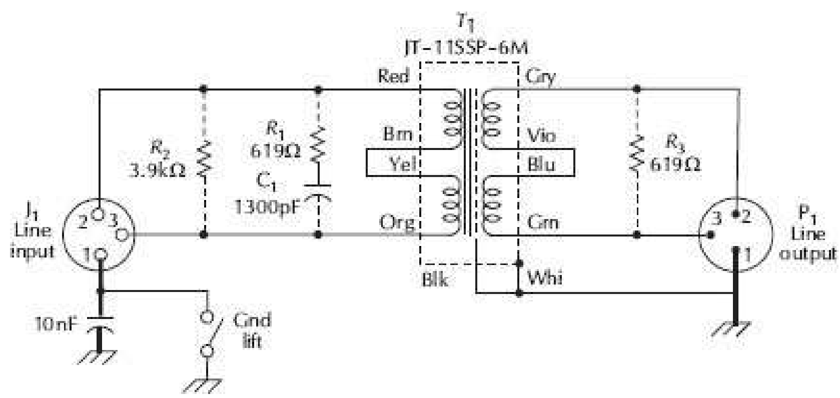


Figure 15-46. Repeat coil ground isolation for 600Ω lines.

15.2.2.7 Telephone Directional Coupling or Hybrid

Telephone hybrid circuits use bridge-nulling principles to separate signals which may be transmitted and received simultaneously or full-duplex on a 2-wire line. This nulling depends critically on well-controlled impedances in all branches of the circuits. This nulling is what suppresses the transmit signal (your own voice) in the receiver of your phone while allowing you to hear the receive signal (the other party).

A two-transformer hybrid network is shown in Fig. 15-47. The arrows and dashed lines show the current flow for a signal from the transmitter TX. Remember that the dots on the transformers show points having the same instantaneous polarity. The transformer turns ratios are assumed to be 1:1:1. When balancing network Z_N has an impedance that matches the line impedance Z_L at all significant frequencies, the currents in the Z_L loop (upper) and Z_N loop (lower) will be equal. Since they flow in opposite directions in the RX transformer (right), there is cancellation and the TX signal does not appear at RX. A signal originating from the line rather than TX is not suppressed and is heard in RX. A common problem with hybrids of any kind is adjusting network Z_N to match the telephone line, which may vary considerably in impedance even over relatively short time spans.

If the transmitter and receiver are electrically connected, the single transformer method, shown in Fig. 15-48, can be used. Any well-designed transformers with accurate turns ratio can be used in hybrid applications.

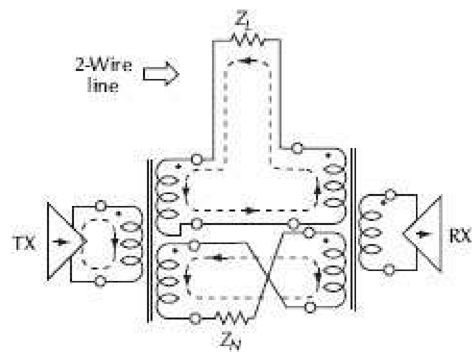


Figure 15-47. Two transformer hybrid.

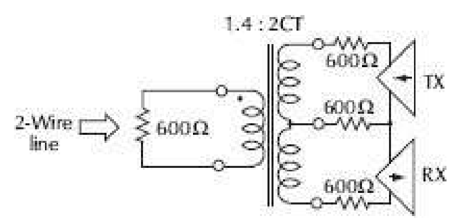


Figure 15-48. Single-transformer hybrid.

15.2.2.8 Moving-Coil Phono Step-Up

Outboard boxes are sometimes used to adapt the output of low-output, low-impedance moving-coil phono pickups to pre-amplifier inputs intended for conventional high-impedance moving-magnet pickups. These pre-amplifiers have a standard input impedance of 47kΩ. Fig. 15-49 shows a 1:37 step-up transformer used for this purpose. It has a voltage gain of 31dB and reflects its 47kΩ pre-amplifier load to the pickup as about 35Ω. This keeps loading loss on a 3Ω pickup to about 1 dB. The series RC network on the secondary provides proper damping for smooth frequency response. Double magnetic shield cans are used because of the very low signal levels involved and the low-frequency gain inherent in the RIAA playback equalization. In these applications, it is extremely important to keep all leads to the pickup tightly twisted to avoid hum from ambient magnetic fields.

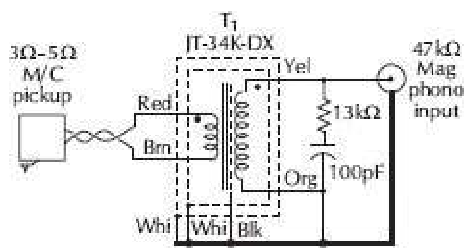


Figure 15-49. Step-up transformer for moving coil phono pickup.

15.3 Measurements and Data Sheets

15.3.1 Testing and Measurements

15.3.1.1 Transmission Characteristics

The test circuits below are the basic setups to determine the signal transmission characteristics of output and input type transformers, respectively, shown in the diagrams as DUT for device under test. In each case, the driving source impedance must be specified and is split into two equal parts for transformers specified for use in balanced systems. For example, if a 600Ω balanced source is specified, the resistors $R_S/2$ become 300Ω each. The generator indicated in both diagrams is understood to have symmetrical voltage outputs. The buffer amplifiers shown are used to provide a zero source impedance, which is not available from most commercial signal sources. The generator could be used in an unbalanced mode by simply connecting the lower end of the DUT primary to ground. The specified load impedance must also be placed on the secondary. For output transformers, the load and meter are often floating as shown in Fig. 15-50. For input transformers, a specified end of the secondary is generally grounded as shown in Fig. 15-51.

These test circuits can be used to determine voltage gain or loss, turns ratio when R_L is infinite, frequency response, and phase response. If the meter is replaced with a distortion analyzer, distortion and maximum operating level may be characterized. Multi-purpose equipment such as the Audio Precision System 1 or System 2 can make such tests fast and convenient. Testing of high-power transformers usually requires an external power amplifier to boost the generator output as well as some hefty power resistors to serve as loads.

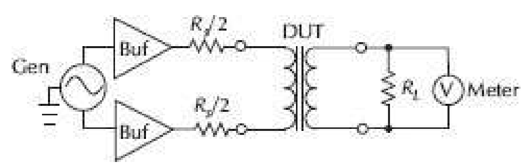


Figure 15-50. Transmission tests for output types.

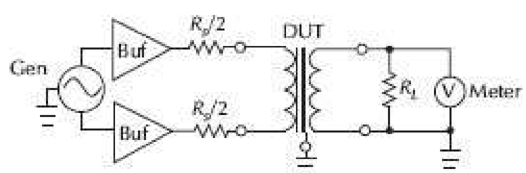


Figure 15-51. Transmission tests for input types.

15.3.1.2 Balance Characteristics

Tests for common-mode rejection are intended to apply a common-mode voltage through some specified resistances to the transformer under test. Any differential voltage developed then represents undesired conversion of common-mode voltage to differential mode by the transformer. In general terms, CMRR or common-mode rejection ratio, is the ratio of the response of a circuit to a voltage applied normally (differentially) to that of the same voltage applied in common-mode through specified impedances. This conversion is generally the result of mismatched internal capacitances in the balanced winding. For output transformers, the most common test arrangement is shown in Fig. 15-52. Common values are 300Ω for RG and values from zero to 300Ω for $R_S/2$. Resistor pairs must be very well matched.

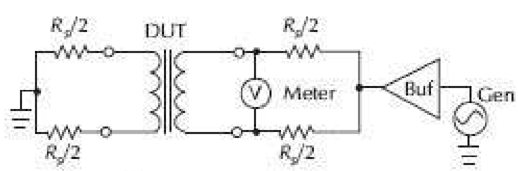


Figure 15-52. Common-mode test for output types.

Traditionally, CMRR tests of balanced input stages involved applying the common-mode voltage through a pair of very tightly-matched resistors. As a result, such traditional tests were not accurate predictors of real-world noise rejection for the overwhelming majority of electronically-balanced inputs. The IEC recognized this in 1998 and solicited suggestions to revise the test. The problem arises from the fact that the common-mode output impedances of balanced sources in typical commercial equipment are not matched with laboratory precision. Imbalances of 10Ω are quite common. This author, through an educational process about balanced interfaces in general, suggested a more realistic test which was eventually adopted by the IEC in their standards document 60268-3 “Testing of Amplifiers” in August, 2000. The “Informative Annex” of this document is a concise summary explaining the nature of a balanced interface. The method of the new test, as shown in Fig. 15-53, is simply to introduce a 10Ω imbalance, first in one line and then in the other. The CMRR is then computed based on the highest differential reading observed.

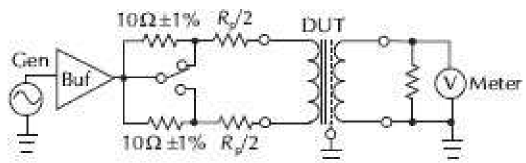


Figure 15-53. IEC Common-mode test for input types.

15.3.1.3 Resistances, Capacitances, and Other Data

Other data which can be very helpful to an equipment or system designer includes resistances of each winding and capacitances from winding to winding or winding to Faraday shield or transformer frame. Do not use an ohmmeter to check winding resistances unless you are able to later demagnetize the part. Ordinary ohmmeters, especially on low-ohm ranges, can weakly magnetize the core. If an ohmmeter simply must be used, use the highest range where the current is least.

Capacitances are usually measured on impedance bridges and, to minimize the effects of winding inductances, with all windings shorted. Total capacitances can be measured this way, but balance of capacitances across a winding must be measured indirectly. CMRR tests are effectively measuring capacitance imbalances.

As shown in Fig. 15-54, sometimes the input impedance of a winding is measured with specified load on other windings. This test includes the effects of primary resistance, secondary resistance, and the parallel loss resistance RC shown in Fig. 15-8 and Fig. 15-13. If specified over a wide frequency range, it also includes the effects of primary inductance and winding capacitances.

Breakdown voltages are sometimes listed as measures of insulation integrity. This is normally done with special equipment, sometimes called a hi-pot tester, which applies a non-destructive high voltage while limiting current to a very low value.

15.3.2 Data Sheets

15.3.2.1 Data to Impress or to Inform?

Data sheets and specifications exist to allow easy comparison of one product with others. But, in a world where marketing seems to supersede all else, honest data sheets and guaranteed specifications are becoming increasingly rare. As with many other audio products, most so-called data sheets and specifications are designed to impress rather than inform. **Specifications offered with unstated measurement conditions are essentially meaningless**, so a degree of skepticism is always appropriate before comparisons are made. A few examples:

- Hum Eliminator and Line Level Shifter products with no noise rejection or CMRR specs at all!
- Line Level Shifter products with no gain spec at all! Section 15.2.2.4 explains why they won't tell you!
- Maximum Power or Maximum Level listed with no frequency and no source impedance specified!

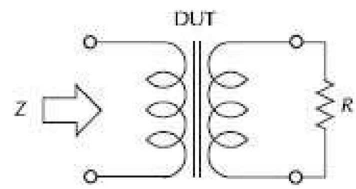


Figure 15-54. Impedance tests.

Other specifications, while technically true, are likely to mislead those not wise in the ways of transformers. For example, Maximum Level and Distortion are commonly specified at 50Hz, 40Hz, or 30Hz instead of the more rigorous 20Hz. Be careful, specs at these higher frequencies will always be much more impressive than those at 20Hz! There is an approximate 6dB per octave relationship at work here. A transformer specified for level or distortion at 40Hz, for example, will handle about 6dB less level at 20Hz and have *at least* twice the distortion!

Seen in transformer-based hum eliminator advertising copy: “Frequency response 10Hz to 40kHz ±1dB into 10kΩ load” and “Distortion less than 0.002% at 1kHz.” What about the source impedance? Response at 10Hz and distortion are always much better when a transformer is driven from a 0Ω source! What happens when a real-world source drives the box? For a full-range audio transformer, measuring distortion at 1kHz is nearly meaningless. Section 15.1.3.1 explains.

15.3.2.2 Comprehensive Data Sheet Example

For reference, Fig. 15-55 is offered as a sample of a data sheet that has been called truly useful and brutally honest. Note that minimum or maximum limits are guaranteed for the most critical specifications!

15.4 Installation and Maintenance

15.4.1 A Few Installation Tips

- Remember that there are very tiny wires inside an audio transformer. Its wire leads should never be used like a handle to pick it up. The internal bonds are strong, but pulling too hard might result in an open winding.
- Be careful with sharp tools. A gouge through the outer wrapper of an output transformer can nick or cut an internal winding.
- When mounting transformers that are in shielded cans, use either the supplied screws or ones no longer than recommended. If the screws are too long, they'll bore right into the windings—big problem!
- Be careful about using magnetized tools. If a screwdriver will pick up a paper clip, it shouldn't be used to install an audio transformer.
- Don't drop a transformer. It can distort the fit of the laminations in output transformers and affect their low-frequency response. Mechanical stress, as in denting of the magnetic shield can of an input transformer will reduce its effectiveness as a shield. For the same reason, don't over-tighten the clamps on transformers mounted with them.
- Twisting helps avoid hum pickup from ambient ac magnetic fields. This is especially true for microphone level lines in splitters, for example. Separately twist the leads from <u>each</u> winding—twisting the leads from *all* windings together can reduce noise rejection or CMRR.

15.4.2 De-Magnetization

Some subtle problems are created when transformer cores and/or their shield cans become magnetized. Generally, cores become magnetized by having dc flow in a winding, even for a fraction of a second. It can leave the core weakly magnetized. Steel cores, because of their wider hysteresis loops, are generally the most prone to such magnetization. The only way to know if the core has some permanent magnetization is to perform distortion measurements. A transformer with an unmagnetized core will exhibit nearly pure third harmonic distortion, with virtually no even order harmonic distortion while magnetized ones will show significant even order distortion, possibly with 2nd harmonic even exceeding 3rd. A test signal at a level about 30 or 40dB below rated maximum operating level at 20 or 30Hz is typically the most revealing because it maximizes the contribution of hysteresis distortion.

Microphone input transformers used with phantom power are exposed to this possibility whenever a microphone is connected or disconnected from a powered input. However, distortion tests before and after exposure to the worst-case 7mA current pulses have shown that the effects are indeed subtle. Third harmonic distortion, which normally dominates transformer distortions, is unaffected. Second harmonic, which normally is near the measurement threshold, is typically increased by about 20dB but is still some 15dB lower than the third harmonic. Is it audible? Some say yes. But even this distortion disappears into the noise floor above a few hundred Hz. In any case, it can be prevented by connecting and disconnecting microphones *only* when phantom power is off. And such magnetized transformers can be de-magnetized.

Demagnetizing of low level transformers can generally be done with any audio generator having a continuously variable output. It may take a booster of some sort to get enough level for output transformers (be sure there's no dc offset at its output!). The idea is to drive the transformer deeply into saturation, 5% THD or more, and then slowly bring the level down to zero. Saturation will, of course, be easiest at a very low frequency. How much level it takes will depend on the transformer. If you're lucky, the level required may not be hazardous to the surrounding electronics and the demagnetizing can be accomplished without disconnecting the transformer. Start with the generator set to 20Hz and its minimum output level, connect it to the transformer, then slowly—over a period of a few seconds—increase the level into saturation—maintain it for a few seconds—then slowly turn it back down to minimum. For the vast majority of transformers, this process will leave them in a demagnetized state.

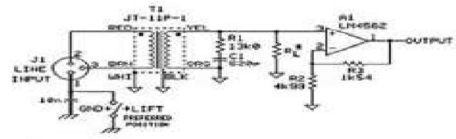
jensen
Transformers, Inc.

JT-11P-1

LINE INPUT TRANSFORMER
1:1 FOR "BALANCED BRIDGING" INPUTS

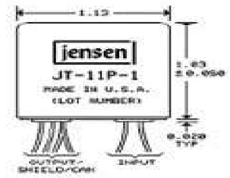
- ***Ideal for balancing any high-impedance unbalanced input***
- ***Wide bandwidth: −3 dB at 0.25 Hz and 100 kHz***
- ***Recommended for levels up to +20 dBu at 20 Hz***
- ***High input impedance: 13 kΩ with 10 kΩ load***
- ***High common-mode rejection: 107 dB at 60 Hz***

This transformer is designed for use in wideband line input stages. Distortion remains very low and CMRR remains high, even when driven by high source impedances. The primary is fully balanced and its leads may be reversed to invert polarity, if required. A 30 dB magnetic shield package is standard.

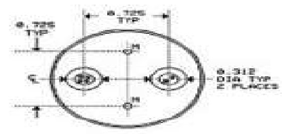


* R_L MUST BE 10 kΩ OR GREATER
OMIT R1 & C1 (DAMPING NETWORK) FOR R_L = 10 kΩ ONLY

LOW NOISE UNITY GAIN INPUT STAGE



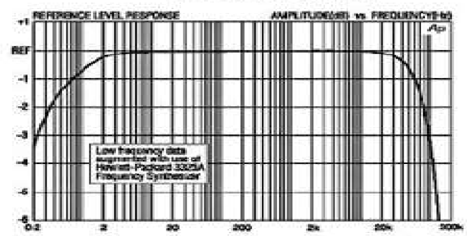
#30 AWG (7×38) UL STYLE 1061 COLOR CODED WIRE LEADS, 8" MINIMUM LENGTH

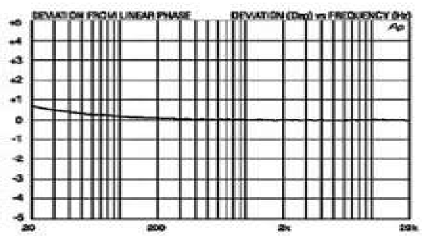


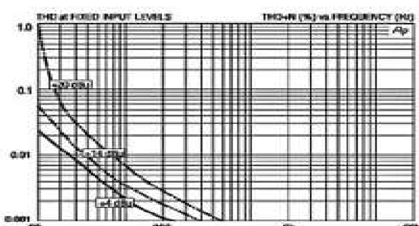
BOTTOM VIEW

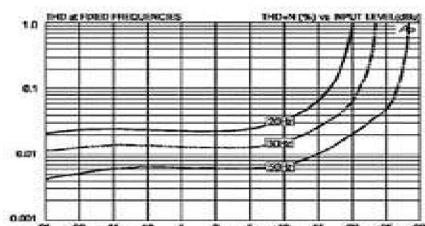
USE ONLY #4 TYPE B SELF TAPPING SCREWS IN HOLES "M". ALLOW NO MORE THAN 0.15" PENETRATION INTO TRANSFORMER HOUSING.

TYPICAL APPLICATION



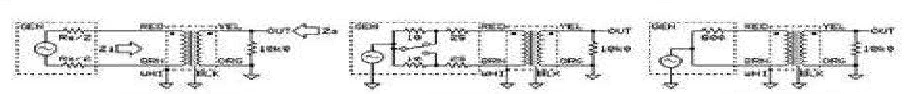






JT-11P-1 SPECIFICATIONS (all levels are input unless noted)

PARAMETER	*CONDITIONS*	*MINIMUM*	*TYPICAL*	*MAXIMUM*
Input impedance, Zi	1 kHz, +4 dBu, test circuit 1	12.3 kΩ	13.0 kΩ	13.7 kΩ
Voltage gain	1 kHz, +4 dBu, test circuit 1	-2.6 dB	-2.3 dB	-2.0 dB
Magnitude response, ref 1 kHz	20 Hz, +4 dBu, test circuit 1, Rs=600 Ω	−0.15 dB	−0.04 dB	0.0 dB
	20 kHz, +4 dBu, test circuit 1, Rs=600 Ω	−0.15 dB	−0.05 dB	0.0 dB
Deviation from linear phase (DLP)	20 Hz to 20 kHz, +4 dBu, test circuit 1, Rs=600 Ω		+0.6°	±2.0°
Distortion (THD)	1 kHz, +4 dBu, test circuit 1, Rs=600 Ω		<0.001%	
	20 Hz, +4 dBu, test circuit 1, Rs=600 Ω		0.025%	0.10%
Maximum 20 Hz input level	1% THD, test circuit 1, Rs=600 Ω	+18 dBu	+20 dBu	
Common-mode rejection ratio (CMRR) 50 Ω balanced source	per IEC 60268-3, 60 Hz, test circuit 2		107 dB	
	per IEC 60268-3, 3 kHz, test circuit 2	65 dB	73 dB	
Common-mode rejection ratio (CMRR) 600 Ω unbalanced source	60 Hz, test circuit 3		100 dB	
	3 kHz, test circuit 3		68 dB	
Output impedance, Zo	1 kHz, test circuit 1, Rs=50 Ω		2.34 kΩ	
DC resistances	primary (RED to BRN)		1.45 kΩ	
	secondary (YEL to ORG)		1.55 kΩ	
Capacitances @ 1 kHz	primary to shield and case		98 pF	
	secondary to shield and case		110 pF	
Turns ratio		0.999:1	1.000:1	1.001:1
Temperature range	operation or storage	0° C		70° C
Breakdown voltage (see IMPORTANT NOTE below)	primary or secondary to shield and case, 60 Hz, 1 minute test duration	250 V RMS		



TEST CIRCUIT 1 TEST CIRCUIT 2 TEST CIRCUIT 3

All minimum and maximum specifications are guaranteed. Unless noted otherwise, all specifications apply at 25°C. Specifications subject to change without notice. All information herein is believed to be accurate and reliable, however no responsibility is assumed for its use nor for any infringements of patents which may result from its use. No license is granted by implication or otherwise under any patent or patent rights of Jensen Transformers, Inc.
IMPORTANT NOTE: This device is NOT intended for use in life support systems or any application where its failure could cause injury or death. The breakdown voltage specification is intended to insure integrity of internal insulation systems; continuous operation at these voltages is NOT recommended. Consult our applications engineering department if you have special requirements.

JENSEN TRANSFORMERS, INC., 9304 Deering Avenue, Chatsworth, CA 91311, USA
(818) 374-5857 • FAX (818) 374-5856 • www.jensen-transformers.com

Figure 15-55. Specification sheet for a quality transformer. Courtesy Jensen Transformers, Inc.

Shield cans are usually magnetized by having a brief encounter with a strongly magnetized tool. Sometimes, transformers are unknowingly mounted on a magnetized chassis. When the shield can of an input transformer becomes magnetized, the result is *microphonic* behavior of the transformer. Even though quality input transformers are potted with a semi-rigid epoxy compound to prevent breakage of very fine wires, vibration between core and can activate what is essentially a variable reluctance microphone. In this case, a good strong tape head de-magnetizer can be used to de-magnetize the can. At the end of the Jensen production line, most transformers are routinely demagnetized with a very strong de-magnetizer just prior to shipment. Although I haven't tried it, I would expect that something like a degausser for 2 inch video tape (remember that!) would also de-magnetize even a large steel-core output transformer.

References

1. Magnetic Shield Corporation, *Frequently Asked Questions*, www.magnetic-shield.com.

2. G.A.V. Sowter, *Soft Magnetic Materials for Audio Transformers: History, Production, and Applications*, Journal of the Audio Engineering Society, October 1987, www.sowter.co.uk/pdf/GAVS.pdf.

3. Bill Whitlock, *Balanced Lines in Audio: Fact, Fiction, and Transformers*, Journal of the Audio Engineering Society, June 1995, pp 454–464.

4. F. Langford Smith, *Radiotron Designer's Handbook*, Wireless Press, Sydney, 4th Edition, 1953, p 208.

5. Lawrence Woolf, *RMS Watt, or Not?*, Electronics World, December 1998, pp 1043–1045.

6. F. Langford Smith, *op. cit.*, p 227.

7. Bill Whitlock, Theory and Construction of Mic "Splitters", AN005, Jensen Transformers, Inc., 1996, www.jensen-transformers.com/an/an005.pdf.

Notes

Co-Netic® is a registered trademark of Magnetic Shield Corp.
HyMu® is a registered trademark of Carpenter Technology Corp.
Mumetal® is a registered trademark of Telcon Metals, Ltd.
Permalloy® is a registered trademark of B & D Industrial & Mining Services, Inc.

Chapter 16

Tubes, Discrete Solid State Devices, and Integrated Circuits

by Glen Ballou, Leslie B. Tyler, and Wayne Kirkwood

16.1 Tubes

In 1883, Edison discovered that electrons flowed in an evacuated lamp bulb from a heated filament to a separate electrode (the Edison effect). Fleming, making use of this principle, invented the

Fleming valve in 1905, but when DeForest, in 1907, inserted the grid, he opened the door to electronic amplification with the *audion*. The millions of vacuum tubes are an outgrowth of the principles set forth by these men.[1]

It was thought that, with the invention of the transistor and integrated circuits, the tube would disappear from audio circuits. This has hardly been the case. Recently tubes have had a revival because some "golden ears" like the smoothness and nature of the tube sound. The 1946 vintage 12AX7 is not dead and is still used today as are miniature tubes in condenser microphones and 6L6s in power amplifiers. It is interesting that many feel that a 50W tube amplifier sounds better than a 250W solid-state amplifier. For this reason, like the phonograph, tubes are still discussed in this handbook.

16.1.1 Tube Elements

Vacuum tubes consist of various elements or electrodes, Table 16-1. The symbols for these elements are shown in Fig. 16-1.

Table 16-1. Vacuum Tube Elements and Their Designation

Filament	The cathode in a directly heated tube that heats and emits electrons. A filament can also be a separate coiled element used to heat the cathode in an indirectly heated tube.
Cathode	The sleeve surrounding the heater that emits electrons. The surface of the cathode is coated with barium oxide or thoriated tungsten to increase the emission of electrons.
Plate	The positive element in a tube and the element from which the output signal is usually taken. It is also called an *anode*.

Control grid	The spiral wire element placed between the plate and cathode to which the input signal is generally applied. This element controls the flow of electrons or current between the cathode and the plate.
Screen grid	The element in a tetrode (four element) or pentode (five element) vacuum tube that is situated between the control grid and the plate. The screen grid is maintained at a positive potential to reduce the capacitance existing between the plate and the control grid. It acts as an electrostatic shield and prevents self-oscillation and feedback within the tube.
Suppressor grid	The gridlike element situated between the plate and screen in a tube to prevent secondary electrons emitted by the plate from striking the screen grid. The suppressor is generally connected to the ground or to the cathode circuit.

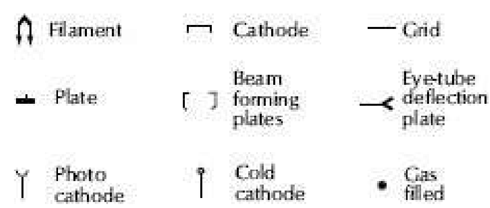


Figure 16-1. Tube elements and their designation.

16.1.2 Tube Types

There are many types of tubes, each used for a particular purpose. All tubes require a type of heater to permit the electrons to flow. Table 16-2 defines the various types of tubes.

Table 16-2. The Eight Types of Vacuum Tubes

Diode	A two-element tube consisting of a plate and a cathode. Diodes are used for rectifying or controlling the polarity of a signal as current can flow in one direction only.
Triode	A three-element tube consisting of a cathode, a control

	grid, and a plate. This is the simplest type of tube used to amplify a signal.
Tetrode	A four-element tube containing a cathode, a control grid, a screen grid, and a plate. It is frequently referred to as a screen-grid tube.
Pentode	A five-element tube containing a cathode, a control grid, a screen grid, a suppressor grid, and a plate.
Hexode	A six-element tube consisting of a cathode, a control grid, a suppressor grid, a screen grid, an injector grid, and a plate.
Heptode	A seven-element tube consisting of a cathode, a control grid, four other grids, and a plate.
Pentagrid	A seven-element tube consisting of a cathode, five grids, and a plate.
Beam-power tube	A power-output tube having the advantage of both the tetrode and pentode tubes. Beam-power tubes are capable of handling relatively high levels of output power for application in the output stage of an audio amplifier. The power-handling capabilities stem from the concentration of the plate-current electrons into beams of moving electrons. In the conventional tube the electrons flow from the cathode to the plate, but they are not confined to a beam. In a beam-power tube the internal elements consist of a cathode, a control grid, a screen grid, and two beam-forming elements that are tied internally to the cathode element. The cathode is indirectly heated as in the conventional tube.

16.1.3 Symbols and Base Diagrams

Table 16-3 gives the basic symbols used for tube circuits. The basing diagrams for various types of vacuum tubes are shown in Fig. 16-2.

Table 16-3. Tube Nomenclature

C	Coupling capacitor between stages

C_{g2}	Screen grid bypass capacitor
C_k	Cathode bypass capacitor
E_{bb}	Supply voltage
E_{ff}	Plate efficiency
E_p	Actual voltage at plate
E_{sg}	Actual voltage at screen grid
E_o	Output voltage
E_{sig}	Signal voltage at input
E_g	Voltage at control grid
E_f	Filament or heater voltage
I_f	Filament or heater current
I_p	Plate current
I_k	Cathode current
I_p	Screen-grid current
I_{pa}	Average plate current
I_{pac}	Average ac plate current
I_{ha}	Average cathode current
I_{sga}	Average screen grid current
g_m	Transconductance (mutual conductance)
mu	Amplification factor (μ)
P_{sg}	Power at screen grid
P_p	Power at plate
P-P	Plate-to-plate or push–-pull amplifier
R_g	Grid resistor
R_k	Cathode resistor
R_l	Plate-load impedance or resistance
R_p	Plate-load resistor
R_{sg}	Screen-dropping resistor
R_d	Decoupling resistor
r_p	Internal plate resistance
V_g	Voltage gain

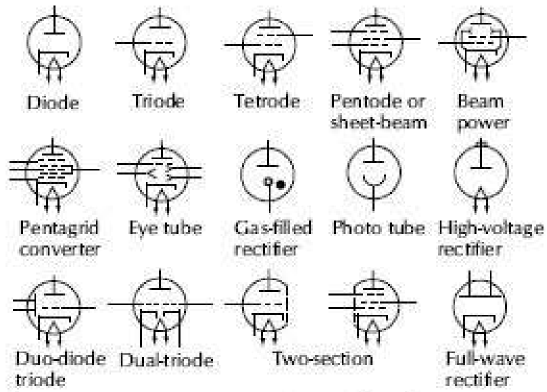


Figure 16-2. Basing diagrams for popular tubes.

16.1.4 Transconductance

Transconductance (g_m) is the change in the value of plate current expressed in microamperes (µA) divided by the signal voltage at the control grid of a tube, and is expressed by conductance. *Conductance* is the opposite of resistance, and the name *mho* (*ohm* spelled backward) was adopted for this unit of measurement. *Siemens* (S) have been adopted as the SI standard for conductance and are currently replacing mhos in measurement.

The basic mho or Siemens is too large for practical usage; therefore, the terms *micromho* (µmho) and *microsiemens* (µS) are used. One micromho is equal to one-millionth of a mho.

The transconductance (g_m) of a tube in µmhos may be found with the equation

$$g_m = \frac{\Delta I_p}{\Delta E_{sig}} \tag{16-1}$$

where,

ΔI_p is the change of plate current,

ΔE_{sig} is the change of control-grid signal voltage,

E_{bb} is the plate supply voltage and is held constant.

For example, a change of 1mA of plate current for a change of 1V at the control grid is equal to a transconductance of 1000µmho. A tube having a change of 2mA plate current for a change of 1V at the control grid would have a transconductance of 2000µmho.

$$g_m = I_{pac} \times 1000 \qquad (16\text{-}2)$$

where,

g_m is the transconductance in µmho or µS,

I_{pac} is the ac plate current.

16.1.5 Amplification Factor

Amplification factor (µ) or *voltage gain* (V_g) is the ratio of the incremental plate voltage change to the control-electrode voltage change at a fixed plate current and constant voltage on all other electrodes. This normally is the amount the signal at the control grid is increased in amplitude after passing through the tube.

Tube voltage gain may be computed using the equation

$$V_g = \frac{\Delta E_p}{\Delta E_g} \qquad (16\text{-}3)$$

where,

V_g is the voltage gain,

ΔE_p is the change in signal plate voltage,

ΔE_g is the change in the signal grid voltage.

If the amplifier consists of several stages, the amount of amplification is multiplied by each stage. The gain of an amplifier stage varies with the type of tube and the interstage coupling used. The general equation for voltage gain is

$$V_{gt} = V_{g1} V_{g2} \ldots V_{gn} \qquad (16\text{-}4)$$

where,

V_{gt} is the total gain of the amplifier,

V_{g1}, V_{g2}, and V_{gn} are the voltage gain of the individual stages.

Triode tubes are classified by their amplification factor. A low-μ tube has an amplification factor less than 10. Medium-μ tubes have an amplification factor from 10 to 50, with a plate resistance of 5Ω to 15,000Ω. High-μ tubes have an amplification factor of 50 to 100 with a plate resistance of 50kΩ to 100kΩ.

16.1.6 Polarity

Polarity reversals take place in a tube. The polarity reversal in electrical degrees between the elements of a self-biased pentode for a given signal at the control grid is shown in Fig. 16-3A. The reversals are the same for a triode. Note that, for an instantaneous positive voltage at the control grid, the voltage polarity between the grid and plate is 180° and will remain so for all normal operating conditions. The control grid and cathode are in polarity. The plate and screen-grid elements are in polarity with each other. The cathode is 180° out of polarity with the plate and screen-grid elements.

The polarity reversal of the instantaneous voltage and current for each element is shown in Fig. 16-3B. For an instantaneous positive sine wave at the control grid, the voltages at the plate and screen grid are negative, and the currents are positive. The voltage and current are both positive in the cathode resistor and are in polarity with the voltage at the control grid. The reversals are the same in a triode for a given element.

16.1.7 Internal Capacitance

The *internal capacitance* of a vacuum tube is created by the close proximity of the internal elements, Fig. 16-4. Unless otherwise stated by the manufacturer, the internal capacitance of a glass tube is measured using a close-fitting metal tube shield around the glass envelope connected to the cathode terminal. Generally, the capacitance is measured with the heater or filament cold and with no voltage applied to any of the other elements.

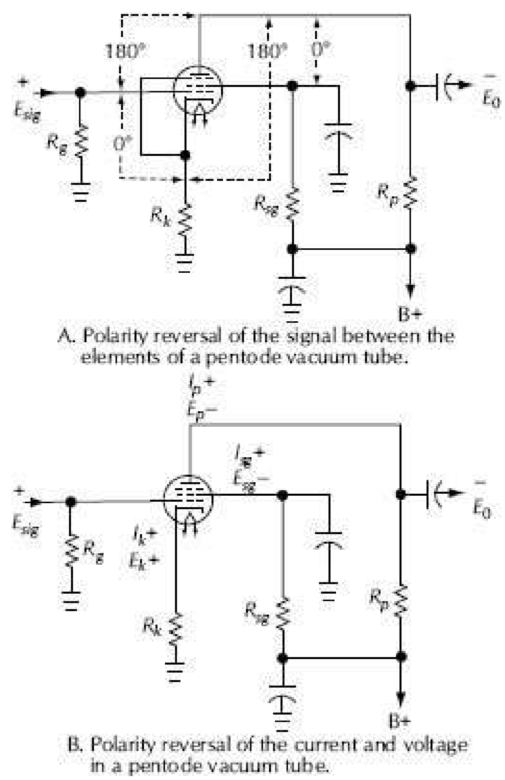


Figure 16-3. Polarity characteristics of a vacuum tube.

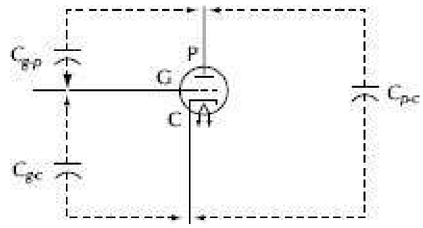


Figure 16-4. Interelectrode capacitance of a triode.

In measuring the capacitance, all metal parts, except the input and output elements, are connected to the cathode. These metal parts include internal and external shields, base sleeves, and unused pins. In testing a midsection tube, elements not common to the section being measured are connected to ground.

Input capacitance is measured from the control grid to all other elements, except the plate, which is connected to ground.

Output capacitance is measured from the plate to all other elements, except the control grid, which is connected to ground.

Grid-to-plate capacitance is measured from the control grid to the plate with all other elements connected to ground.

16.1.8 Plate Resistance

The *plate resistance* (r_p) of a vacuum tube is a constant and denotes the internal resistance of the tube or the opposition offered to the passage of electrons from the cathode to the plate. Plate resistance may be expressed in two ways: the dc resistance and the ac resistance. Dc resistance is the internal opposition to the current flow when steady values of voltage are applied to the tube elements and may be determined simply by using Ohm's Law

$$r_{p_{dc}} = \frac{E_p}{I_p} \qquad (16\text{-}5)$$

where,

E_p is the dc plate voltage,

I_p is the steady value of plate current.

The ac resistance requires a family of plate-current curves from which the information may be extracted. As a rule, this information is included with the tube characteristics and is used when calculating or selecting components for an amplifier. The equation for calculating ac plate resistance is

$$r_{p_{ac}} = \frac{\Delta E_p}{\Delta I_p} \qquad (16\text{-}6)$$

where,

ΔE_p is the change in voltage at the plate,

ΔI_p is the change in plate current,

E_{sig} is the control grid signal voltage and is held constant.

The values of E_p and I_p are taken from the family of curves supplied by the manufacturer for the particular tube under consideration.

16.1.9 Grid Bias

Increasing the plate voltage or decreasing the grid-bias voltage decreases the plate resistance. The six methods most commonly used to bias a tube are illustrated in Fig. 16-5. In Fig. 16-5A bias cell (battery) is connected in series with the control grid. In Fig. 16-5B the tube is self-biased by the use of a resistor connected in the cathode circuit. In Fig. 16-5C the circuit is also a form of self-bias; however, the bias voltage is obtained by the use of a grid capacitor and grid-leak resistor connected between the control grid and ground. In Fig. 16-5D the bias voltage is developed by a grid-leak resistor and capacitor in parallel, connected in series with the control grid. The method illustrated in Fig. 16-5E is called *combination bias* and consists of self-bias and battery bias. The resultant bias voltage is the negative voltage of the battery, and the bias created by the self-bias resistor in the cathode circuit. Another combination bias circuit is shown in Fig. 16-5F. The bias battery is connected in series with the grid-leak resistor. The bias voltage at the control grid is that developed by the battery and the self-bias created by the combination of the grid resistor and capacitor.

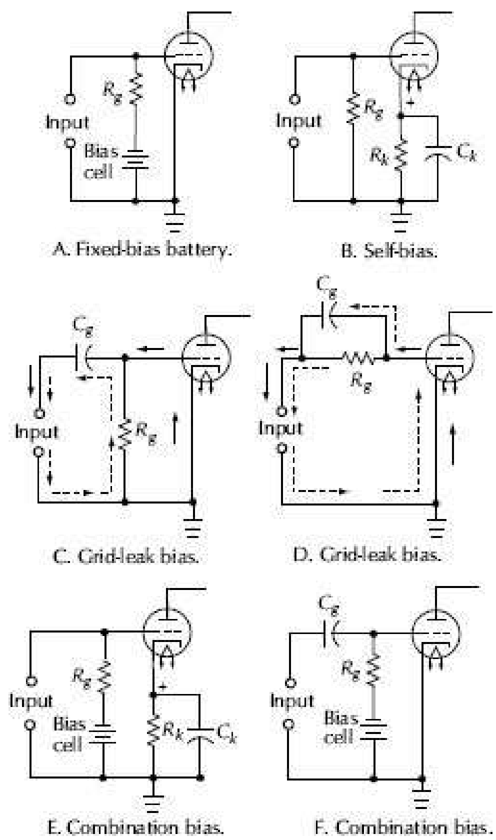


Figure 16-5. Various methods of obtaining grid bias.

If the control grid becomes positive with respect to the cathode, it results in a flow of current between the control grid and the cathode through the external circuits. This condition is unavoidable because the wires of the control grid, having a positive charge, attract electrons passing from the cathode to the plate. It is important that the control-grid voltage is kept negative, reducing grid current and distortion.

Grid-current flow in a vacuum tube is generally thought of as being caused by driving the control grid into the positive region and causing the flow of grid current.

The grid voltage, plate-current characteristics are found through a series of curves supplied by the tube manufacturer, as shown in Fig. 16-6.

The curves indicate that for a given plate voltage the plate current and grid bias may be determined. For example, the manufacturer states that for a plate voltage of 250V and a negative grid bias of −8V, the plate current will be 9mA, which is indicated at point A on the 250V curve. If it is desired to operate this tube with a plate voltage of 150V and still maintain a plate current of 9mA, the grid bias will have to be changed to a −3V.

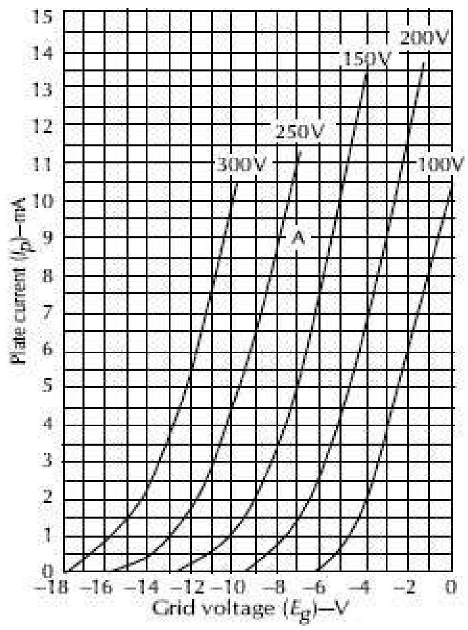


Figure 16-6. Grid voltage, plate-current curves for a triode tube.

16.1.10 Plate Efficiency

The *plate efficiency* (E_{ff}) is calculated by the equation:

$$E_{ff} = \frac{watts}{E_{pa}I_{pa}} \times 100 \qquad (16\text{-}7)$$

where,

watts is the power output,

E_{pa} is the average plate voltage,

I_{pa} is the average plate current.

The measurement is made with a load resistance in the plate circuit equal in value to the plate resistance stated by the manufacturer.

16.1.11 Power Sensitivity

Power sensitivity is the ratio of the power output to the square of the input voltage, expressed in mhos or Siemens and is determined by the equation

$$\text{Power sensitivity} = \frac{P_o}{E_{in}^{\;2}} \qquad (16\text{-}8)$$

where,

P_o is the power output of the tube in W,

E_{sig} is the rms signal voltage at the input.

16.1.12 Screen Grid

The *screen grid* series-dropping resistance is calculated by referring to the data sheet of the manufacturer and finding the maximum voltage that may be applied and the maximum power that may be dissipated by the screen grid. These limitations are generally shown graphically as in Fig. 16-7. The value of the resistor may be calculated using the equation

$$R_{sg} = \frac{E_{sg} \times (E_{bb} - E_{sg})}{P_{sg}} \qquad (16\text{-}9)$$

where,

R_{sg} is the minimum value for the screen-grid voltage-dropping resistor in Ω,

E_{sg} is the selected value of screen-grid voltage,

E_{bb} is the screen-grid supply voltage,

P_{sg} is the screen-grid input in watts corresponding to the selected value of E_{sg}.

16.1.13 Plate Dissipation

Plate dissipation is the maximum power that can be dissipated by the plate element before damage and is found with the equation

$$\text{Watts dissipation} = E_p I_p \qquad (16\text{-}10)$$

where,

E_p is the voltage at the plate,

I_p is the plate current.

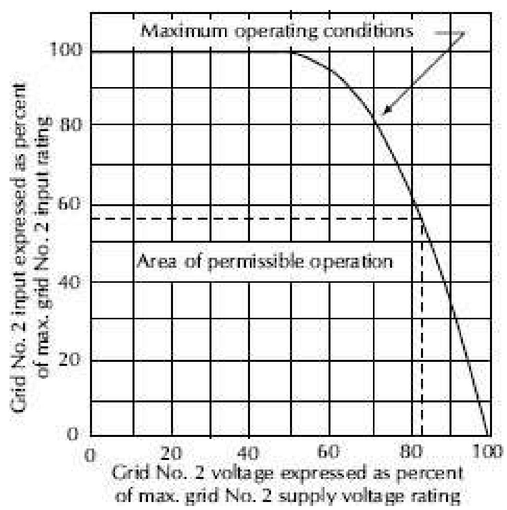


Figure 16-7. Typical graph for determining the maximum power dissipated by the screen grid.

16.1.14 Changing Parameters

If a tube is to operate at a different plate voltage than published, the new values of bias, screen voltage, and plate resistance can be calculated by the use of conversion factors F_1, F_2, F_3, F_4, and F_5. Assume the following conditions are specified for a single beam-power tube:

Plate voltage250.0V

Screen voltage250.0V

Grid voltage –12.5V

Plate current 45.0mA
Screen current 4.5 mA
Plate resistance 52,000.0Ω
Plate load 5000.0Ω
Transconductance4100.0μS
Power output 4.5 W

F_1 is used to find the new plate voltage

$$F_1 = \frac{E_{p_{new}}}{E_{p_{old}}} \qquad (16\text{-}11)$$

For example, the new plate voltage is to be 180V. The conversion factor F_1 for this voltage is obtained by dividing the new plate voltage by the published plate voltage Eq. 16-11:

$$F_1 = \frac{180}{250}$$
$$= 0.72$$

The screen and grid voltage will be proportional to the plate voltage:

$$E_g = F_1 \times old\ grid\ voltage \qquad (16\text{-}12)$$

$$E_{sg} = F_1 \times old\ screen\ voltage \qquad (16\text{-}13)$$

In the example,

$$E_g = 0.72 \times (-12.5)$$
$$= -9\ V$$

$$E_{sg} = 0.72 \times 250$$
$$= 180\ \text{V}$$

F_2 is used to calculate the plate and screen currents

$$F_2 = F_1\sqrt{F_1} \quad (16\text{-}14)$$

$$I_P = F_2 \times old\ plate\ current \quad (16\text{-}15)$$

$$I_s = F_2 \times old\ screen\ current \quad (16\text{-}16)$$

In the example,

$$\begin{aligned} F_2 &= 0.72 \times 0.848 \\ &= 0.61 \end{aligned}$$

$$\begin{aligned} I_P &= 0.61 \times 45\,\text{mA} \\ &= 27.4\,\text{mA} \end{aligned}$$

$$\begin{aligned} I_{sg} &= 0.61 \times 4.5\,\text{mA} \\ &= 2.74\,\text{mA} \end{aligned}$$

The plate load and plate resistance may be calculated by use of factor F_3:

$$F_3 = \frac{F_1}{F_2} \quad (16\text{-}17)$$

$$r_p = F_3 \times old\ internal\ plate\ resistance \quad (16\text{-}18)$$

$$R_L = F_3 \times old\ \ plateload\ resistance \quad (16\text{-}19)$$

In the example,

$$\begin{aligned} F_3 &= \frac{0.720}{0.610} \\ &= 1.18 \end{aligned}$$

$$\begin{aligned} r_p &= 1.18 \times 52{,}000 \\ &= 61{,}360\,\Omega \end{aligned}$$

$$\begin{aligned} R_L &= 1.18 \times 5000 \\ &= 5900\,\Omega \end{aligned}$$

F_4 is used to find the power output

$$F_4 = F_1F_2 \qquad (16\text{-}20)$$

$$Power\ output = F_4 \times old\ power\ output \qquad (16\text{-}21)$$

In the example:

$$F_4 = 0.72 \times 0.610$$
$$= 0.439$$

$$Power\ output = 0.439 \times 4.5$$
$$= 1.97\ \text{W}$$

F_5 is used to find the transconductance where

$$F_5 = \frac{1}{F_3} \qquad (16\text{-}22)$$

$$transconductance = F_5 \times old\ transconductance \qquad (16\text{-}23)$$

In the example,

$$F_5 = \frac{1}{1.18}$$
$$= 0.847$$

$$transconductance = 0.847 \times 4100$$
$$= (3472\,\mu\text{mho or } \mu\text{S})$$

The foregoing method of converting for voltages other than those originally specified may be used for triodes, tetrodes, pentodes, and beam-power tubes, provided the plate and grid 1 and grid 2 voltages are changed simultaneously by the same factor. This will apply to any class of tube operation, such as class A, AB_1, AB_2, B, or C. Although this method of conversion is quite satisfactory in most instances, the error will be increased as the conversion factor departs from unity. The most satisfactory region of operation will be between 0.7 and 2.0. When the factor falls outside this region, the accuracy of operation is reduced.

16.1.15 Tube Heater

The data sheets of tube manufacturers generally contain a warning that the heater voltage should be maintained within ±10% of the rated voltage. As a rule, this warning is taken lightly, and little attention is paid to heater voltage variations, which have a pronounced effect on the tube characteristics. Internal noise is the greatest offender. Because of heater-voltage variation, emission life is shortened, electrical leakage between elements is increased, heater-to-cathode leakage is increased, and grid current is caused to flow. Thus, the life of the tube is decreased with an increase of internal noise.

16.2 Discrete Solid-State Devices

16.2.1 Semiconductors

Conduction in solids was first observed by Munck and Henry in 1835 and later in 1874 by Braum. In 1905, Col. Dunwoody invented the crystal detector used in the detection of electromagnetic waves. It consisted of a bar of silicon carbide or carborundum held between two contacts. However, in 1903, Pickard filed a patent application for a crystal detector in which a fine wire was placed in contact with the silicon. This was the first mention of a silicon rectifier and was the forerunner of the present-day silicon rectifier. Later, other minerals such as galena (lead sulfide) were employed as detectors. During World War II, intensive research was conducted to improve crystal detectors used for microwave radar equipment. As a result of this research, the original point-contact transistor was invented at the Bell Telephone Laboratories in 1948.

A s*emiconductor* is an electronic device whose main functioning part is made from materials, such as germanium and silicon, whose conductivity ranges between that of a conductor and an insulator.

Germanium is a rare metal discovered by Winkler in Saxony, Germany, in 1896. Germanium is a by-product of zinc mining. Germanium crystals are grown from germanium dioxide powder. Germanium in its purest state behaves much like an insulator because it has very few electrical charge carriers. The conductivity of germanium may be increased by the addition of small amounts of an impurity.

Silicon is a nonmetallic element used in the manufacture of diode rectifiers and transistors. Its resistivity is considerably higher than that of germanium.

The relative position of pure germanium and silicon is given in Fig. 16-8. The scale indicates the resistance of conductors, semiconductors, and insulators per cubic centimeter. Pure germanium has a resistance of approximately $60\Omega/cm^3$. Germanium has a higher conductivity or less resistance to current flow than silicon and is used in low- and medium-power diodes and transistors.

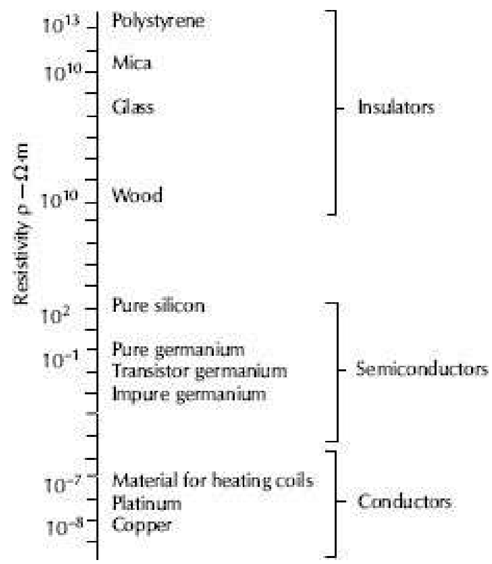


Figure 16-8. Resistance of various materials per cubic centimeter.

The base elements used to make semiconductor devices are not usable as semiconductors in their pure state. They must be subjected to a complex chemical, metallurgical, and photolithographical process wherein the base element is highly refined and then modified with the addition of specific impurities. This precisely controlled process of diffusing impurities into the pure base element is called *doping* and converts the pure base material into a semiconductor material. The semiconductor mechanism is achieved by the application of a voltage across the device with the proper polarity so as to have the device act either as an extremely low resistance (the forward biased or conducting mode) or as an extremely high resistance (reversed bias or nonconducting mode). Because the device is acting as both a good conductor of electricity and also, with the proper reversal of voltage, as a good electrical nonconductor or insulator, it is called a *semiconductor*.

Some semiconductor materials are called *p* or positive type because they are processed to have an excess of positively charged ions. Others are called *n* or negative type because they are processed to have an excess of negatively charged electrons. When a *p*-type of material is brought into contact with an n-type of material, a *pn junction* is formed. With the application of the proper external voltage, a low-resistance path is produced between the *n* and *p* material. By reversing the previously applied voltage, an extremely high-resistance called the *depletion layer* between the *p* and *n* types results. A diode is an example because its conduction depends upon the polarity of the externally applied voltage. Combining several of these *pn* junctions together in a single device produces semiconductors with extremely useful electrical properties.

The theory of operation of a semiconductor device is approached from its atomic structure. The outer orbit of a germanium atom contains four electrons. The atomic structure for a pure germanium crystal is shown in Fig. 16-9A. Each atom containing four electrons forms covalent bonds with adjacent atoms, therefore, there are no “free” electrons. Germanium in its pure state is a poor conductor of electricity. If a piece of “pure” germanium (the size used in a transistor) has a voltage applied to it, only a few microamperes of current caused by electrons that have been broken away from their bonds by thermal agitation will flow in the circuit. This current will increase at an exponential rate with an increase of temperature.

When an atom with five electrons, such as antimony or arsenic, is introduced into the germanium crystal, the atomic structure is changed to that of Fig. 16-9B. The extra electrons (called free electrons) will move toward the positive terminal of the external voltage source.

When an electron flows from the germanium crystal to the positive terminal of the external voltage source, another electron enters the crystal from the negative terminal of the voltage source. Thus, a continuous stream of electrons will flow as long as the external potential is maintained.

The atom containing the five electrons is the *doping agent* or *donor*. Such germanium crystals are classified as *n-type germanium*.

Using a doping agent of indium, gallium, or aluminum, each of which contains only three electrons in its outer orbit, causes the germanium crystal to take the atomic structure of Fig. 16-9C. In this structure, there is a *hole* or *acceptor*. The term *hole* is used to denote a mobile particle that has a positive charge and that simulates the properties of an electron having a positive charge.

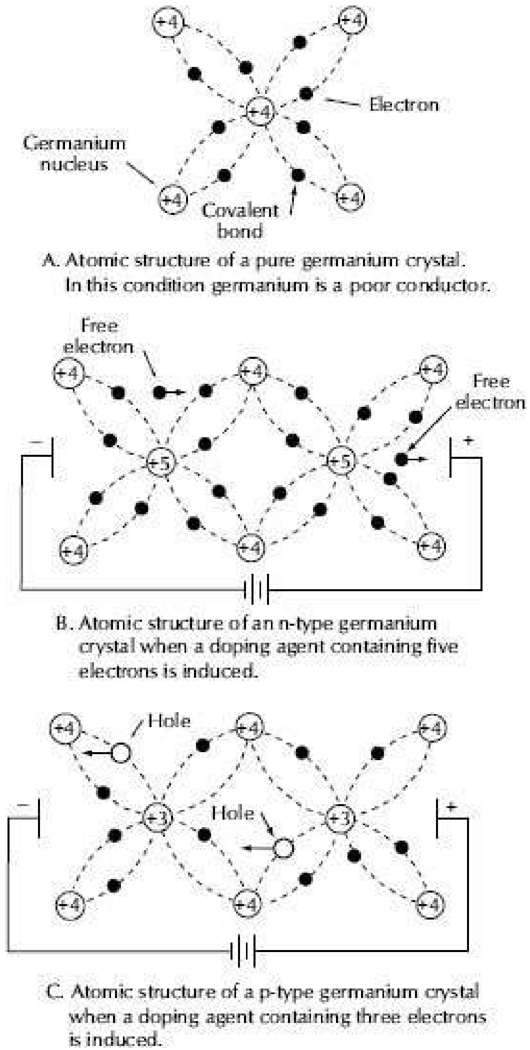


Figure 16-9. Atomic structure of germanium.

When a germanium crystal containing holes is subjected to an electrical field, electrons jump into the holes, and the holes appear to move toward the negative terminal of the external voltage source.

When a hole arrives at the negative terminal, an electron is emitted by the terminal, and the hole is canceled. Simultaneously, an electron from one of the covalent bonds flows into the positive terminal of the voltage source. This new hole moves toward the negative terminal causing a continuous flow of holes in the crystal.

Germanium crystals having a deficiency of electrons are classified *p-type germanium*. Insofar as the external electrical circuits are concerned, there is no difference between electron and hole current flow. However, the method of connection to the two types of transistors differs.

When a germanium crystal is doped so that it abruptly changes from an *n*-type to a *p*-type, and a positive potential is applied to the *p*-region, and a negative potential is applied to the *n*-region, the holes move through the junction to the right and the electrons move to the left, resulting in the voltage-current characteristic shown in Fig. 16-10A. If the potential is reversed, both electrons and holes move away from the junction until the electrical field produced by their displacement counteracts the applied electrical field. Under these conditions, zero current flows in the external circuit. Any minute amount of current that might flow is caused by thermal-generated hole pairs. Fig. 16-10B is a plot of the voltage versus current for the reversed condition. The leakage current is essentially independent of the applied potential up to the point where the junction breaks down.

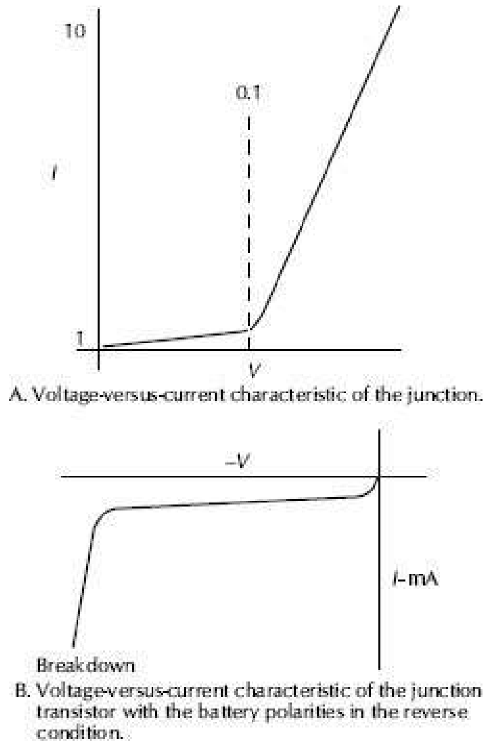


Figure 16-10. Voltage-versus-current characteristics.

16.2.2 Diodes

The *diode* is a device that exhibits a low resistance to current flow in one direction and a high resistance in the other. Ideally, when reverse biasing the diode (connecting the negative of the supply to the diode anode), no current should flow regardless of the value of voltage impressed across the diode. A forward-biased diode presents a very low resistance to current flow.

Fig. 16-11 shows the actual diode characteristics. Starting with the diode reverse biased, a small reverse current does flow. The size of this reverse-leakage current has been exaggerated for clarity and typically is in the order of nanoamperes. The forward resistance is not constant, and therefore it does not yield a straight-line forward-conduction curve. Instead, it begins high and drops rapidly at relatively low applied voltage. Above a 0.5 to 1V drop it approaches a steep straight line slope (i.e., low resistance).

In the reverse-biased region of Fig. 16-11, when the applied voltage (–V) becomes large enough, the leakage current suddenly begins to increase very rapidly, and the slope of the characteristic curves becomes very steep. Past the *knee* in the characteristic, even a small increase in reverse voltage causes a large increase in the reverse current. This steep region is called the *breakdown* or *avalanche* region of the diode characteristic.

The application of high reverse voltage causes the diode to break down and stop behaving like a diode. *Peak-reverse-voltage* rating, or *prv* is one of the two most important diode parameters. This is also referred to as the *peak-inverse-voltage* rating, or *piv*. This rating indicates how high the reverse voltage can be without approaching the knee and risking breakdown. Additional diode parameters are:

Maximum average current	Causes overheating of the device
Peak repetitive current	Maximum peak value of current on a repetitive basis
Surge current	Absolute maximum allowed current even if just momentary

The maximum average current is limited by power dissipation in the junction. This power dissipation is represented by the product of forward voltage drop (V_F) and the forward current (I_F):

$$P = V_F I_F \qquad (16\text{-}24)$$

Selenium Rectifiers and Diodes. A *selenium rectifier cell* consists of a nickel-plated aluminum baseplate coated with selenium, over which a low-temperature alloy is sprayed. The aluminum base serves as a negative electrode and the alloy as the positive. Current flows from the base plate to the alloy but encounters high resistance in the opposite direction. The efficiency of conversion depends to some extent on the ratio of the resistance in the conducting direction to that of the blocking direction. Conventional rectifiers generally have ratios from 100:1 to 1000:1.

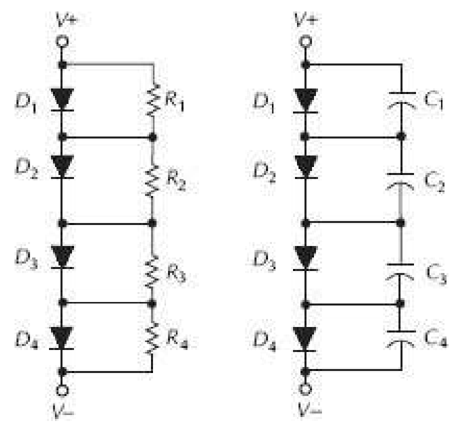


Figure 16-11. Actual diode characteristics.

Selenium rectifiers may be operated over temperatures of −55°C to +150°C (−67°F to +302°F). Rectification efficiency is on the order of 90% for three-phase bridge circuits and 70% for single-phase bridge circuits. As a selenium cell ages, the forward and reverse resistance increases for approximately one year and then stabilizes, decreasing the output voltage by approximately 15%. The internal impedance of a selenium rectifier is low and exhibits a nonlinear characteristic with respect to the applied voltage, maintaining a good voltage regulation. They are often used for battery charging.

Selenium rectifiers, because of their construction, have considerable internal capacitance which limits their operating range to audio frequencies. Approximate capacitance ranges are 0.10 to 0.15μF/in^2 of rectifying surface.

The minimum voltage required for conduction in the forward direction is termed the *threshold voltage* and is about 1V, therefore, selenium rectifiers cannot be used successfully below that voltage.

Silicon Rectifiers and Diodes. The high forward-to-reverse current characteristic of the silicon diode produces an efficiency of about 99%. When properly used, silicon diodes have long life and are not affected by aging, moisture, or temperature when used with the proper heat sink.

As an example, four individual diodes of 400Vpiv may be connected in series to withstand a piv of 1600V. In a series arrangement, the most important consideration is that the applied voltage be equally distributed between the several units. The voltage drops across each individual unit must be very nearly identical. If the instantaneous voltage is not equally divided, one of the units may be subjected to a voltage exceeding its rated value, causing it to fail. This causes the other rectifiers to absorb the piv, often creating destruction of all the rectifiers.

Uniform voltage distribution can be obtained by the connection of capacitors or resistors in parallel with the individual rectifier unit, Fig. 16-12. Shunt resistors are used for steady-state applications, and shunt capacitors are used in applications where transient voltages are expected. If the circuit is exposed to both dc and ac, both shunt capacitors and resistors should be employed.

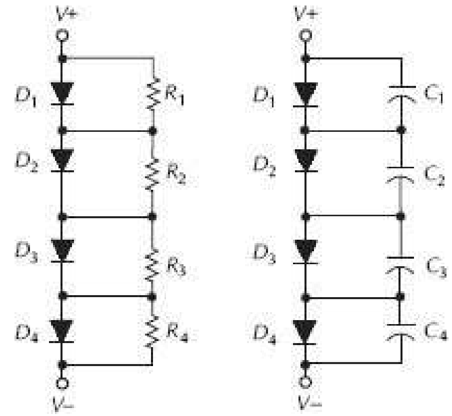


Figure 16-12. Rectifiers connected in series.

When the maximum current of a single diode is exceeded, two or more units may be connected in parallel. To avoid differences in voltage drop across the individual units, a resistor or small inductor is connected in series with each diode, Fig. 16-13. Of the two methods, the inductance is favored because of the lower voltage drop and consumption of power.

Zener and Avalanche Diodes. When the reverse voltage is increased beyond the breakdown knee of the diode characteristics as shown in Fig. 16-11, the diode impedance suddenly drops sharply to a very low value. If the current is limited by an external circuit resistance, operating in the "zener region" is normal for certain diodes specifically designed for the purpose. In *zener diodes*, sometimes simply called *zeners*, the breakdown characteristic is deliberately made as vertical as possible in the zener region so that the voltage across the diode is essentially constant over a wide reverse-current range, acting as a voltage regulator. Since its zener region voltage can be made highly repeatable and very stable with respect to time and temperature, the zener diode can also function as a voltage reference. Zener diodes come in a wide variety of voltages, currents, and powers, ranging from 3.2V to hundreds of volts, from a few milliamperes to 10 A or more, and from about 250mW to over 50W.

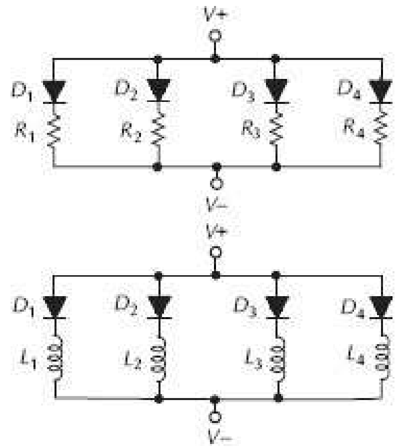


Figure 16-13. Rectifiers connected in parallel.

Avalanche diodes are diodes in which the shape of the breakdown knee has been controlled, and the leakage current before breakdown has been reduced so that the diode is especially well suited to two applications: high-voltage stacking and clamping. In other words, they prevent a circuit from exceeding a certain value of voltage by causing breakdown of the diode at or just below that voltage.

Small-Signal Diodes. *Small-signal diodes* or *general-purpose diodes* are low-level devices with the same general characteristics as power diodes. They are smaller, dissipate much less power, and are not designed for high-voltage, high-power operation. Typical rating ranges are:

I_F (forward current)	1 to 500mA
V_F (forward voltage drop at I_F)	0.2 to 1.1V
piv or *prv*	6 to 1000V
I_R (leakage current at 80% *prv*)	0.1 to 1.0μA

Switching Diodes. *Switching diodes* are small-signal diodes used primarily in digital-logic and control applications in which the voltages may change very rapidly so that speed, particularly reverse-recovery time, is of paramount importance. Other parameters of particular importance are low shunt capacitance, low and uniform V_F (forward voltage drop), low I_R (reverse leakage current), and in control circuits, *prv*.

Noise Diodes. *Noise diodes* are silicon diodes used in the avalanche mode (reverse biased beyond the breakdown knee) to generate broadband noise signals. All diodes generate some noise; these, however, have special internal geometry and are specially processed so as to generate uniform noise power over very broad bands. They are low-power devices (typically, 0.05 to 0.25W) and are available in several different bandwidth classes from as low as 0kHz to 100kHz to as high as 1000 to 18,000MHz.

Varactor Diodes. *Varactor diodes* are made of silicon or gallium arsenide and are used as adjustable capacitors. Certain diodes, when operated in the reverse-biased mode at voltages below the breakdown value, exhibit a shunt capacitance that is inversely proportional to the applied voltage. By varying the applied reverse voltage, the capacitance of the varactor varies. This effect can be used to tune circuits, modulate oscillators, generate harmonics, and mix signals. Varactors are sometimes referred to as *voltage-tunable trimmer capacitors*.

Tunnel Diodes. The *tunnel diode* takes its name from the tunnel effect, a process where a particle can disappear from one side of a barrier and instantaneously reappear on the other side as though it had tunneled through the barrier element.

Tunnel diodes are made by heavily doping both the p and n materials with impurities, giving them a completely different voltage-current characteristic from regular diodes. This characteristic makes them uniquely useful in many high-frequency amplifiers as well as pulse generators and radio frequency oscillators, Fig. 16-14.

What makes the tunnel diode work as an active element is the negative-resistance region over the voltage range V_d (a small fraction of a volt). In this region, increasing the voltage decreases the current, the opposite of what happens with a normal resistor. Tunnel diodes conduct heavily in the reverse direction; in fact, there is no breakdown knee or leakage region.

16.2.3 Thyristors

Stack four properly doped semiconductor layers in series, pnpn (or npnp), and the result is a four-layer, or Shockley breakover diode. Adding a terminal (gate) to the second layer creates a gate-controlled, reverse-blocking *thyristor*, or *silicon-controlled rectifier* (SCR), as shown in Fig. 16-15A.

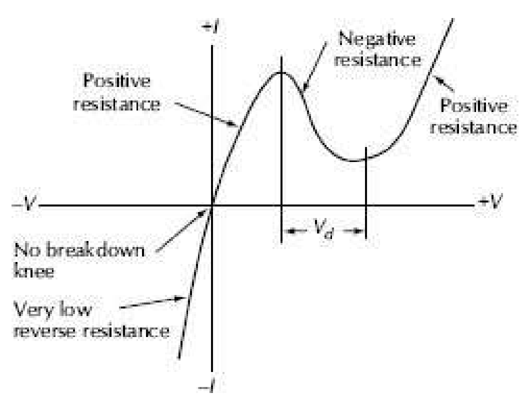


Figure 16-14. Tunnel-diode characteristics showing negative region (tunnel region).

The four-layer diode connects (fires) above a specific threshold voltage. In the SCR, the gate controls this firing threshold voltage, called the *forward blocking voltage*.

To understand how four-layer devices work, separate the material of the layers into two three-layer transistor devices. Fig. 16-15B is an equivalent two-transistor representation in a positive-feedback connection. Assuming a_1 and a_2 are the current gains of the two transistor sections with each gain value less than unity, the total base current I_b into the $n_1p_2n_2$ transistor is

$$I_b = a_1 a_2 I_b + I_o + I_g \qquad (16\text{-}25)$$

where,

a_1 and a_2 are the transistor current gains,

I_b is the total base current,

I_o is the leakage current into the base of the $n_1p_2n_2$ transistor,

I_g is the current into the gate terminal.

The circuit turns on and becomes self-latching after a certain turn-on time needed to stabilize the feedback action, when the equality of Eq. 16-25 is achieved. This result becomes easier to understand by solving for I_b, which gives

$$I_b = \frac{I_o + I_g}{1 - a_1 a_2} \qquad (16\text{-}26)$$

When the product a_1a_2 is close to unity, the denominator approaches zero and I_b approaches a large value. For a given leakage current I_o, the gate current to fire the device can be extremely small. Moreover, as I_b becomes large, I_g can be removed, and the feedback will sustain the on condition since a_1 and a_2 then approach even closer to unity.

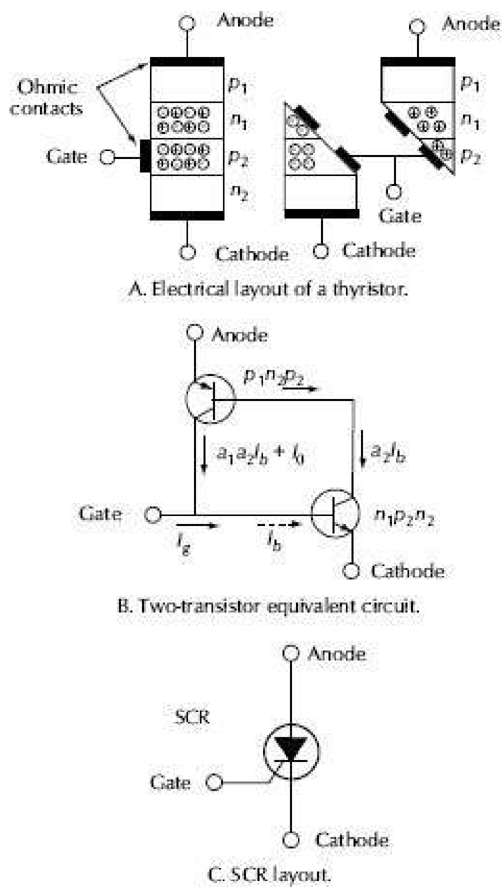


Figure 16-15. Thyristor schematics.

As applied anode voltage increases in the breakover diode, where I_g is absent, I_o also increases. When the quality of Eq. 16-26 is established, the diode fires. The thyristor fires when the gate current I_g rises to establish equality in the equation with the anode voltage fixed. For a fixed I_g, the anode voltage can be raised until the thyristor fires, with I_g determining the firing voltage, Fig. 16-16.

Once fired, a thyristor stays on until the anode current falls below a specified minimum holding current for a certain turnoff time. In addition, the gate loses all control once a thyristor fires. Removal or even reverse biasing of the gate signal will not turn off the device although reverse biasing can help speed turnoff. When the device is used with an ac voltage on the anode, the unit automatically turns off on the negative half of the voltage cycle. In dc switching circuits, however, complex means must often be used to remove, reduce, or reverse the anode voltage for turnoff.

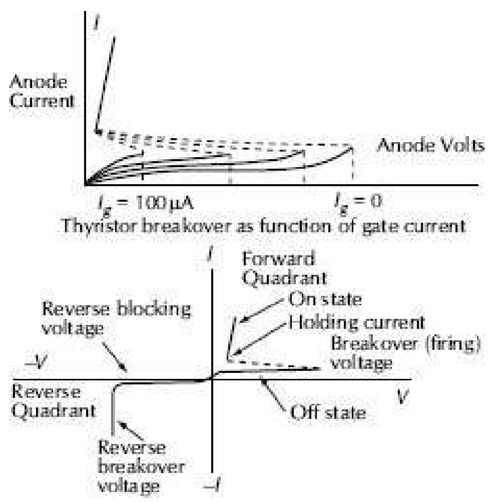


Figure 16-16. Thyristor breakover as a function of gate current and forward voltage.

Fig. 16-17 shows a bilaterally conductive arrangement that behaves very much like two four-layer diodes (*diacs*), or two SCRs (*triacs*), parallel and oppositely conductive. When terminal A is positive and above the breakover voltage, a path through $p_1n_1p_2n_2$ can conduct; when terminal B is positive, path $p_2n_1p_1n_3$ can conduct. When terminal A is positive and a third element, terminal G, is sufficiently positive, the $p_1n_1p_2n_2$ path will fire at a much lower voltage than when G is zero. This action is almost identical with that of the SCR. When terminal G is made negative and terminal B is made positive, the firing point is lowered in the reverse, or $p_2n_1p_1n_3$, direction.

Because of low impedances in the on condition, four-layer devices must be operated with a series resistance in the anode and gate that is large enough to limit the anode-to-cathode or gate current to a safe value.

To understand the low-impedance, high-current capability of the thyristor, the device must be examined as a whole rather than by the two-transistor model. In Fig. 16-17B the $p_1n_1p_2$ transistor has holes injected to fire the unit, and the $n_1p_2n_2$ transistor has electrons injected. Considered separately as two transistors, the space-charge distributions would produce two typical transistor saturation-voltage forward drops, which are quite high when compared with the actual voltage drop of a thyristor.

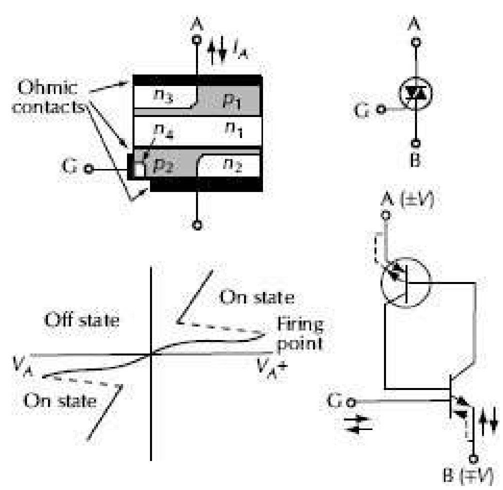


Figure 16-17. Bilateral arrangement to create a triac or ac operating device.

However, when the thyristor shown in Fig. 16-17A is considered, the charges of both polarities exist simultaneously in the same n_1 and p_2 regions. Therefore, at the high injection levels that exist in thyristors, the mobile-carrier concentration of minority carriers far exceeds that from the background-doping density. Accordingly, the space charge is practically neutralized so that the forward drop becomes almost independent of the current density to high current levels. The major resistance to current comes from the ohmic contacts of the unit and load resistance.

The price paid for this low-impedance capability in a standard thyristor is a long turnoff time relative to turn-on time necessary to allow the high level of minority current carriers to dissipate. This long turnoff time limits the speed of a thyristor. Fortunately, this long turnoff time does not add significantly to switching power losses the way that a slow turnon time would.

Turnoff time is the minimum time between the forward anode current ceasing and the device being able to block reapplied forward voltage without turning on again.

Reverse-recovery time is the minimum time after forward conduction ceases that is needed to block reverse-voltage with ac applied to the anode-cathode circuit.

A third specification, *turnon time*, is the time a thyristor takes from the instant of triggering to when conduction is fully on.

These timing specifications limit the operating frequency of a thyristor. Two additional important specifications, the derivative of voltage with respect to time (dv/dt) and the derivative of current with respect to time ($di/d\,t$) limit the rates of change of voltage and current application to thyristor terminals.

A rapidly varying anode voltage can cause a thyristor to turn on even though the voltage level never exceeds the forward breakdown voltage. Because of capacitance between the layers, a current large enough to cause firing can be generated in the gated layer. Current through a capacitor is directly proportional to the rate of change of the applied voltage; therefore, the $dv/d\,t$ of the anode voltage is an important thyristor specification.

Turnon by the *dv/d t* can be accomplished with as little as a few volts per microsecond in some units, especially in older designs. Newer designs are often rated in tens to hundreds of volts per microsecond.

The other important rate effect is the anode-current *di/dt* rating. This rating is particularly important in circuits that have low inductance in the anode-cathode path. Adequate inductance would limit the rate of current rise when the device fires.

When a thyristor fires, the region near the gate conducts first; then the current spreads to the rest of the semiconductor material of the gate-controlled layer over a period of time. If the current flow through the device increases too rapidly during this period because the input-current *di/dt* is too high, the high concentration of current near the gate could damage the device do to localized overheating. Specially designed gate structures can speed up the turnon time of a thyristor, and thus its operational frequency, as well as alleviate this hot-spot problem.

Silicon-Controlled Rectifiers. The SCR thyristor can be considered a solid-state latching relay if dc is used as the supply voltage for the load. The gate current turns on the SCR, which is equivalent to closing the contacts in the load circuit.

If ac is used as the supply voltage, the SCR load current will reduce to zero as the positive ac wave shape crosses through zero and reverses its polarity to a negative voltage. This will shut off the SCR. If the positive gate voltage is also removed it will not turn on during the next positive half cycle of applied ac voltage unless positive gate voltage is applied.

The SCR is suitable for controlling large amounts of rectifier power by means of small gate currents. The ratio of the load current to the control current can be several thousand to one. For example, a 10A load current might be triggered on by a 5mA control current.

The major time-related specification associated with SCRs is the d*v*/d*t* rating. This characteristic reveals how fast a transient spike on the power line can be before it false-triggers the SCR and starts its conducting without gate control current. Apart from this time-related parameter and its gate characteristics, SCR ratings are similar to those for power diodes.

SCRs can be used to control dc by using *commutating circuits* to shut them off. These are not needed on ac since the anode supply voltage reverses every half cycle. SCRs can be used in pairs or sets of pairs to generate ac from dc in inverters. They are also used as protective devices to protect against excessive voltage by acting as a short-circuit switch. These are commonly used in power supply *crowbar* overvoltage protection circuits. SCRs are also used to provide switched power-amplification, as in solid-state relays.

Triacs. The triac in Fig. 16-18 is a three-terminal semiconductor that behaves like two SCRs connected back to front in parallel so that they conduct power in both directions under control of a single gate-control circuit. Triacs are widely used to control ac power by phase shifting or delaying the gate-control signal for some fraction of the half cycle during which the power diode could be conducting. Light dimmers found in homes and offices and variable-speed drills are good examples of triac applications.

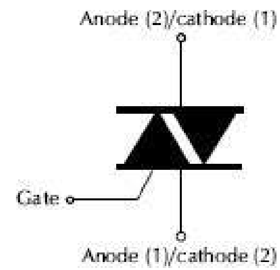


Figure 16-18. Schematic of a triac.

Light-Activated Silicon-Controlled Rectifiers. When sufficient light falls on the exposed gate junction, the SCR is turned on just as if the gate-control current were flowing. The gate terminal is also provided for optional use in some circuits. These devices are used in projector controls, positioning controls, photo relays, slave flashes, and security protection systems.

Diacs. The diac is shown in Fig. 16-19. It acts as two zener (or avalanche) diodes connected in series, back to back. When the voltage across the diac in either direction gets large enough, one of the zeners breaks down. The action drops the voltage to a lower level, causing a current increase in the associated circuit. This device is used to trigger SCRs or triacs.

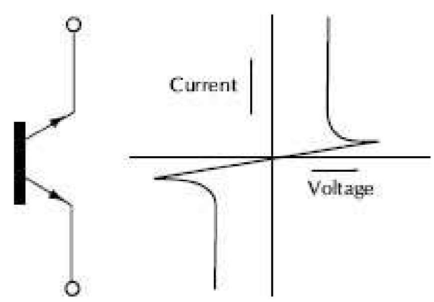


Figure 16-19. Schematic of a diac.

Opto-Coupled Silicon-Controlled Rectifiers. An *optocoupled SCR* is a combination of a light-emitting diode (LED) and a photo silicon-controlled rectifier (photo-SCR). When sufficient current is forced through the LED, it emits an infrared radiation that triggers the gate of the photo-SCR. A small control current can regulate a large load current, and the device provides insulation and isolation between the control circuit (the LED) and the load circuit (the SCR). Opto-coupled transistors and Darlington transistors that operate on the same principle will be discussed later.

16.2.4 Transistors

There are many different types of transistors,[1] and they are named by the way they are grown, or made. Fig. 16-20A shows the construction of a *grown-junction* transistor. An *alloy-junction* transistor is shown in Fig. 16-20B. During the manufacture of the material for a grown junction, the impurity content of the semiconductor is altered to provide *npn* or *pnp* regions. The grown material is cut into small sections, and contacts are attached to the regions. In the alloy-junction type, small dots of *n*- or *p*-type impurity elements are attached to either side of a thin wafer of *p*- or *n*-type semiconductor material to form regions for the emitter and collector junctions. The base connection is made to the original semiconductor material.

Drift-field transistors, Fig. 16-20C, employ a modified alloy junction in which the impurity concentration in the wafer is diffused or graded. The drift field speeds up the current flow and extends the frequency response of the alloy-junction transistor. A variation of the drift-field transistor is the *microalloy diffused* transistor, as shown in Fig. 16-20D. Very narrow base dimensions are achieved by etching techniques, resulting in a shortened current path to the collector.

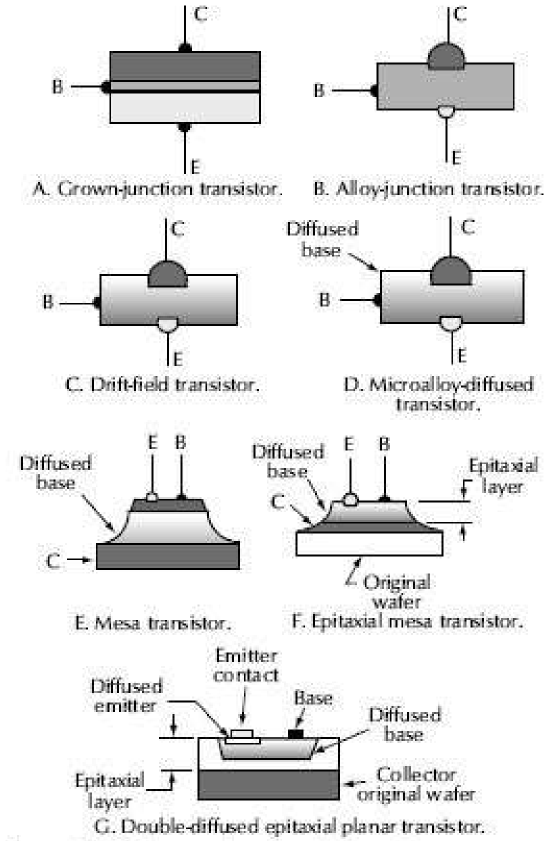


Figure 16-20. Construction of various transistors.

Mesa transistors shown in Fig. 16-20E use the original semiconductor material as the collector, with the base material diffused into the wafer and an emitter dot alloyed into the base region. A flat-topped peak or mesa is etched to reduce the area of the collector at the base junction. Mesa devices have large power-dissipation capabilities and can be operated at very high frequencies. *Double-diffused epitaxial mesa transistors* are grown by the use of vapor deposition to build up a crystal layer on a crystal wafer and will permit the precise control of the physical and electrical dimensions independently of the nature of the original wafer. This technique is shown in Fig. 16-20F.

The *planar transistor* is a highly sophisticated method of constructing transistors. A limited area source is used for both the base diffusion and emitter diffusion, which provides a very small active area, with a large wire contact area. The advantage of the planar construction is its high dissipation, lower leakage current, and lower collector cut-off current, which increases the stability and reliability. Planar construction is also used with several of the previously discussed base designs. A double-diffused epitaxial planar transistor is shown in Fig. 16-20G.

The principle difference between a conventional transistor and the *field-effect transistor* is the transistor is a current-controlled device, while the FET is voltage controlled, similar to the vacuum tube. Conventional transistors also have a low-input impedance, which may at times complicate the circuit designer's problems. The FET has a high-input impedance with a low-output impedance, much like a vacuum tube.

The basic principles of the FET operation can best be explained by the simple mechanism of a *pn* junction. The control mechanism is the creation and control of a depletion layer, which is common to all reverse-biased junctions. Atoms in the *n* region possess excess electrons that are available for conduction, and the atoms in the *p* region have excess holes that may also allow current to flow. Reversing the voltage applied to the junction and allowing time for stabilization, very little current flows, but a rearrangement of the electrons and holes will occur. The positively charged holes will be drawn toward the negative terminals of the voltage source, and the electrons, which are negative, will be attracted to the positive terminal of the voltage source. This results in a region being formed near the center of the junction having a majority of the carriers removed and therefore called the *depletion regions*.

Referring to Fig. 16-21A, a simple bar composed of *n*-type semiconductor material has a nonrectifying contacts at each end. The resistance between the two end electrodes is

$$R = \frac{PL}{WT} \tag{16-27}$$

where,

P is the function of the material sensitivity,

L is the length of the bar,

W is the width,

T is the thickness.

Varying one or more of the variables of the resistance of the semiconductor changes the bar. Assume a p-region in the form of a sheet is formed at the top of the bar shown in Fig. 16-21B. A *pn* junction is formed by diffusion, alloying, or epitaxial growth creating a reverse voltage between the *p* and *n*-material producing two depletion regions. Current in the n-material is caused primarily by means of excess electrons.

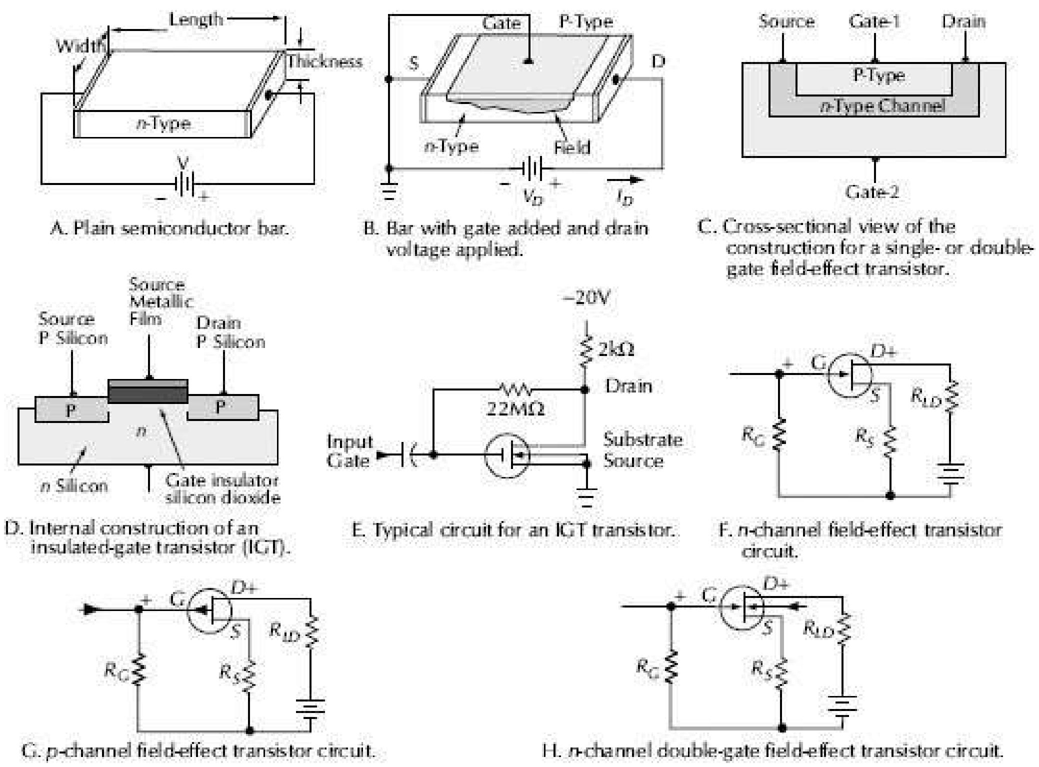


Figure 16-21. Field-effect transistors (FETs).

By reducing the concentration of electrons or majority carriers, the resistivity of the material is increased. Removal of the excess electrons by means of the depletion region causes the material to become practically nonconductive.

Disregarding the *p* region and applying a voltage to the ends of the bar cause a current and create a potential gradient along the length of the bar material, with the voltage increasing toward the right, with respect to the negative end or ground. Connecting the *p* region to ground causes varying amounts of reverse-bias voltage across the *pn* junction, with the greatest amount developed toward the right end of the *p* region. A reverse voltage across the bar will produce the same depletion regions. If the resistivity of the p-type material is made much smaller than that of the *n*-type material, the depletion region will then extend much farther into the *n* material than into the *p* material. To simplify the following explanation, the depletion of *p* material will be ignored.

The general shape of the depletion is that of a wedge, increasing the size from left to right. Since the resistivity of the bar material within the depletion area is increased, the effective thickness of the conducting portion of the bar becomes less and less, going from the end of the *p* region to the right end. The overall resistance of the semiconductor material is greater because the effective thickness is being reduced. Continuing to increase the voltage across the ends of the bar, a point is reached where the depletion region is extended practically all the way through the bar, reducing the effective thickness to zero. Increasing the voltage beyond this point produces little change in current.

The *p* region controls the action and is termed a *gate*. The left end of the bar, being the source of majority carriers, is termed the *source*. The right end, being where the electrons are drained off, is called the *drain*. A cross-sectional drawing of a typical FET is shown in Fig. 16-21C, and three basic circuits are shown in Fig. 16-21F–H.

Insulated-gate transistors (IGT) are also known as field-effect transistors, metal-oxide silicon or semiconductor field-effect transistors (MOSFET), metal-oxide silicon or semiconductor transistors (MOST), and insulated-gate field-effect transistors (IGFET). All these devices are similar and are simply names applied to them by the different manufacturers.

The outstanding characteristics of the IGT are its extremely high input impedance, running to $10^{15}\Omega$. IGTs have three elements but four connections—the gate, the drain, the source, and an *n*-type substrate, into which two identical *p*-type silicon regions have been diffused. The source and drain terminals are taken from these two p regions, which form a capacitance between the *n* substrate and the silicon-dioxide insulator and the metallic gate terminals. A cross-sectional view of the internal construction appears in Fig. 16-21D, with a basic circuit shown in Fig. 16-21E. Because of the high input impedance, the IGT can easily be damaged by static charges. Strict adherence to the instructions of the manufacturer must be followed since the device can be damaged even before putting it into use.

IGTs are used in electrometers, logic circuits, and ultrasensitive electronic instruments. They should not be confused with the conventional FET used in audio equipment.

Transistor Equivalent Circuits, Current Flow, and Polarity. Transistors may be considered to be a T configuration active network, as shown in Fig. 16-22

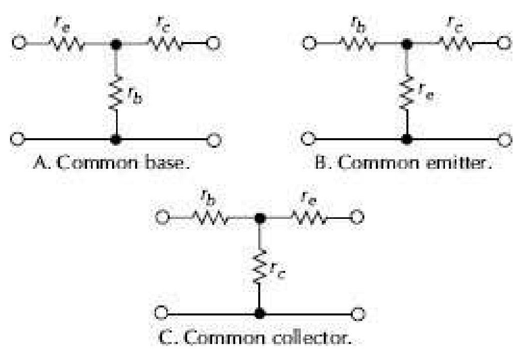


Figure 16-22. Equivalent circuits for transistors.

The current flow, phase, and impedances of the *npn* and *pnp* transistors are shown in Fig. 16-23 for the three basic configurations, common emitter, common base and common collector. Note phase reversal only takes place in the common-emitter configuration.

The input resistance for the common-collector and common-base configuration increases with an increase of the load resistance R_L. For the common emitter, the input resistance decreases as the load resistance is increased; therefore, changes of input or output resistance are reflected from one to the other.

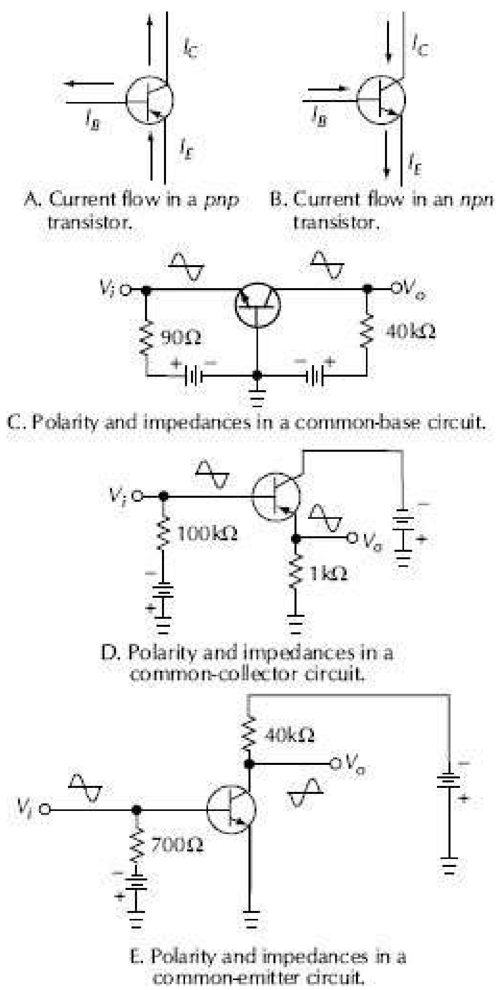


Figure 16-23. Current, polarity. and impedance relationships.

Fig. 16-24 shows the signal-voltage polarities of a *p*-channel field-effect transistor. Note the similarity to tube characteristics.

Voltage, power, and current gains for a typical transistor using a common-emitter configuration are shown in Fig. 16-25. The current gain decreases as the load resistance is increased, and the voltage gain increases as the load resistance is increased. Maximum power gain occurs when the load resistance is approximately 40,000Ω, and it may exceed unity.

For the common-collector connection, the current gain decreases as the load resistance is increased and the voltage gain increases as the load resistance is increased, but it never exceeds unity. Curves such as these help the designer to select a set of conditions for a specific result.

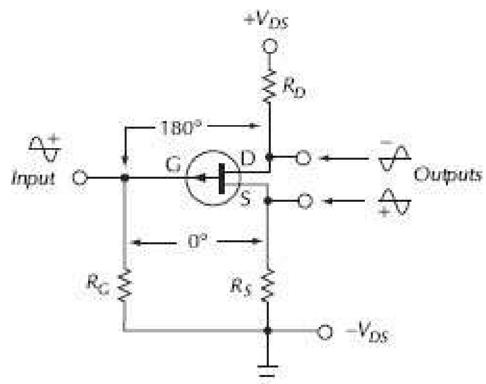


Figure 16-24. Signal-voltage polarities in a p-channel field-effect transistor (FET).

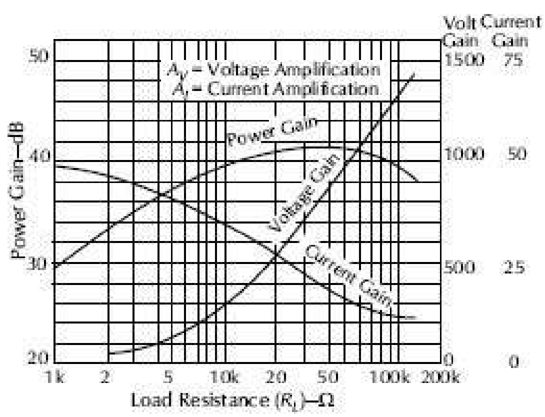


Figure 16-25. Typical voltage, power, and current gains for a conventional transistor using a common-emitter configuration.

The power gain varies as the ratio of the input to output impedance and may be calculated with the equation

$$dB = 10\log\frac{Z_o}{Z_{in}} \qquad (16\text{-}28)$$

where,

Z_o is the output impedance in Ω,

Z_{in} is the input impedance in Ω.

Forward-Current-Transfer Ratio. An important characteristic of a transistor is its *forward-current-transfer ratio*, or the ratio of the current in the output to the current in the input element. Because of the many different configurations for connecting transistors, the forward transfer ratio is specified for a particular circuit configuration. The forward-current-transfer ratio for the common-base configuration is often referred to as *alpha* (α) and the common-emitter forward-current-transfer ratio as *beta* (β). In common-base circuitry, the emitter is the input element, and the collector is the output element. Therefore, α_{dc} is the ratio of the dc collector current I_C to the dc emitter current I_E. For the common emitter, the β_{dc} is then the ratio of the dc collector current I_C to the base current I_B. The ratios are also given in terms of the ratio of signal current, relative to the input and output, or in terms of ratio of change in the output current to the input current, which causes the change.

The terms α and β are also used to denote the frequency cutoff of a transistor and is defined as the frequency at which the value of α for a common-base configuration, or β for a common-emitter circuit, falls to 0.707 times its value at a frequency of 1000Hz.

Gain-bandwidth product is the frequency at which the common-emitter forward-current-transfer ratio β is equal to unity. It indicates the useful frequency range of the device and assists in the determination of the most suitable configuration for a given application.

Bias Circuits. Several different methods of applying bias voltage to transistors are shown in Fig. 16-26, with a master circuit for aiding in the selection of the proper circuit shown in Fig. 16-27. Comparing the circuits shown in Fig. 16-26, their equivalents may be found by making the resistors in Fig. 16-27 equal to zero or infinity for analysis and study. As an example, the circuit of Fig. 16-26D may be duplicated in Fig. 16-27 by shorting out resistors R4 and R5 in Fig. 16-27.

The circuit Fig. 16-26G employs a split voltage divider for R_2. A capacitor connected at the junction of the two resistors shunts any ac feedback current to ground. The stability of circuits A, D, and G in Fig. 16-26 may be poor unless the voltage drop across the load resistor is at least one-third the value of the power supply voltage V_{cc}. The final determining factors will be gain and stability.

Stability may be enhanced by the use of a thermistor to compensate for increases in collector current with increasing temperature. The resistance of the thermistor decreases as the temperature increases, decreasing the bias voltage so the collector voltage tends to remain constant. Diode biasing may also be used for both temperature and voltage variations. The diode is used to establish the bias voltage, which sets the transistor idling current or the current flow in the quiescent state.

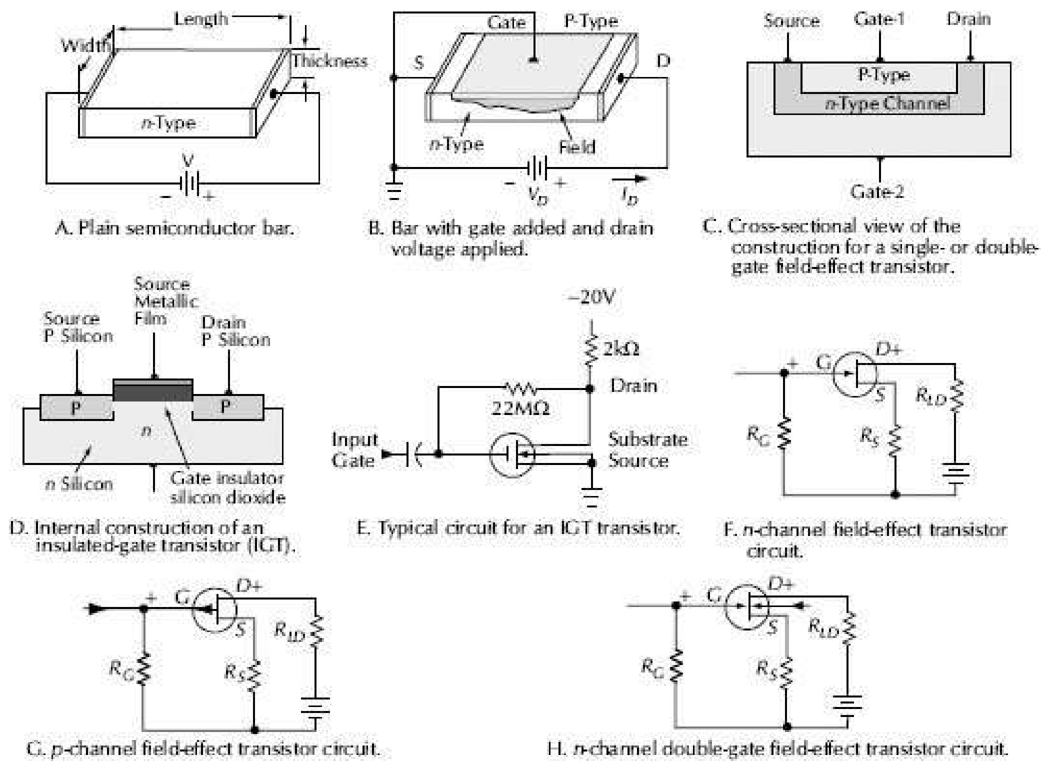


Figure 16-26. Basic design circuit for transistor bias circuits.

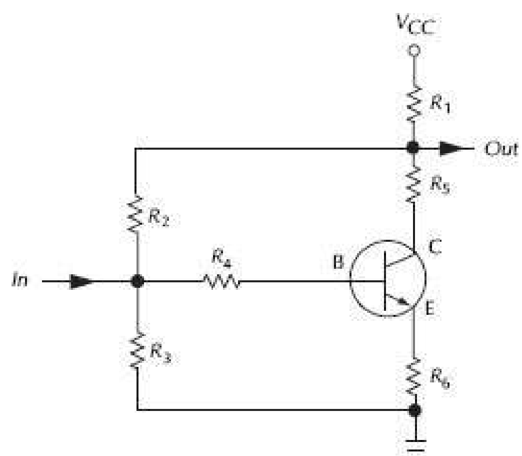


Figure 16-27. Basic bias circuits for transistors

When a transistor is biased to a nonconducting state, small reverse dc currents flow, consisting of leakage currents that are related to the surface characteristics of the semiconductor material and saturation currents. Saturation current increases with temperature and is related to the impurity concentration in the material. Collector-cutoff current is a dc current caused when the collector-to-base circuit is reverse biased and the emitter-to-base circuit is open. Emitter-cutoff current flows when the emitter to base is reverse biased and the collector-to-base circuit is open.

Small- and Large-Signal Characteristics. The transistor, like the vacuum tube, is nonlinear and can be classified as a nonlinear active device. Although the transistor is only slightly nonlinear, these nonlinearities become quite pronounced at very low and very high current and voltage levels. If an ac signal is applied to the base of a transistor without a bias voltage, conduction will take place on only one-half cycle of the applied signal voltage, resulting in a highly distorted output signal. To avoid high distortion, a dc-biased voltage is applied to the transistor, and the operating point is shifted to the linear portion of the characteristic curve. This improves the linearity and reduces the distortion to a value suitable for small-signal operation. Even though the transistor is biased to the most linear part of the characteristic curve, it can still add considerable distortion to the signal if driven into the nonlinear portion of the characteristic.

Small-signal swings generally run from less than 1μV to about 10mV so it is important that the dc-biased voltage be large enough that the applied ac signal is small compared to the dc bias current and voltage. Transistors are normally biased at current values between 0.1mA and 10mA. For large-signal operation, the design procedures become quite involved mathematically and require a considerable amount of approximation and the use of nonlinear circuit analysis.

It is important to provide an impedance match between cascaded stages because of the wide difference of impedance between the input and output circuits of transistors. If the impedances are not matched, an appreciable loss of power will take place.

The maximum power amplification is obtained with a transistor when the source impedance matches the internal input resistance, and the load impedance matches the internal output resistance. The transistor is then said to be image matched.

If the source impedance is changed, it affects the internal output resistance of the transistor, requiring a change in the value of the load impedance. When transistor stages are connected in tandem, except for the grounded-emitter connection, the input impedance is considerably lower than the preceding stage output impedance. Therefore, an interstage transformer should be used to supply an impedance match in both directions.

When working between a grounded base and a grounded-emitter circuit, a step-down transformer is used. Working into a grounded-collector stage, a step-up transformer is used. Grounded-collector stages can also be used as an impedance-matching device between other transistor stages.

When adjusting the supply voltages for a transistor amplifier employing transformers, the *battery* voltage must be increased to compensate for the dc voltage drop across the transformer windings. The data sheets of the manufacturer should be consulted before selecting a transformer to determine the source and load impedances.

Transistor Noise Figure (*nf*). In a low-level amplifier, such as a preamplifier, noise is the most important single factor and is stated as the *SNR* or *nf*. Most amplifiers employ resistors in the input circuit which contribute a certain amount of measurable noise because of thermal activity. This power is generally about –160dB, re: 1W, for a bandwidth of 10,000Hz. When the input signal is amplified, the noise is also amplified. If the ratio of the signal power to noise power is the same, the amplifier is noiseless and has a noise figure of unity or more. In a practical amplifier some noise is present, and the degree of impairment is the noise figure (*nf*) of the amplifier, expressed as the ratio of signal power to noise power at the output:

$$nf = \frac{S_1 \times N_o}{S_o \times N_1} \qquad (16\text{-}29)$$

where,

S_1 is the signal power,

N_1 is the noise power,

S_o is the signal power at the output,

N_o is the noise at the output.

$$nf_{dB} = 10\log(nf \text{ of the power ratio}) \qquad (16\text{-}30)$$

For an amplifier with various *nf*, the *SNR* would be:

nf	*SNR*
1dB	1.26
3dB	2
10dB	10
20dB	100

An amplifier with an *nf* below 6 dB is considered excellent.

Low *nf* can be obtained by the use of an emitter current of less than 1mA, a collector voltage of less than 2V, and a signal-source resistance below 2000Ω.

Internal Capacitance. The paths of internal capacitance in a typical transistor are shown in Fig. 16-28. The width of the *pn* junction in the transistor varies in accordance with voltage and current, and the internal capacitance also varies. Variation of collector-base capacitance *C* with collector voltage and emitter current is shown in Figs. 16-28B and C. The increase in the width of the *pn* junction between the base and collector, as the reverse bias voltage (V_{CB}) is increased, is reflected in lower capacitance values. This phenomenon is equivalent to increasing the spacing between the plates of a capacitor. An increase in the emitter current, most of which flows through the base-collector junction, increases the collector-base capacitance (C_{CB}). The increased current through the *pn* junction may be considered as effectively reducing the width of the pn junction. This is equivalent to decreasing the spacing between the plates of a capacitor, therefore increasing the capacitance.

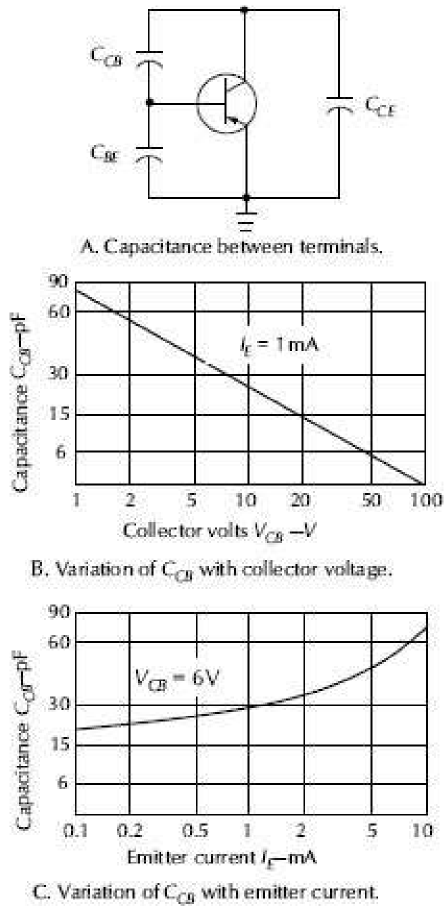


Figure 16-28. Internal capacitance of a transistor.

The average value of collector-base capacitance (C_{CB}) varies from 2 to 50pF, depending on the type transistor and the manufacturing techniques. The collector-emitter capacitance is caused by the pn junction. It normally is five to ten times greater than that of the collector-base capacitance and will vary with the emitter current and collector voltage.

Punch-Through. *Punch-through* is the widening of the space charge between the collector element and the base of a transistor. As the potential V_{CB} is increased from a low to a high value, the collector-base space charge is widened. This widening effect of the space charge narrows the effective width of the base. If the diode space charge does not avalanche before the space charge spreads to the emitter section, a phenomenon termed punch-through is encountered, as shown in Fig. 16-29.

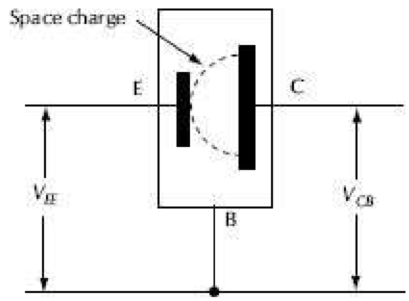


Figure 16-29. Spreading of the space charge between the emitter and the collector, which creates punch-through.

The effect is the base disappears as the collector-base space-charge layer contacts the emitter, creating relatively low resistance between the emitter and the collector. This causes a sharp rise in the current. The transistor action then ceases. Because there is no voltage breakdown in the transistor, it will start functioning again if the voltage is lowered to a value below where punch-through occurs.

When a transistor is operated in the punch-through region, its functioning is not normal, and heat is generated internally that can cause permanent damage to the transistor.

Breakdown Voltage. *Breakdown voltage* is that voltage value between two given elements in a transistor at which the crystal structure changes and current begins to increase rapidly. Breakdown voltage may be measured with the third electrode open, shorted, or biased in either the forward or reverse direction. A group of collector characteristics for different values of base bias are shown in Fig. 16-30. The collector-to-emitter breakdown voltage increases as the base-to-emitter bias is decreased from the normal forward values through zero to reverse. As the resistance in the base-to-emitter circuit decreases, the collector characteristics develop two breakdown points. After the initial breakdown, the collector-to-emitter voltage decreases with an increasing collector current, until another breakdown occurs at the lower voltage.

Breakdown can be very destructive in power transistors. A breakdown mechanism, termed *second breakdown*, is an electrical and thermal process in which current is concentrated in a very small area. The high current, together with the voltage across the transistor, causes intense heating, melting a hole from the collector to the emitter. This causes a short circuit and internal breakdown of the transistor.

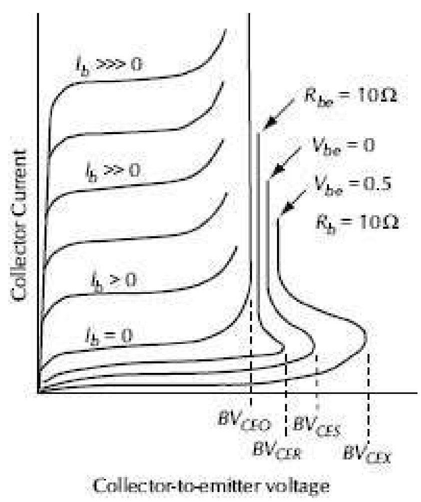


Figure 16-30. Typical collector characteristic curves showing locations of various breakdown voltages.

The fundamental limitation to the use of transistors is the breakdown voltage (BV_{cer}). The breakdown voltage is not sharp so it is necessary to specify the value of collector current at which breakdown will occur. This data is obtained from the data sheet of the manufacturer.

Transistor Load Lines. Transistor load lines are used to design circuits. An example of circuit design uses a transistor with the following characteristics:

Maximum collector current	10mA
Maximum collector voltage	−22V
Base current	0 to 300μA
Maximum power dissipation	300mW

The base current curves are shown in Fig. 16-31A. The amplifier circuit is to be Class A, using a common-emitter circuit, as shown in Fig. 16-31B. By proper choice of the operating point, with respect to the transistor characteristics and supply voltage, low-distortion Class A performance is easily obtained within the transistor power ratings.

The first requirement is a set of collector-current, collector-voltage curves for the transistor to be employed. Such curves can generally be obtained from the data sheets of the manufacturer. Assuming that such data is at hand and referring to Fig. 16-31A, a curved line is plotted on the data sheet, representing the maximum power dissipation by the use of the equation

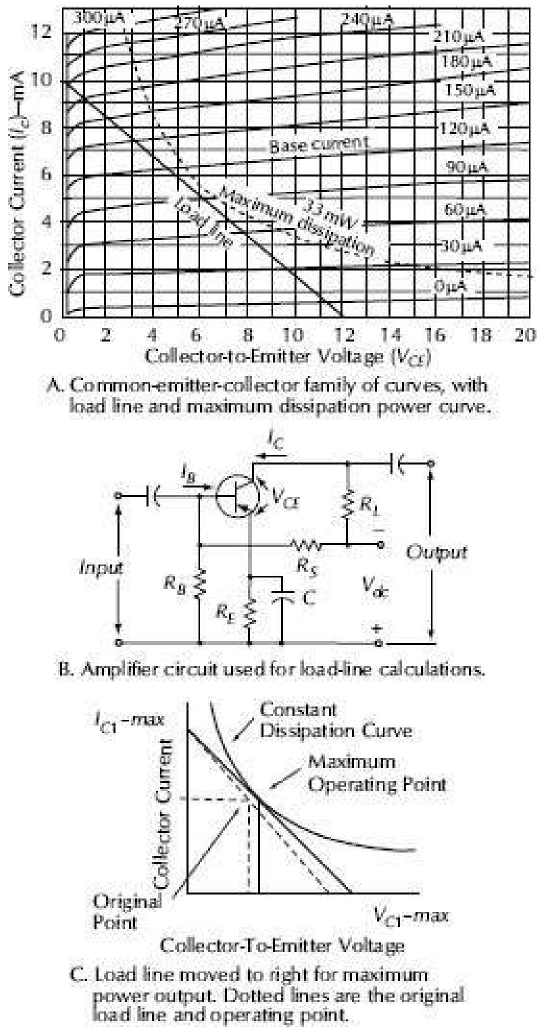


Figure 16-31. Load-line calculation curves.

$$I_c = \frac{P_c}{V_c} \tag{16-31}$$

or

$$V_c = \frac{P_c}{I_c} \tag{16-32}$$

where,

I_c is the collector current,

P_c is the maximum power dissipation of the transistor,

V_c is the collector voltage.

At any point on this line at the intersection of $V_c I_c$, the product equals 0.033W or 33mW. In determining the points for the dissipation curve, voltages are selected along the horizontal axis and the corresponding current is equated using

$$I_C = \frac{P_C}{V_{CE}} \qquad (16\text{-}33)$$

The current is determined for each of the major collector-voltage points, starting at 16V and working backward until the upper end of the power curve intersects the 300μA base current line. After entering the value on the graph for the power dissipation curve, the area to the left of the curve encompasses all points within the maximum dissipation rating of the transistor. The area to the right of the curve is the overload region and is to be avoided.

The operating point is next determined. A point that results in less than a 33mW dissipation is selected somewhere near the center of the power curve. For this example, a 5mA collector current at 6V, or a dissipation of 30mW, will be used. The selected point is indicated on the graph and circled for reference. A line is drawn through the dot to the maximum collector current, 10mA, and downward to intersect the V_{CE} line at the bottom of the graph, which, for this example, is 12V. This line is termed the *load line*. The load resistance R_L may be computed with

$$R_L = \frac{dV_{CE}}{dI_C} \qquad (16\text{-}34)$$

where,

R_L is the load resistance,

dV_{CE} is the range of collector-to-emitter voltage,

dI_C is the range of collector current.
In the example,

$$R_L = \frac{0-12}{0-0.01}$$
$$= \frac{12}{0.01}$$
$$= 1200\,\Omega$$

Under these conditions, the entire load line dissipates less than the maximum value of 33mW, with 90μA of base current and 5mA of collector current. The required base current of 90μA may be obtained by means of one of the biasing arrangements shown in Fig. 16-26.

To derive the maximum power output from the transistor, the load line may be moved to the right and the operating point placed in the maximum dissipation curve, as shown in Fig. 16-31C. Under these conditions, an increase in distortion may be expected. As the operating point is now at 6.5V and 5mA, the dissipation is 33mW. Drawing a line through the new operating point and 10mA (the maximum current), the voltage at the lower end of the load line is 13.0V; therefore, the load impedance is now 1300Ω.

16.3 Integrated Circuits

An *integrated circuit* (IC) is a device consisting of hundreds and even thousands of components in one small enclosure, and came into being when manufacturers learned how to grow and package semiconductors and resistors.

The first ICs were small scale and usually too noisy for audio circuits; however, as time passed, the noise was reduced, stability increased, and the operational amplifier (op-amp) IC became an important part of the audio circuit. With the introduction of medium-scale integration (MSI) and large-scale integration (LSI) circuits, power amplifiers were made on a single chip with only capacitors, gain, and frequency compensation components externally connected.

Typical circuit components might use up a space 4mils × 6mils (1 mil = 0.001 in) for a transistor, 3mils × 4mils for a diode, and 2mils × 12mils for a resistor. These components are packed on the surface of the semiconductor wafer and interconnected by a metal pattern that is evaporated into the top surface. Leads are attached to the wafer that is then sealed and packaged in several configurations, depending on their complexity.

ICs can be categorized by their method of fabrication or use. The most common are monolithic or hybrid and linear or digital. Operational amplifiers and most analog circuits are linear while flip-flops and on–off switch circuits are digital.

An IC is considered *monolithic* if it is produced on one single chip and *hybrid* if it consists of more than one monolithic chip tied together and/or includes discrete components such as transistors, resistors, and capacitors.

With only a few external components, ICs can perform math functions, such as trigonometry, squaring, square roots, logarithms and antilogarithms, integration, and differentiation. ICs are well suited to act as voltage comparators, zero-crossing detectors, ac and dc amplifiers, audio and video amplifiers, null detectors, and sine-, square-, or triangular-wave generators, and all at a fraction of the cost of discrete-device circuits.

16.3.1 Monolithic Integrated Circuits

All circuit elements, both active and passive, are formed at the same time on a single wafer. The same circuit can be repeated many times on a single wafer and then cut to form individual 50mil^2 ICs.

Bipolar transistors are often used in ICs and are fabricated much like the discrete transistor by the planar process. The differences are the contact-to-the-collector region is through the top surface rather than the substrate, requiring electrical isolation between the substrate and the collector. The integrated transistor is isolated from other components by a pn junction that creates capacitance, reducing high-frequency response and increasing leakage current, which in low-power circuits can be significant.

Integrated diodes are produced the same way as transistors and can be regarded as transistors whose terminals have been connected to give the desired characteristics.

Resistors are made at the same time as transistors. The resistance is characterized in terms of its sheet resistance, which is usually 100–200Ω/square material for diffused resistors and 50–150Ω/square material for deposited resistors. To increase the value of a resistor, square materials are simply connected in series.

It is very difficult to produce resistors with much closer tolerance than 10%; however, it is very easy to produce two adjacent resistors to be almost identical. When making comparator-type circuits, the circuits are balanced and are made to perform on ratios rather than absolute values. Another advantage is uniformity in temperature. As the temperature of one component varies, so does the temperature of the other components, allowing good tracking between components and circuits so ICs are usually more stable than discrete circuits.

Capacitors are made as thin-film integrated capacitors or junction capacitors. The thin-film integrated capacitor has a deposited metal layer and an *n*+ layer isolated with a carrier-free region of silicon dioxide. In junction capacitors, both layers are diffused low-resistance semiconductor materials. Each layer has a dopant of opposite polarity; therefore, the carrier-free region is formed by the charge-depleted area at the *pn* junction.

The MOSFET transistor has many advantages over the bipolar transistor for use in ICs as it occupies only $\frac{1}{25}$ the area of the bipolar equivalent due to lack of isolation pads. The MOSFET acts like a variable resistor and can be used as a high-value resistor. For instance, a 100kΩ resistor might occupy only 1mil^2 as opposed to 250mil^2 for a diffused resistor.

The chip must finally be connected to terminals or have some means of connecting to other circuits, and it must also be packaged to protect it from the environment. Early methods included using fine gold wire to connect the chip to contacts. This was later replaced with aluminum wire ultrasonically bonded.

Flip-chip and beam-lead methods eliminate the problems of individually bonding wires. Relatively thick metal is deposited on the contact pads before the ICs are separated from the wafer. The deposited metal is then used to contact a matching metal pattern on the substrate. In the flip-chip method, globules of solder deposited on each contact pad ultrasonically bond the chip to the substrate.

In the beam-lead method, thin metal tabs lead away from the chip at each contact pad. The bonding of the leads to the substrate reduces heat transfer into the chip and eliminates pressure on the chip.

The chip is finally packaged in either hermetically sealed metal headers or is encapsulated in plastic, which is an inexpensive method of producing ICs.

16.3.2 Hybrid Integrated Circuits

Hybrid circuits combine monolithic and thick- and thin-film discrete components for obtaining the best solution to the design.

Active components are usually formed as monolithics; however, sometimes discrete transistors are soldered into the hybrid circuit.

Passive components such as resistors and capacitors are made with thin- and thick-film techniques. Thin films are 0.001–0.1mil thick, while thick films are normally 60mils thick. Resistors can be made with a value from ohms to megohms with a tolerance of 0.05% or better.

High-value capacitors are generally discrete, miniature components that are welded or soldered into the circuit, and low-value capacitors can be made as film capacitors and fabricated directly on the substrate.

Along with being certain that the components will fit into the hybrid package, the temperature must also be taken into account. The temperature rise T_R of the package can be calculated with the following equation

$$\begin{aligned} T_R &= T_C - T_A \\ &= P_T T \theta_{CA} \end{aligned} \qquad (16\text{-}35)$$

where,

T_C is the case temperature,

T_A is the ambient temperature,

P_T is the total power dissipation,

θ_{CA} is the case-to-ambient thermal resistance.

The θ_{CA} for a package in free air can be approximated at 35°C/W/in^2 or a device will have a 35°C rise in temperature above ambient if 1W is dissipated over an area of 1in^2.

16.3.3 Operational Voltage Amplifiers (Op-Amp)

One of the most useful ICs for audio is the op-amp. Op-amps can be made with discrete components, but they would be very large and normally unstable to temperature and external noise.

An op-amp normally has one or more of the following features:

- Very high input impedance (>10^6 to $10^{12}\Omega$).
- Very high open-loop (no feedback) gain.
- Low output impedance (<200Ω).
- Wide frequency response (>100MHz).
- Low input noise.
- High symmetrical slew rate and/or high input dynamic range.
- Low inherent distortion.

By adding external feedback paths, gain, frequency response, and stability can be controlled.

Op-amps are normally two-input differential devices; one input inverting the signal and the second input not inverting the signal, and hence called *noninverting*. Several typical op-amp circuits are shown in Fig. 16-32.

Because there are two inputs of opposite polarity, the output voltage is the difference between the inputs where

$$E_{O(+)} = A_V E_2 \qquad (16\text{-}36)$$

$$E_{O(-)} = A_V E_1 \qquad (16\text{-}37)$$

E_O is calculated with the equation

$$E_O = A_V \times (E_1 - E_2) \qquad (16\text{-}38)$$

Often one of the inputs is grounded, either through a direct short or a capacitor. Therefore, the gain is either

$$E_O = A_V E_1 \qquad (16\text{-}39)$$

or

$$E_O = A_V E_2 \qquad (16\text{-}40)$$

To provide both a positive and negative output with respect to ground, a positive and negative power supply is required, as shown in Fig. 16-33. The supply should be regulated and filtered. Often a + and – power supply is not available, such as in an automobile, so the op-amp must operate on a single supply, as shown in Fig. 16-34. In this supply, the output dc voltage is set by adjusting R_1 and R_2 so the voltage at the noninverting input is about one-third the power supply voltage.

The diodes and zener diodes in Fig. 16-35 are used to protect the op-amp from damage caused by transients, reverse voltage, and overdriving. D_6 and D_7 clip the inputs before overdriving, D_1 and D_2 protect against reverse polarity, D_4 and D_5 regulate the supply, and D_3 limits the total voltage across the op-amp.

The dc error factors result in an output offset voltage E_{Oo}, which exists between the output and ground when it should be zero. The dc offset error is most easily corrected by supplying a voltage differential between the inverting and noninverting inputs, which can be accomplished by one of several methods, Fig. 16-36. Connecting the feedback resistor R_f usually causes an offset and can be found with the equation

$$E_{Oo} = I_{bias} R_f \qquad (16\text{-}41)$$

To obtain minimum offset, make the compensating resistor shown in Fig. 16-36A equal to

$$R_{comp} = \frac{R_f R_{in}}{R_f + R_{in}} \qquad (16\text{-}42)$$

If this method is not satisfactory, the methods of Figs. 16-36B or C might be required.

Many op-amps are internally compensated. Often it is advantageous to compensate a device externally to optimize bandwidth and slew rate, lowering distortion. Internally compensated op-amp ICs come in standard packages—the 8 pin TO-99 metal can, the 8 pin dual-in-line package (MINI DIP), and the 14 pin DIP.

Inverting Amplifiers. In the *inverting amplifier* the + input is grounded and the signal is applied to the – input, Fig. 16-37. The output of the circuit is determined by the input resistor R_1 and the feedback resistor R_f.

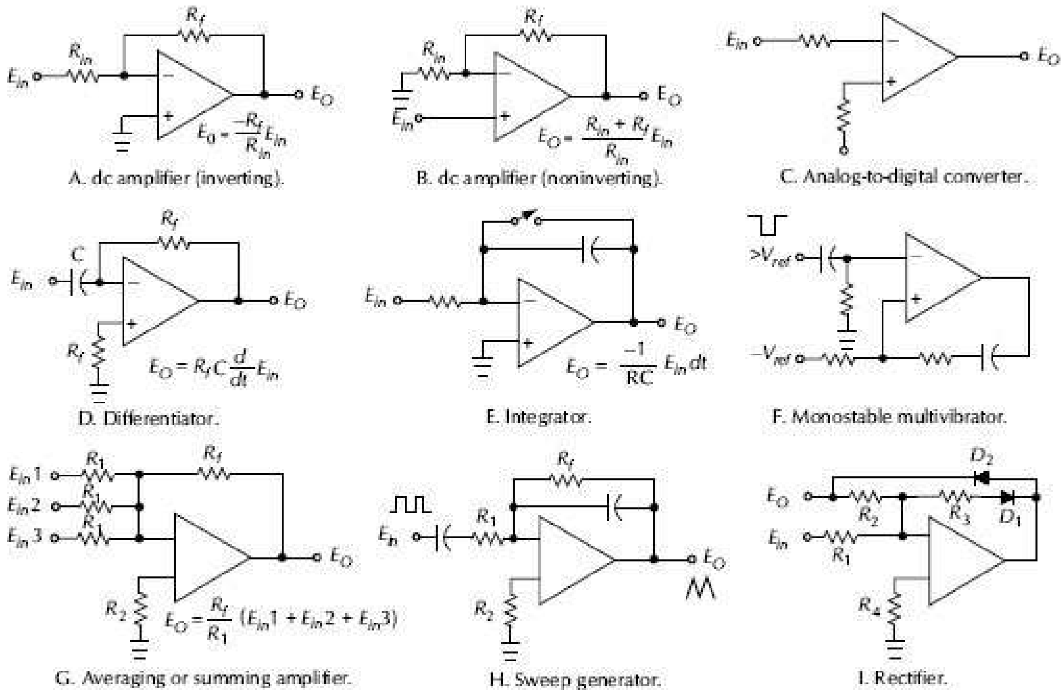


Figure 16-32. Typical op-amp circuits.

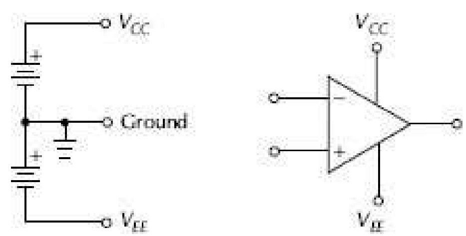


Figure 16-33. Positive- and negative-type power supply.

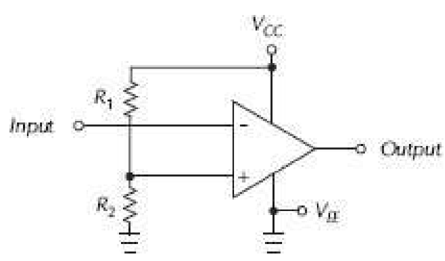


Figure 16-34. Simple circuit for operating on a single-ended power supply.

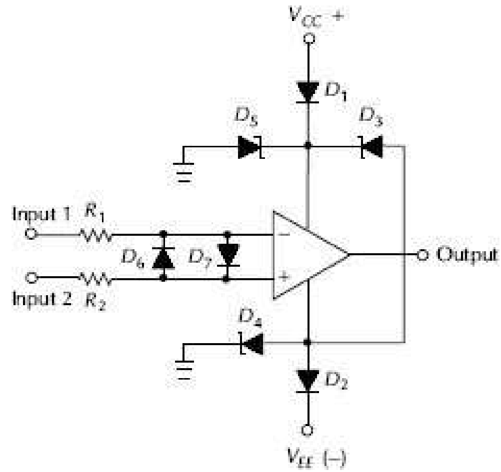


Figure 16-35. Diode protection circuits for op-amps.

$$E_O = E_{in}\left(\frac{-R_f}{R_1}\right) \tag{16-43}$$

where,

E_{in} is the signal input voltage in V,

R_f is the feedback resistor in Ω,

R_1 is the input resistor in Ω.

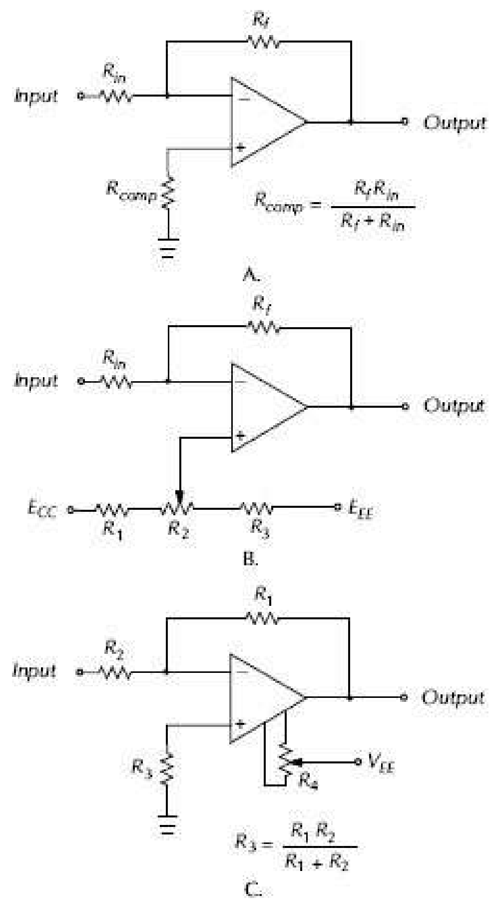


Figure 16-36. Various methods of correcting dc error.

The low frequency roll-off is

$$f_C = \frac{1}{2\pi R_1 C_1} \tag{16-44}$$

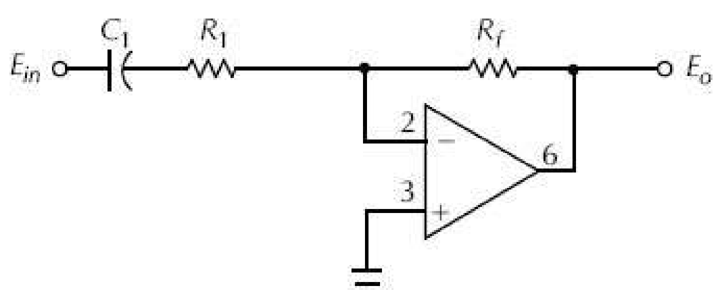


Figure 16-37. A simple inverting amplifier.

Noninverting Amplifier. In the *noninverting amplifier*, Fig. 16-

38, the signal is applied to the + input, while the minus input is part of the feedback loop. The output is

$$E_O = I_{in}\left(\frac{1+R_f}{R_1}\right) \qquad (16\text{-}45)$$

The low-frequency roll-off is in two steps.

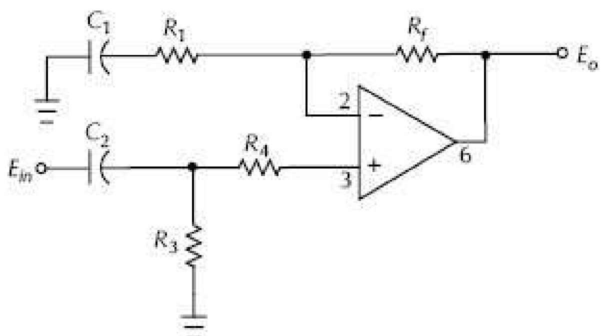


Figure 16-38. A simple noninverting amplifier.

$$f_{C_1} = \frac{1}{2\pi R_1 C_1} \qquad (16\text{-}46)$$

$$f_{C_2} = \frac{1}{2\pi R_3 C_2} \qquad (16\text{-}47)$$

To keep low-frequency noise gain at a minimum, keep $f_{C1} > f_{C2}$.

Power Supply Compensation. The power supply for wideband op-amp circuits should be bypassed with capacitors, Fig. 16-39A, between the plus and minus pin and common. The leads should be as short as possible and as close to the IC as possible. If this is not possible, bypass capacitors should be on each printed circuit board.

Input Capacitance Compensation. Stray input capacitance can lead to oscillation in feedback op-amps because it represents a potential phase shift at the frequency of

$$f = \frac{1}{2\pi R_f C_s} \tag{16-48}$$

where,

R_f is the feedback resistor,

C_s is the stray capacitance.

One way to reduce this problem is to keep the value of R_f low. The most useful way, however, is to add a compensation capacitor, C_f, across R_f as shown in Fig. 16-39B. This makes C_f/R_f and C_s/R_{in} a frequency compensated divider.

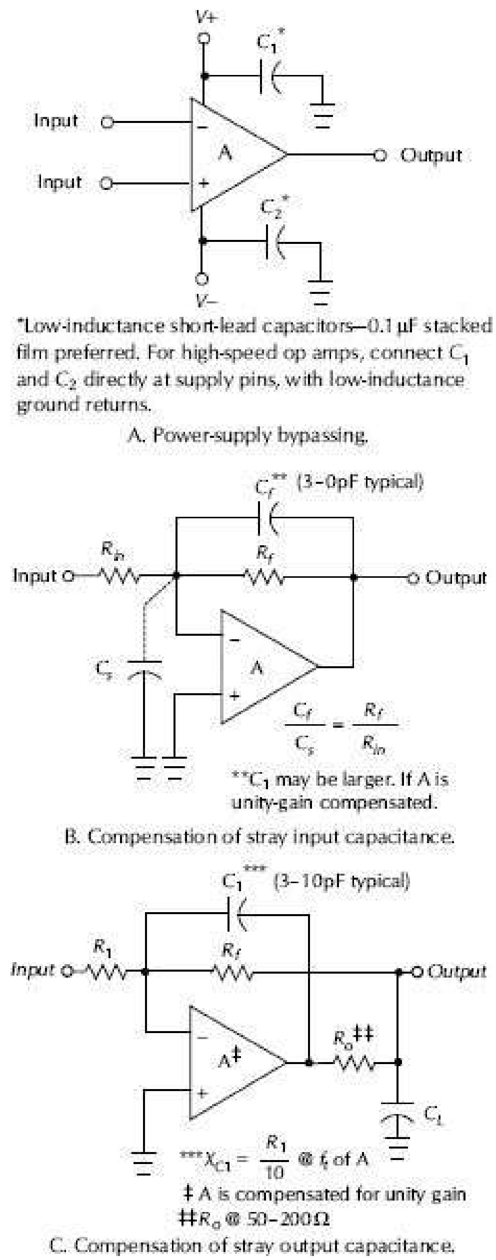


Figure 16-39. Stability enhancement techniques.

Output Capacitance Compensation. Output capacitance greater than 100pF can cause problems, requiring a series resistor R_o being installed between the output of the IC and the load and stray capacitance as shown in Fig. 16-39C. The feedback resistor (R_f) is connected after R_o to compensate for the loss in signal caused by R_o. A compensating capacitor (C_f) bypasses R_f to reduce gain at high frequencies.

Gain and Bandwidth. A perfect op-amp would have infinite gain and infinite bandwidth. In real life however, the dc open loop voltage gain is around 100,000 or 100dB and the bandwidth where gain is 0 is 1MHz, Fig. 16-40.

To determine the gain possible in an op-amp, for a particular bandwidth, determine the bandwidth, follow vertically up to the open loop gain response curve and horizontally to the voltage gain. This, of course, is with no feedback at the upper frequency. For example, for a frequency bandwidth of 0 to 10kHz, the maximum gain of the op-amp in Fig. 16-40 is 100. To have lower distortion, it would be better to have feedback at the required upper frequency limit. To increase this gain beyond 100 would require a better op-amp or two op-amps with lower gain connected in series.

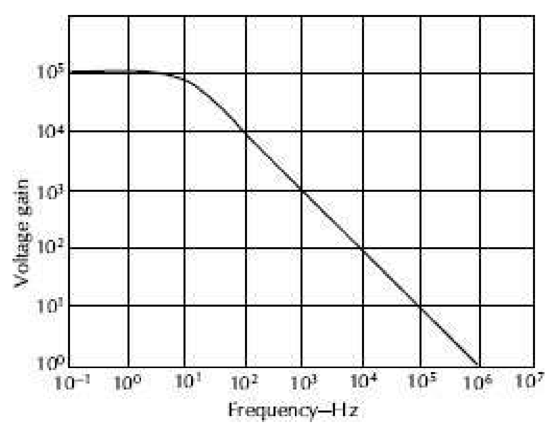


Figure 16-40. Typical open loop gain response.

Differential Amplifiers. Two *differential amplifier* circuits are shown in Fig. 16-41. The ability of the differential amplifier to block identical signals is useful to reduce hum and noise that is picked up on input lines such as in low-level microphone circuits. This rejection is called common-mode rejection and sometimes eliminates the need for an input transformer.

In Fig. 16-41A, capacitors C_1 and C_2 block dc from the previous circuit and provide a 6dB/octave roll-off below

$$f_{C_1} = \frac{1}{2\pi R_1 C_1} \quad (16\text{-}49)$$

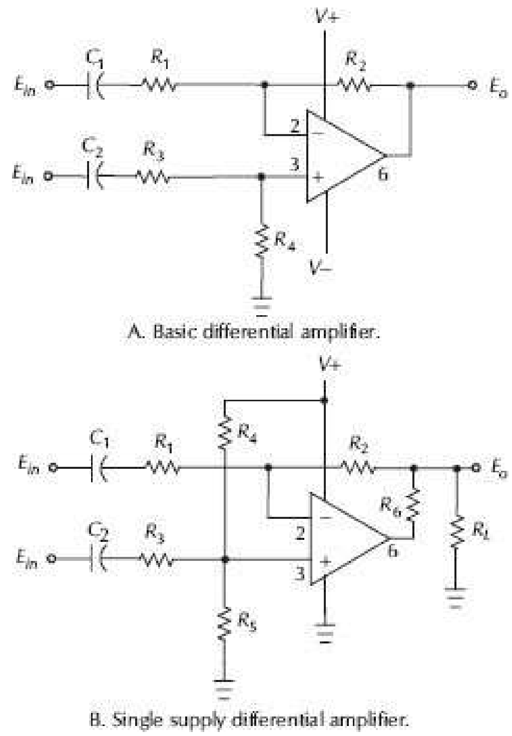


Figure 16-41. Differential amplifiers.

$$f_{C_2} = \frac{1}{2\pi(R_3+R_4)C_2} \quad (16\text{-}50)$$

The output voltage is

$$E_O = (E_{in_2} - E_{in_1})\frac{R_2}{R_1} \quad (16\text{-}51)$$

To reduce the common mode rejection ratio (CMRR),

$$\frac{R_2}{R_1} \equiv \frac{R_4}{R_3} \tag{16-52}$$

and

$$f_{C_1} = f_{C_2} \tag{16-53}$$

Summing Inverter Amplifiers. In the summing inverter, Fig. 16-32G, the virtual ground characteristic of the amplifier's summing point is used to produce a scaling adder. In this circuit, I_{in} is the algebraic sum of the number of inputs, Eq. 16-55.

The output voltage is found with the equation

$$E_O = \left[R_{in_1}\left(\frac{R_f}{R_{in_1}}\right) + R_{in_2}\left(\frac{R_f}{R_{in_2}}\right) + \ldots\ R_{in_n}\left(\frac{R_f}{R_{in_n}}\right) \right] \tag{16-54}$$

$$\begin{aligned} I_{in_1} &= \frac{E_{in_1}}{R_{in_1}} \\ I_{in_2} &= \frac{E_{in_2}}{R_{in_2}} \\ I_{in_n} &= \frac{E_{in_n}}{R_{in_n}} \end{aligned} \tag{16-55}$$

and the total input current is

$$I_{in} = I_{in_1} + I_{in_2} + \ldots I_{in_n} = I_f \tag{16-56}$$

and

$$I_f = \frac{-E_O}{R_f} \tag{16-57}$$

Therefore

$$I_{in_1} + I_{in_2} + \ldots I_{in_n} = \frac{-E_o}{R_f} \tag{16-58}$$

It is interesting that even though the inputs mix at one point, all signals are isolated from each other and one signal does not affect the others and one impedance does not affect the rest.

Operational Transconductance Amplifiers. The *operational transconductance amplifier* (OTA) provides transconductance gain and current output rather than voltage gain and output as in an operational amplifier. The output is the product of the input voltage and amplifier transconductance, and it can be considered an infinite impedance current generator.

Varying the bias current on the OTA can completely control the open-loop gain of the device and can also control the total power input.

OTAs are useful as multipliers, automatic gain control (agc) amplifiers, sample and hold circuits, multiplexers, and multivibrators to name a few.

16.3.4 Dedicated Analog Integrated Circuits for Audio Applications

By Leslie B. Tyler and Wayne Kirkwood, THAT

Corporation

The first ICs used in audio applications were general-purpose op-amps like the famous Fairchild μA741. Early op-amps like the classic 741 generally had drawbacks that limited their use in professional audio, from limited slew rate to poor clipping behavior.

Early on, IC manufacturers recognized that the relatively high-volume consumer audio market would make good use of dedicated ICs tailored to specific applications such as phono preamplifiers and companders. The National LM381 preamplifier and Signetics NE570 compander addressed the needs of consumer equipment makers producing high-volume products such as phono preamplifiers and cordless telephones. Operational Transconductance Amplifiers, such as the RCA CA3080, were introduced around 1970 to primarily serve the industrial market. It was not long before professional audio equipment manufacturers adapted OTAs for professional audio use as early voltage-controlled amplifiers (VCAs). However, through the 1970s all these integrated circuits were intended more for use in consumer and industrial applications than professional audio.

In the mid-1970s, semiconductor manufacturers began to recognize that professional audio had significantly different requirements from the needs of consumer audio or industrial products. The Philips TDA1034 was the first op-amp to combine low noise, 600Ω drive capability and high slew rate—all important characteristics to pro audio designers. Shortly after its introduction, Philips transferred production of the TDA1034 to the newly purchased Signetics division which re-branded it the NE5534. At about the same time, Texas Instruments and National Semiconductor developed general-purpose op-amps using a combination of bipolar and FET technology (the TI TL070- and TL080- series, and the National LF351-series, sometimes called “BIFET”). These parts offered high slew rates, low distortion, and modest noise (though not the 600Ω drive capability of the 5534). While not specifically aimed at pro audio, these characteristics made them attractive to pro audio designers. Along with the NE5534, these op-amps became pro audio industry standards much like the 12AX7 of the vacuum tube era.

In 2006 National Semiconductor, later to be acquired by TI, introduced a line of audio-focused op-amps to challenge the NE5532 and 5534. National recognized the continued need for audio op-amps operating from 30 and 36V supplies. The LME49710 single, 49720 dual and extended voltage LME49860 dual provided a modern replacement for the NE5532 and NE5534-series with greatly improved performance. Interestingly, the LME4562 won Electronic Product’s “Product of the Year award” for 2006.

The advent of the integrated Delta-Sigma analog-to-digital converter in the 1990s drove the development of fully-differential audio op-amps. Delta-Sigma converters require an external modulator capacitor to be driven differentially. To meet this requirement, manufacturers like Analog Devices, Linear Technology, and Texas Instruments have developed op-amps that have differential outputs in addition to the differential inputs of conventional op-amps.[1] (Note, there seems to be some difference in nomenclature for these parts among various manufacturers, including “ADC Drivers” and “Differential ADC Drivers.”) The TI OPA1632, introduced around 2003, met this requirement providing low voltage noise, high output current and the ability to pass a dc common mode bias voltage directly to the converter inputs. An example of a fully differential op-amp driving an A/D converter is shown in Fig. 16-42.

Op-amps are fundamentally general-purpose devices. The desire to control gain via a voltage, and the application of such technology to tape noise reduction, in particular, created a market for ICs that were dedicated to a specific function. This paralleled the way that phono preamplifiers spawned ICs designed for preamplification. In many ways, the VCA drove the development of early pro audio ICs.

The design of audio VCAs benefitted from the early work of Barrie Gilbert, inventor of the "Gilbert Cell" multiplier, who in 1968 published "a precise four-quadrant multiplier with sub-nanosecond response."[2] Gilbert discovered a current mode analog multiplication cell using current mirrors that was linear with respect to both of its inputs. Although its primary appeal at the time was to communications system designers working at RF frequencies, Gilbert laid the groundwork for many audio VCA designs.

In 1972, David E. Blackmer received U.S. Patent 3,681,618 for an "RMS Circuit with Bipolar Logarithmic Converter" and in the following year patent 3,714,462 for a "Multiplier Circuit" useful as an audio voltage-controlled amplifier. Unlike Gilbert, Blackmer used the logarithmic properties of bipolar transistors to perform the analog computation necessary for gain control and rms level detection. Blackmer's development was targeted at professional audio.[3,4] Blackmer's timing could not have been better as the number of recording tracks expanded and, due to reduced track width coupled with the effect of summing many tracks together, tape noise increased. The expanded number of recorded tracks also increased mix complexity. Automation became a desirable feature for recording consoles because there just were not enough hands available to operate the faders.

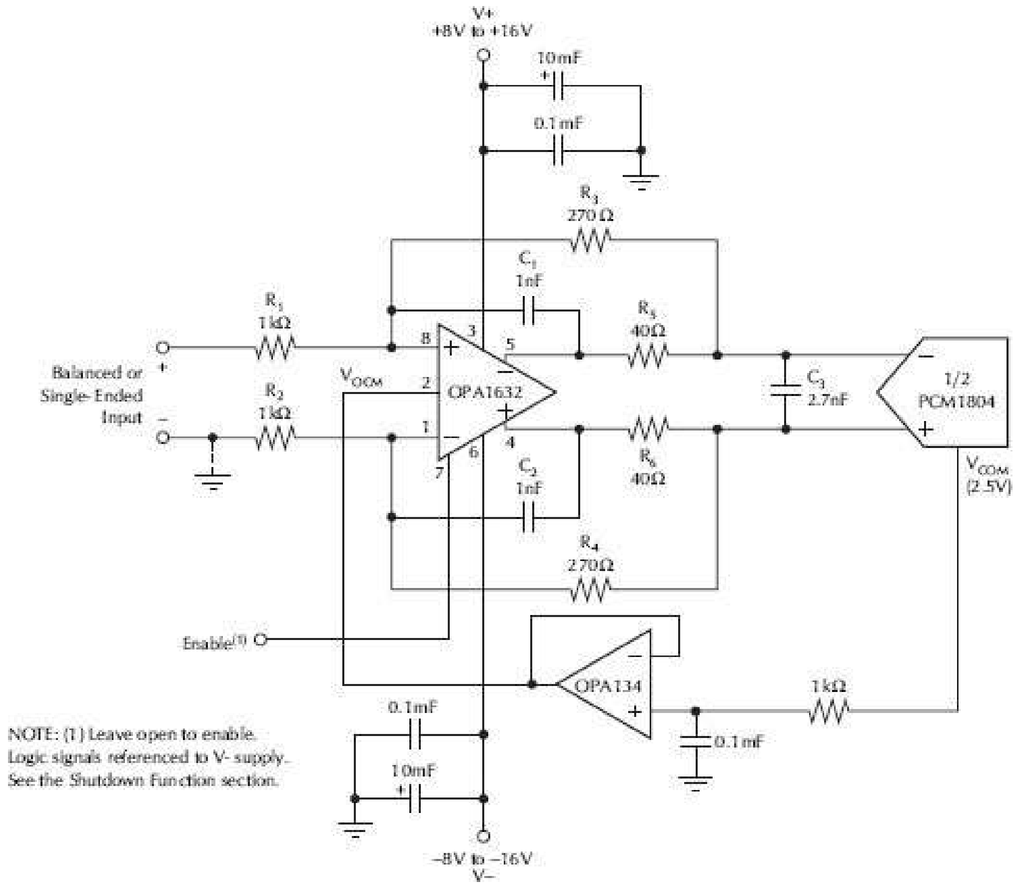


Figure 16-42. A fully differential op-amp driving an A/D converter. Courtesy Texas Instruments Incorporated.

Companies such as dbx Inc. and Dolby Laboratories benefited from this trend with tape noise reduction technologies and, in the case of dbx, VCAs for console automation. Blackmer's discrete transistor-based rms level detectors and VCAs, made by dbx, were soon used in companding multitrack tape noise reduction and console automation systems.

The early Blackmer VCAs used discrete NPN and PNP transistors that required careful selection to match each other. Later Blackmer's design would benefit greatly from integration into monolithic form, though for some time this proved to be very difficult. Nonetheless, Blackmer's discrete audio VCAs and Gilbert's transconductance cell laid the groundwork for dedicated audio ICs. VCAs became a major focus of audio IC development.

Electronic music, not professional recording, primarily drove the early integration of monolithic VCAs and dedicated audio ICs. In 1976, Ron Dow of Solid State Music (SSM) and Dave Rossum of E-mu Systems developed some of the first monolithic ICs for analog synthesizers. SSM's first product was the SSM2000 monolithic VCA.[5] Solid State Music, later to become Solid State Microtechnology, developed an entire line of audio ICs including microphone preamplifiers, VCAs, voltage-controlled filters, oscillators, and level detectors. Later, Douglas Frey developed a VCA topology known as the o*perational voltage-controlled element* (OVCE) that was first used in the SSM2014.[6] Doug Curtis, of Interdesign and later founder of Curtis Electro Music (CEM), also developed a line of monolithic ICs for the synthesizer market that proved to be very popular with manufacturers such as Oberheim, Moog, and ARP.[7] VCAs produced for electronic music relied on NPN transistor gain cells to simplify integration.

In the professional audio market, Paul Buff of Valley People, David Baskind and Harvey Rubens of VCA Associates, and others in addition to Blackmer also advanced discrete VCA technology. Baskind and Rubens eventually produced a VCA IC that ultimately became the Aphex/VCA Associates "1537."[8]

Blackmer's VCAs and rms detectors used the precise logarithmic characteristics of bipolar transistors to perform mathematical operations suitable for VCAs and rms detection. The SSM, CEM, and Aphex products used variations on the linear multiplier, where a differential pair, or differential quad, is varied to perform VCA functions. Close transistor matching and control of temperature-related errors are required for low distortion and control feed-through in all VCA topologies.

The Gilbert multiplier, the CA3080-series of OTAs, and the VCAs produced by SSM, CEM, and Aphex all relied solely on NPN transistors as the gain cell elements. This greatly simplified the integration of the circuits. Blackmer's log-antilog VCAs required, by contrast, precisely matched NPN and PNP transistors. This made Blackmer's VCAs the most difficult to integrate. dbx finally introduced its 2150-series monolithic VCAs in the early 1980s, almost six years after the introduction of the SSM2000.[9]

Many of the earlier developers of VCAs changed ownership or left the market as analog synthesis faded. Analog Devices currently produces many of the SSM products after numerous ownership changes. THAT Corporation assumed the patent portfolio of dbx Inc. Today Analog Devices, Texas Instruments, and THAT Corporation are the primary manufacturers making analog ICs specifically for the professional audio market.

16.3.4.1 Voltage-Controlled Amplifiers

Modern IC VCAs take advantage of the inherent and precise matching of monolithic transistors that, when combined with on-chip trimming, lowers distortion to very low levels. Two types of IC audio VCAs are commonly used and manufactured today: those based on Douglas Frey's Operational Voltage Controlled Element (OVCE)[10] and those based on David Blackmer's bipolar log-antilog topology.[3]

The Analog Devices SSM2018. The Frey OVCE gain cell was first introduced in the SSM2014 manufactured by Solid State Microtechnology (SSM).[11] SSM was acquired by Precision Monolithics, Inc., which was itself acquired by Analog Devices, who currently offers a Frey OVCE gain cell branded the SSM2018T. Frey's original patents, U.S. 4,471,320 and U.S. 4,560,947, built upon the work of David Baskind and Harvey Rubens (see U.S. Patent 4,155,047) by adding corrective feedback around the gain cell core.[12,13,14] Fig. 16-43 shows a block diagram of the SSM2018T VCA.

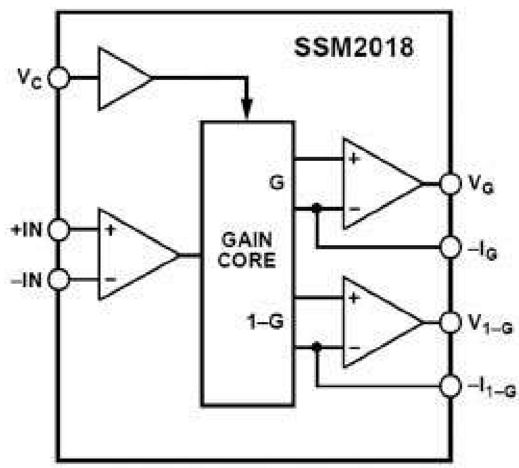


Figure 16-43. A block diagram of the SSM2018T VCA. Courtesy Analog Devices, Inc.

The OVCE is unique in that it has two outputs: V_G and $V_{1\text{-}G}$. As the V_G output increases gain in response to control voltage, the $V_{1\text{-}G}$ output attenuates. The result is that the audio signal pans from one output to the other as the control voltage is changed.

The following expressions show how this circuit works mathematically:

$$V_{out1} = V_G = 2K \times V_{in} \quad (16\text{-}59)$$

and

$$V_{out2} = V_{1-G} = 2(1-K) \times V_{in} \quad (16\text{-}60)$$

where,

K varies between 0 and 1 as the control voltage is changed from full attenuation to full gain.

When the control voltage is 0V, $K = 0.5$ and both output voltages equal the input voltage. The value K is exponentially proportional to the applied control voltage; in the SSM2018T, the gain control constant in the basic VCA configuration is – 30mV/dB, so the decibel gain is directly proportional to the applied control voltage. This makes the part especially applicable to audio applications.

The SSM2018 has many applications as a VCA, but its use as a voltage-controlled panner (VCP) is perhaps one of the most unique, Fig. 16-44.

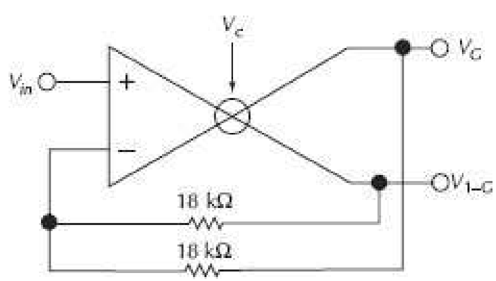


Figure 16-44. SSM2018 as a VCP. Courtesy Analog Devices, Inc.

THAT Corporation's 2180 and 2181 VCAs. The Blackmer VCAs now offered by THAT Corporation (which registered the trademark "Blackmer" for this application) exploit the mathematical property that adding a constant to the logarithm of a number is equivalent, in the linear domain, to multiplying the number by the antilog of the constant.

The equation for determining the output is

$$\begin{aligned} I_{out} &= \text{antilog}[(\log I_{in}) + E_C] \\ &= I_{in} \times [\text{antilog} E_C] \end{aligned} \tag{16-61}$$

I_{in} is multiplied by the antilog of E_C to produce I_{out}. Conveniently, and fortunately for Blackmer, the exponential response of E_C is linear in dB.

Consider the unity-gain case when $E_C = 0$.

$$\begin{aligned} I_{out} &= \text{antilog}[(\log I_{in}) + 0] \\ &= I_{in} \times [\text{antilog} 0] \\ &= I_{in} \times 1 \\ I_{out} &= I_{in} \end{aligned}$$

Blackmer VCAs exploit the logarithmic properties of a bipolar junction transistor (BJT). In the basic Blackmer circuit, the input signal I_{in} (the Blackmer VCA works in the current, not the voltage domain) is first converted to its log-domain equivalent. A control voltage, E_C, is added to the log of the input signal. Finally, the antilog is taken of the sum to provide an output signal I_{out}. This multiplies I_{in} by a control constant, E_C. When needed, the input signal voltage is converted to a current via an input resistor, and the output signal current is converted back to a voltage via an op-amp and feedback resistor.

Like the Frey OVCE, the Blackmer VCA's control voltage (E_C) is exponentiated in the process. This makes the control law exponential, or linear in dB. Many of the early embodiments of

VCAs for electronic music were based on linear multiplication and required exponential converters, either external or internal to the VCA, to obtain this desirable characteristic.[15] Fig. 16-45 shows the relationship between gain and E_C for a Blackmer VCA.

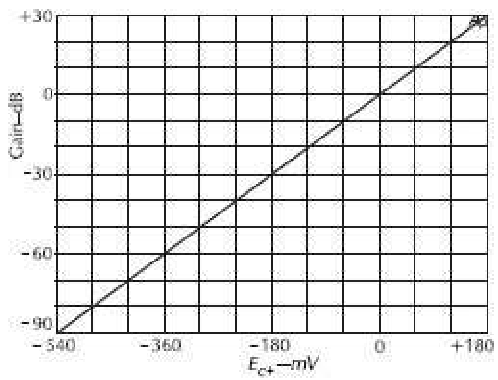


Figure 16-45. THAT 2180 gain versus E_C+. Courtesy THAT Corporation.

Audio signals are of both polarities; that is, the sign of I_{in} in the above equations will be either positive or negative at different times. Mathematically, the log of a negative number is undefined, so the circuit must be designed to handle both polarities. The essence of David Blackmer's invention was to handle each phase—positive and negative—of the signal waveform with different "genders" of transistors—NPN and PNP—and to provide a class A-B bias scheme to deal with the crossover region between the two. This made it possible to generate a sort of bipolar log and antilog. A block diagram of a Blackmer VCA is shown in Fig. 16-46.

Briefly, the circuit functions as follows. An ac input signal current I_{IN} flows in pin 1, the input pin. An internal operational transconductance amplifier (OTA) maintains pin 1 at virtual ground potential by driving the emitters of Q_1 and (through the Voltage Bias Generator) Q_3. Q_3/D_3 and Q_1/D_1 act to log the input current, producing a voltage (V_3) that represents the bipolar logarithm of the input current. (The voltage at the junction of D_3 and D_4 is the same as V_3, but shifted down by four forward V_{be} drops.)

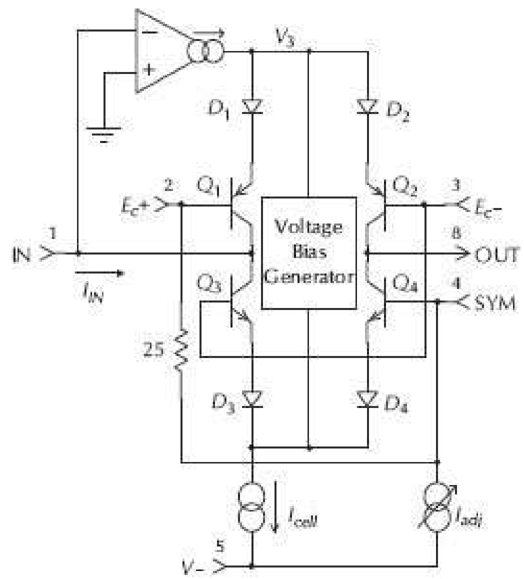


Figure 16-46. THAT 2180 equivalent schematic. Courtesy THAT Corporation.

Pin 8, the output, is usually connected to a virtual ground. As a result, Q_2/D_2 and Q_4/D_4 take the bipolar antilog of V_3, creating an output current flowing to the virtual ground, which is a precise replica of the input current. If pin 2 (E_C+) and pin 3 (E_C-) are held at ground, the output current will equal the input current. For pin 2 positive or pin 3 negative, the output current will be scaled larger than the input current. For pin 2 negative or pin 3 positive, the output current is scaled smaller than the input.

The log portion of the VCA, D_1/Q_1 and D_3/Q_3, and the antilog stages, D_2/Q_2 and D_4/Q_4 in Fig. 16-45, require both the NPN and the PNP transistors to be closely matched to maintain low distortion. As well, all the devices (including the bias network) must be at the same temperature. Integration solves the matching and temperature problems, but conventional "junction-isolated" integration is notorious for offering poor-performing PNP devices. Frey and others avoided this problem by basing their designs exclusively on NPN devices for the critical multiplier stage. Blackmer's design required "good" PNPs as well as NPNs.

One way to obtain precisely matched PNP transistors that provide discrete transistor performance is to use an IC fabrication technology known as *dielectric isolation*. THAT Corporation uses dielectric isolation to fabricate integrated PNP transistors that equal or exceed the performance of NPNs. With dielectric isolation, the bottom layers of the devices are available early in the process, so both N- and P-type collectors are possible. Furthermore, each transistor is electrically insulated from the substrate and all other devices by an oxide layer, which enables discrete transistor performance with the matching and temperature characteristics only available in monolithic form.

In Fig. 16-46, it can also be seen that the Blackmer VCA has two E_C inputs having opposite control response—E_C+ and E_C–. This unique characteristic allows both control inputs to be used simultaneously. Individually, gain is exponentially proportional to the voltage at pin 2, and exponentially proportional to the negative of the voltage at pin 3. When both are used simultaneously, gain is exponentially proportional to the difference in voltage between pins 2 and 3. Overall, because of the exponential characteristic, the control voltage sets gain *linearly* in *decibels* at 6mV/dB.

Fig. 16-47 shows a typical VCA application based on a THAT 2180 IC. The audio input to the VCA is a current; an input resistor converts the input voltage to a current. The VCA output is also a current. An op-amp and its feedback resistor serve to convert the

VCA’s current output back to a voltage.

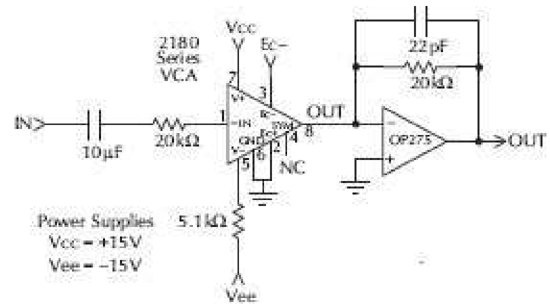


Figure 16-47. Basic THAT 2180 VCA application. Courtesy THAT Corporation.

As with the basic topologies from Gilbert, Dow, Frey, Curtis, and other transconductance cells, the current input/output Blackmer VCA can be used as a variable conductance to tune oscillators, filters, and the like. An example of a VCA being used to control a first-order state-variable filter is shown in Fig. 16-48 with the response plot in Fig. 16-49. When combined with audio level detectors, VCAs can be used to form a wide range of dynamics processors, including compressors, limiters, gates, duckers, companding noise reduction systems, and signal-controlled filters.

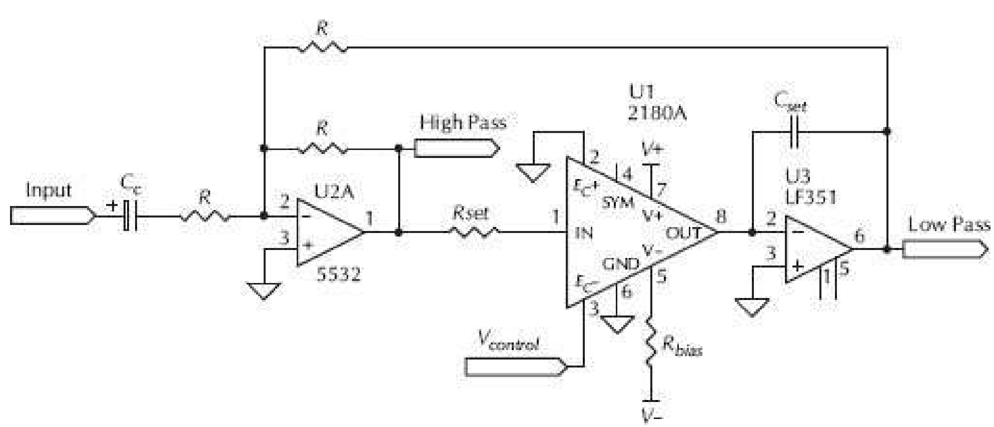


Figure 16-48. VCA state-variable filter. Courtesy THAT Corporation.

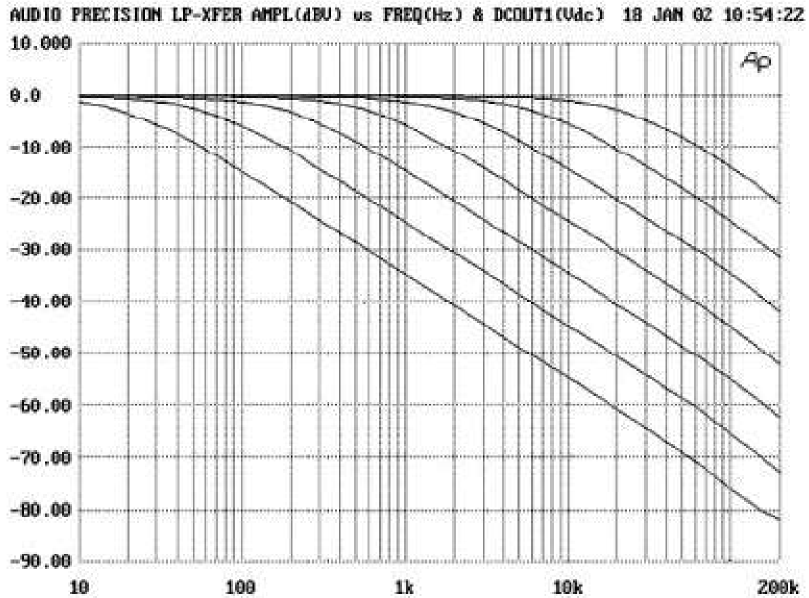


Figure 16-49. State-variable filter response. Courtesy THAT Corporation.

16.3.4.2 Peak, Average, and RMS Level Detection

It is often desirable to measure audio level for display, dynamics control, noise reduction, instrumentation, etc. Level detectors take different forms: among the most common are those that represent peak level, some form of average level over time, and *root-mean-square* (more simply known as *rms* level).

Peak signal level is usually interpreted to mean the highest instantaneous level within the audio bandwidth. Measuring peak level involves a detector with very fast charge (attack) response and much slower decay. Peak levels are often used for headroom and overload indication and in audio limiters to prevent even brief overload of transmission or storage media. However, peak measurements do not correlate well with perceived loudness, since the ear responds not only to the amplitude, but also to the duration of a sound.

Average-responding level detectors generally average out (or smooth) full or half-wave rectified signals to provide envelope information. While a pure average response (that of an R-C circuit) has equal rise (attack) and fall (decay) times, in audio applications, level detectors often have faster attack than decay. The familiar VU meter is average responding, with a response time and return time of the indicator both equal to 300ms. The peak program meter (PPM), commonly used in Europe for audio program level measurement, combines a specific quick attack response with an equally specific, slow fall time. PPM metering provides a reliable indication of meaningful peak levels.[16]

Rms level detection is unique in that it provides an ac measurement suitable for the calculation of signal power. Rms measurements of voltage, current, or both indicate effective power. *Effective power* is the heating power of a dc signal equivalent to that offered by an ac signal. True rms measurements are not affected by the signal waveform complexity, while peak and average readings vary greatly depending on the nature of the waveform. For example, a resistor heated by a 12Vac rms signal produces the same number of watts—and heat—as a resistor connected to 12Vdc. This is true regardless of whether the ac waveform is a pure sinusoid, a square wave, a triangle wave or music. In instrumentation, rms is often referred to as *true rms* to distinguish it from average-responding instruments that are calibrated to read *rms* only for sinusoidal inputs. Importantly, in audio signal-processing applications, rms response is thought to closely approximate the human perception of loudness.[2]

16.3.4.3 Peak and Average Detection with Integrated Circuits

The fast response of a peak detector is often desirable for overload indication or dynamics control when a signal needs to be limited to fit the strict level confines of a transmission or storage medium. A number of opamp-based circuits detect peak levels using full or half-wave rectification. General-purpose op-amps are quite useful for constructing peak detectors and are discussed in section 16.3.3. The now discontinued Analog Devices PKD01 was perhaps the only peak detector IC suited for audio applications.

Average-responding level detection is performed by rectification followed by a smoothing resistor/capacitor (R-C) filter whose time constants are chosen for the application. If the input is averaged over a sufficiently long period, the signal envelope is detected. Again, general-purpose op-amps serve quite well as rectifiers with R-C networks or integrators serving as averaging filters.

Other than meters, most simple electronic audio level detectors use an asymmetrical averaging response that attacks more quickly than it decays. Such circuits usually use diode steering to charge a capacitor quickly through a relatively small-value resistor, but discharge it through a larger resistor. The small resistor yields a fast attack, and the large resistor yields a slower decay.

16.3.4.4 Rms Level Detection Basics

Rms detection has many applications in acoustic and industrial instrumentation. As mentioned previously, rms level detectors are thought to respond similarly to the human perception of loudness. This makes them particularly useful for audio dynamics control.

Rms is mathematically defined as the square root of the mean of the square of a waveform. Electronically, the mean is equal to the average, which can be approximated by an R-C network or an op-amp-based integrator. However, calculating the square and square root of waveforms is more difficult.

Designers have come up with a number of clever techniques to avoid the complexity of numerical rms calculation. For example, the heat generated by a resistive element may be used to measure power. Power is directly proportional to the square of the voltage across, or current through, a resistor, so the heat given off is proportional to the square of the applied signal level. To measure large amounts of power having very complex waveforms, such as the RF output of a television transmitter, a resistor dummy load is used to heat water. The temperature rise is proportional to the transmitter power. Such caloric instruments are naturally slow to respond, and impractical for the measurement of sound. Nonetheless, solid-state versions of this concept have been integrated, as, for example U.S. Patent 4,346,291, invented by Roy Chapel and Macit Gurol.[17] This patent, assigned to instrumentation manufacturer Fluke, describes the use of a differential amplifier to match the power dissipated in a resistive element, thus measuring the true rms component of current or voltage applied to the element. While very useful in instrumentation, this technique has not made it into audio products due to the relatively slow time constants of the heating element.

To provide faster time constants to measure small rms voltages or currents with complex waveforms such as sound, various analog computational methods have been employed. Computing the square of a signal generally requires extreme dynamic range, which limits the usefulness of direct analog methods in computing rms value. As well, the square and square-root operations require complex analog multipliers, which have traditionally been expensive to fabricate.

As with VCAs, the analog computation required for rms level detection is simplified by taking advantage of the logarithmic properties of bipolar junction transistors. The seminal work on computing rms values for audio applications was developed by David E. Blackmer, who received U.S. Patent 3,681,618 for an “RMS Circuit with Bipolar Logarithmic Converter.”[3] Blackmer’s circuit, discussed later, took advantage of two important log-domain properties to compute the square and square root. In the log domain, a number is squared by multiplying it by 2; the square root is obtained by dividing it by 2.

For example, to square the signal V_{in} use

$$V_{in}^{2} = \text{antilog}[(\log V_{in}) \times 2] \qquad (16\text{-}62)$$

To take the square root of V log,

$$\sqrt{V \log} = \text{antilog}\left[\frac{\log(V \log)}{2}\right] \qquad (16\text{-}63)$$

16.3.4.5 Rms Level Detection ICs

Because rms level detectors are more complex than either peak- or average-responding circuits, they benefit greatly from integration. Fortunately, a few ICs are suitable for the professional audio applications. In 2014, after over 30 years of production, THAT Corporation announced the end of life for its standalone 2252 rms level detector, pointing to the rms detectors in its Analog Engine® series of parts as a follow-on migration path. Solid State Music offered the SSM2110 rms-to-dc converter, but it was discontinued after Analog Devices acquired Precision Monolithics Inc., which had acquired Solid State Music. Analog Devices still offers its AD636.[18]

Analog Devices AD636. The AD636 has enjoyed wide application in audio and instrumentation. Its predecessor, the AD536, was used in the channel dynamics processor of the SSL 4000 series console in conjunction with a dbx VCA.

The AD636 shown in Fig. 16-50 provides both a linear-domain rms output and a dB-scaled logarithmic output. The linear output at pin 8 is ideal for applications where the rms input voltage must be read with a dc meter. Suitably scaled, 1Vrms input can produce 1Vdc at the buffer output, pin 6.

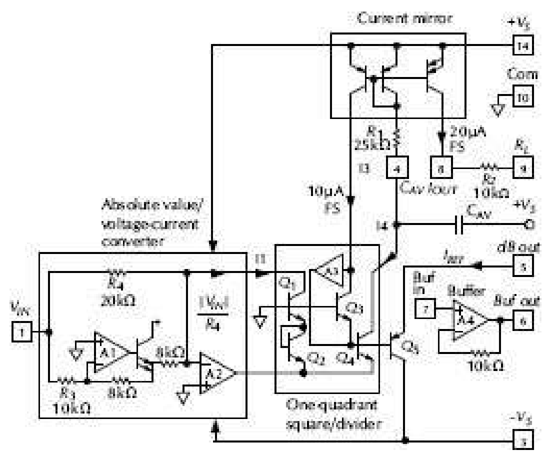


Figure 16-50. The AD636 block diagram. Courtesy Analog Devices, Inc.

In audio applications such as signal processors, it is often most useful to express the signal level in dB. The AD636 also provides a dB-scaled current output at pin 5. The linear dB output is particularly useful with exponentially controlled VCAs such as the SSM2018 or THAT 2180 series.

Averaging required to calculate the mean of the sum of the squares is performed by a capacitor, C_{AV}, connected to pin 4. Fig.

16-51 shows an AD636 used as an audio dB meter for measurement applications.

THAT 2252. The recently discontinued 2252 IC uses the technique taught by David Blackmer to provide wide dynamic range, logarithmic linear dB output, and relatively fast time constants. Blackmer's detector delivers a fast attack with a slow linear dB decay characteristic in the log domain.[3] Because it was specifically developed for audio applications, it has become a standard for use in companding noise reduction systems and VCA-based compressor/limiters. This design now lives on in the form of THAT's 43xx series of Analog Engine® ICs (see section 16.3.4.6), which all contain at least one VCA and one rms detector.

A simplified schematic of Blackmer's rms detector, used in the THAT 2252, is shown in Fig. 16-52.

The audio input is first converted to a current I_{in} by an external resistor (not shown in Fig. 16-52). I_{in} is full-wave rectified by a current mirror rectifier formed by OA1 and Q_1-Q_3, such that IC_4 is a full-wave rectified version of I_{in}. Positive input currents are forced to flow through Q_1, and mirrored to Q_2 as IC_2; negative input currents flow through Q_3 as IC_3; both IC_2 and IC_3 thus flow through Q_4. (Note that pin 4 is normally connected to ground through an external 20Ω resistor.)

Performing the absolute value before logarithmic conversion avoids the problem that, mathematically, the log of a negative number is undefined. This eliminates the requirement for bipolar logarithmic conversion and the PNP transistors required for log-domain VCAs.

OA2, together with Q_4 and Q_5, forms a log amplifier. Due to the two diode-connected transistors in the feedback loop of OA2, the voltage at its output is proportional to twice the log of IC_4. This voltage, V_{log}, is therefore proportional to the log of I_{in^2} (plus the bias voltage V_2).

To average $V_{\log}$, pin 6 is usually connected to a capacitor C_T and a negative current source R_T, see Fig. 16-53. The current source establishes a quiescent dc bias current, I_T, through diode-connected Q_6. Over time, C_T charges to $1V_{be}$ below $V_{\log}$.

Q_6's emitter current is proportional to the antilog of its V_{be}. The potential at the base (and collector) of Q_6 represents the log of I_{in^2} while the emitter of Q_6 is held at ac ground via the capacitor. Thus, the current in Q_6 is proportional to the square of the instantaneous change in input current. This dynamic antilogging causes the capacitor voltage to represent the log of the mean of the square of the input current. Another way to characterize the operation of Q_6, C_T, and R_T is that of a "log domain" filter.[18]

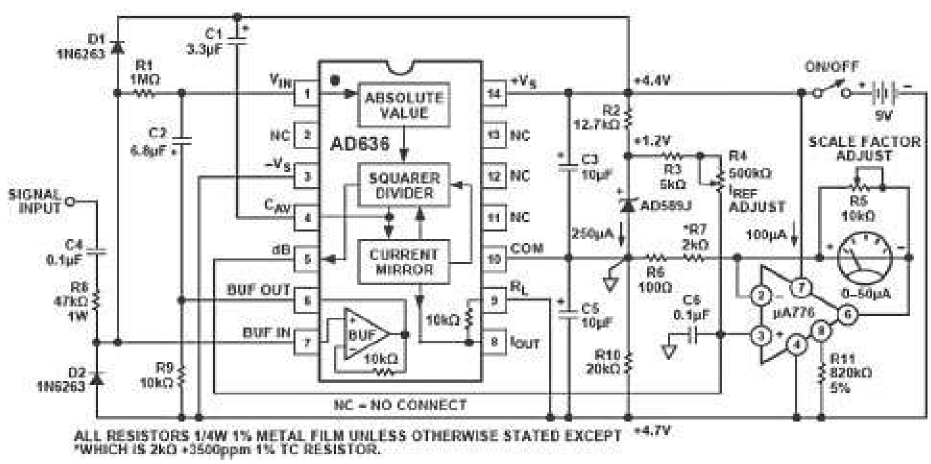


Figure 16-51. AD636 as an audio dB meter. Courtesy Analog Devices, Inc.

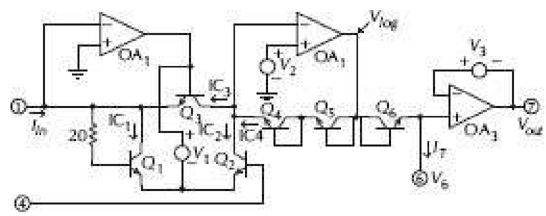


Figure 16-52. Block diagram of a THAT 2252 IC. Courtesy THAT Corporation.

In the THAT 2252, the square root portion of the rms calculation is not computed explicitly but is implied by the constant of proportionality for the output. Since, in the log domain, taking the square root is equivalent to dividing by two, the voltage at the output (pin 7) is proportional to the mean of the square at approximately 3mV/dB and proportional to the square root of the mean of the square at approximately 6mV/dB.

The attack and release times of rms detectors are locked in a relationship to each other and separate controls for each are not possible while still maintaining rms response. Varying the value of C_T and R_T in the THAT 2252, and C_{AV} in the AD636 allow the time constant to be varied to suit the application. More complex approaches, such as a nonlinear capacitor, are possible with additional circuitry.[19]

Fig. 16-53 shows a typical application for the THAT 2252. The input voltage is converted to a current by R_{in}. C_{in} blocks input dc and internal op-amp bias currents. The network around pin 4 sets the waveform symmetry for positive versus negative input currents. Internal bias for the THAT 2252 is set by R_b and bypassed by a 1μF capacitor. R_T and C_T set the timing of the log-domain filter. The output signal (pin 7) is 0V when the input signal current equals a reference current determined by I_{bias} and I_T. It varies in dc level above and below this value to represent the dB input level at the rate of ~6mV/dB.

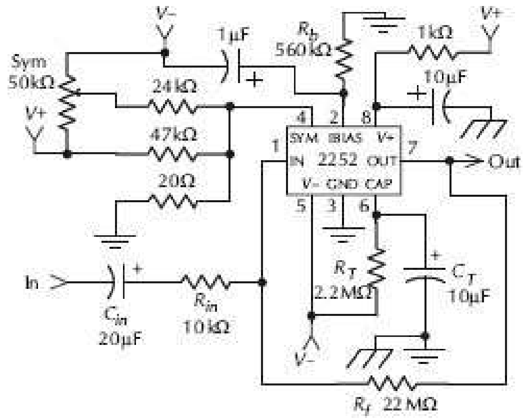


Figure 16-53. Typical application of a THAT 2252 IC. Courtesy THAT Corporation.

Fig. 16-54 shows the tone burst response of the THAT rms detectors, while Fig. 16-55 is a plot of their output level versus input level. THAT's detectors have linear dB response over up to a 100dB range.

The Analog Devices AD636 and the detectors in THAT Corporation Analog Engines provide precise, low-cost rms detection due to their integration into monolithic form. On their own, rms detectors are very useful at monitoring signal level, controlling instrumentation, and other applications. When combined with VCAs for gain control, many different signal processing functions can be realized including noise reduction, compression, and limiting.

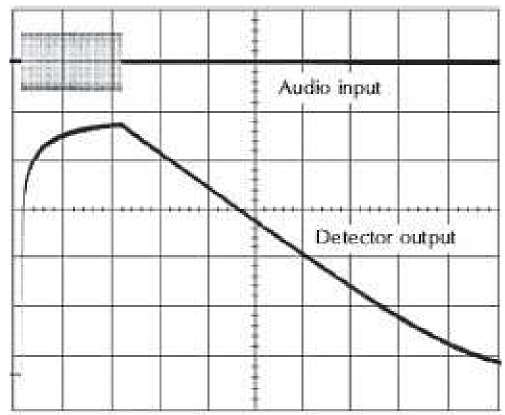


Figure 16-54. THAT 2252 tone burst response. Courtesy THAT Corporation

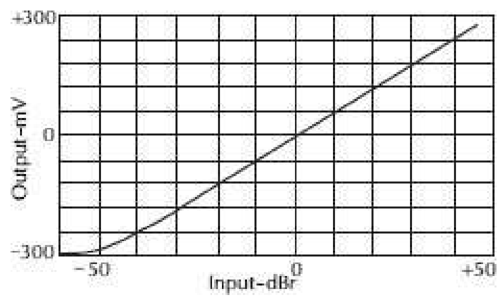


Figure 16-55. THAT 2252 input versus output. Courtesy THAT Corporation.

16.3.4.6 Monolithic VCA + Detector ICs

In 1976, Craig Todd (then of Signetics) developed the NE750 compander IC. Originally intended as the compander for cordless phones, this part contained two audio channels, each with a VCA and a level detector. Each channel could be configured as either a compressor or expander. Typically, a compressor was used in the handset to compress the microphone signal before broadcasting it to the base unit, where it would be expanded by a reciprocal expander. The other halves were used to compress the incoming audio in the base unit before transmitting it to the handset, where it was expanded by the other half of the NE570. Signetics broadened this concept to a family of parts, including the 571, 572, and 575, with various different features. Eventually, many manufacturers, including Toko and Mitsumi would introduce parts aimed at the cordless telephone market; the original Signetics' designs were eventually acquired by Philips, and most recently On Semiconductor.

A VCA plus a level detector was useful in many ways, from the companding application originally envisioned by Signetics and Todd to a variety of analog signal conditioning, particularly dynamics processors. By the 1990s, Solid State Music had produced the SSM2120 which included two VCAs and two log-domain averaging level detectors. This part differed from the companders aimed at cordless phones in the critical area that the detector output was not internally tied to the VCA control input. This allowed designers more flexibility in compression and expansion ratios, facilitating complex side chain designs to accommodate different ratios at different levels and other "tricks" of the dynamics processor art.

In the early 1990s, THAT Corporation introduced THAT 4301, to which the company applied the trademarked name "Analog Engine". Intended for single-channel operation, the 4301 contains an rms-level detector (derived from THAT's 2252), a "Blackmer" VCA with dedicated current-to-voltage converter op-amp, and two additional op-amps. Like THAT's 2181 series, the VCA in the 4301 uses an external trim to adjust distortion and dc feedthrough.

THAT has since introduced several more Analog Engines, including several specifically intended for low-voltage and low-power operation, including the 4315, 4316, and 4320, Fig. 16-56. The lower-power versions are often used to form the compander in wireless microphones and belt packs for pro-audio applications. The SSM2120 was obsoleted by Analog Devices in 1999, and THAT's 4311 (a low power Analog Engine) is no longer recommended for new designs. But, the rest of the THAT Analog Engines remain in active production.

16.3.5 Integrated Circuit Preamplifiers

The primary applications of preamplifiers for professional audio in the post-tape era are for use with microphones. Before the development of monolithic ICs dedicated to the purpose, vacuum tubes, discrete bipolar or field-effect transistors,[21] or general-purpose audio op-amps were used as preamplifiers.[22] Dynamic microphones generally produce very small signal levels and have low output impedance. Ribbon microphones are notorious for low output levels. For many audio applications, significant gain (40 to 60dB) is required to bring these microphone level signals up to pro audio levels. Condenser microphones, powered by phantom power, external power supplies, or batteries, often produce higher signal levels requiring less gain.

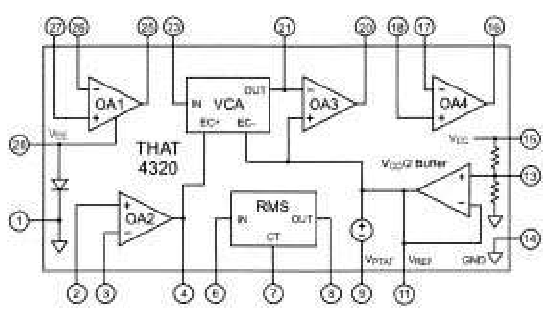


Figure 16-56. THAT 4320 Analog Engine. Courtesy THAT Corporation.

To avoid adding significant noise to the microphone's output, professional audio preamplifiers must have very low input noise. Transformer-coupled preamps ease the requirement for very low noise amplification, since they take advantage of the voltage step-up possible within the input transformer. Early transformerless, or active, designs required performance that eluded integration until the early 1980s. Until semiconductor process and design improvements permitted it and the market developed to generate sufficient demand, most microphone preamplifiers were based on discrete transistors or discrete transistors augmented with commercially available op-amps.

Virtually all professional microphones use two signal lines to produce a balanced output. This allows a preamplifier to distinguish the desired differential audio signal—which appears as a voltage *difference* between the two signal lines—from hum and noise pickup—which appears as a "common-mode" signal with the *same* amplitude and polarity on both signal lines. Common mode rejection quantifies the ability of the preamplifier to reject common mode interference while accepting differential signals.

Therefore, one goal of a pro-audio mic preamp is to amplify differential signals in the presence of common-mode hum. As well, the preamp should ideally add no more noise than the thermal noise of the source impedance—well below the self-noise of the microphone and ambient acoustic noise.

Phantom power is required for many microphones, especially professional condenser types. This is usually a +48Vdc power supply applied to both polarities of the differential input through 6.8kΩ resistors (one for each input polarity). Dc supply current from the microphone returns through the ground conductor. Phantom power appears in common mode essentially equal on both inputs. The voltage is used to provide power to the circuitry inside the microphone.

16.3.5.1 Transformer Input Microphone Preamplifiers

Many microphone preamplifiers use transformers at their inputs. Transformers, although costly, provide voltage gain that can ease the requirements for low noise in the subsequent amplifier. The transformer's voltage gain is determined by the turns ratio of the secondary versus the primary. This ratio also transforms impedance, making it possible to match a low-impedance microphone to a high-impedance amplifier without compromising noise performance.

A transformer's voltage gain is related to its impedance ratio by the following equation

$$Gain = 20\log\sqrt{\frac{Z_s}{Z_p}} \qquad (16\text{-}64)$$

where,

Gain is the voltage gain in dB of the transformer,

Z_p is the impedance in Ω that is connected to the primary of the transformer (i.e., the microphone source impedance),

Z_s is the impedance in Ω that the secondary presents (as a source) to its load.

A properly designed transformer with a 10:1 turns ratio produces 20dB of free voltage gain without adding noise save a small contribution from resistance in the windings. If its primary is connected to a 150Ω source, it will present a 15kΩ source at the output of its secondary. Well-made transformers also provide high common-mode rejection, which helps avoid hum and noise pickup. This is especially important with the low output voltages and long cable runs common with professional microphones. In addition, transformers provide galvanic isolation by electrically insulating the primary circuit from the secondary while allowing signal to pass. While usually unnecessary in microphone applications, this provides a true ground lift, which can eliminate ground loops in certain difficult circumstances.

Transformer isolation is also useful when feeding phantom power (a +48Vdc current-limited voltage to power internal circuitry in the microphone) down the mic cable from the preamp input terminals. Phantom power may be connected through a center tap on the primary to energize the entire primary to +48Vdc, or supplied through resistors (usually 6.8kΩ) to each end of the primary of the transformer. (The latter connection avoids dc currents in the coils, which can lead to premature saturation of the core magnetics.) The galvanic isolation of the transformer avoids any possibility of the 48Vdc signal from reaching the secondary windings.

16.3.5.2 Active Microphone Preamplifiers Eliminate Input Transformers

As is common in electronic design, transformers do have drawbacks. Perhaps the most prominent one is cost: a Jensen Transformer, Inc. JT-115K-E costs approximately $75 US or $3.75 per dB of gain.[23] From the point of view of signal, transformers add distortion due to core saturation. Transformer distortion has a unique sonic signature that is considered an asset or a liability—depending on the transformer and whom you ask. Transformers also limit frequency response at both ends of the audio spectrum.

Furthermore, they are susceptible to picking up hum from stray electromagnetic fields.

Well-designed active *transformerless* preamplifiers can avoid these problems, lowering cost, reducing distortion, and increasing bandwidth. However, transformerless designs require far better noise performance from the active circuitry than transformer-based preamps do. Active mic preamps usually require capacitors (and other protection devices) to block potentially damaging effects of phantom power.[24,25]

16.3.5.3 The Evolution of Active Microphone Preamplifier ICs

Active balanced-input microphone preamplifier ICs were not developed until the early 1980s. Early IC fabrication processes did not permit high-quality low-noise devices, and semiconductor makers were uncertain of the demand for such products.

Active transformerless microphone preamplifiers must have fully differential inputs because they interface to balanced microphones. The amplifiers described here, both discrete and IC, use a current feedback (CFB) topology with feedback returned to one (or both) of the differential input transistor pair's emitters. Among its many attributes, current feedback permits differential gain to be set by a single resistor.

Current feedback amplifiers have a history rooted in instrumentation amplifiers. The challenges of amplifying low-level instrumentation signals are very similar to microphones. The current feedback instrumentation amplifier topology, known at least since Demrow's 1968 paper,[26] was integrated as early as 1982 as the Analog Devices AD524 developed by Scott Wurcer.[27] A simplified diagram of the AD524 is shown in Fig. 16-57. Although the AD524 was not designed as an audio preamp, the topology it used later became a de-facto standard for IC microphone preamps. Demrow and Wurcer both used a bias scheme and fully balanced topology in which they wrapped op-amps around each of the two input transistors to provide both ac and dc feedback. Gain is set by a single resistor connected between the emitters (shown as 40Ω, 404Ω, and 4.44kΩ, and feedback is provided by two resistors (R_{56} and R_{57}). The input stage is fully symmetrical and followed by a precision differential amplifier to convert the balanced output to single ended. Wurcer's AD524 required laser-trimmed thin film resistors with matching to 0.01% for an 80dB common mode rejection ratio at unity gain.

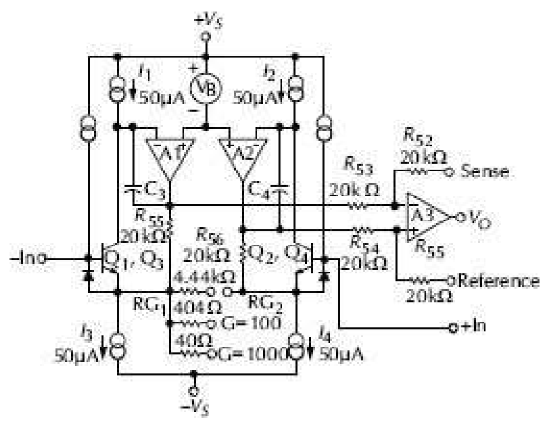


Figure 16-57. AD524 block diagram. Courtesy Analog Devices, Inc.

Audio manufacturers, using variations on current feedback and the Demrow/Wurcer instrumentation amp, produced microphone preamps based on discrete low-noise transistor front ends as early as 1978; an example is the Harrison PC1041 module.[28] In December of 1984, Graeme Cohen also published his discrete transistor topology; it was remarkably similar to the work of Demrow, Wurcer, and the Harrison preamps.[29]

Solid State Music, or SSM, which later became Solid State Microtechnology, developed the first active microphone preamp IC for professional audio around 1982.[30] SSM specialized in producing niche-market semiconductors aimed at the professional audio business. The SSM2011 was almost completely self-contained, requiring only a handful of external resistors and capacitors to provide a complete preamp system. One unique feature of the SSM2011 was an on-chip LED overload and signal presence indicator.

SSM later produced the SSM2015 and the SSM2016 designed by Derek Bowers.[31] The SSM2016, and the SSM2011 and 2015 that preceded it, did not use a fully balanced topology like Wurcer's AD524 and the Harrison PC1041. The SSM parts used an internal op-amp to convert the differential stage output to single-ended. This allowed external feedback resistors to be used, eliminating the performance penalty of on-chip diffused resistors. The SSM2016 was highly successful but required external precision resistors and up to three external trims. SSM was later acquired by Precision Monolithics and eventually by Analog Devices (ADI). In the mid-1990s, ADI discontinued the SSM2016 which then became highly sought after.

Analog Devices introduced the SSM2017 self contained preamp, also designed by Bowers, as a replacement for the SSM2016. The SSM2017 used internal laser-trimmed thin-film resistors that permitted the fully balanced topology of the AD524 and discrete preamps to be realized as an IC. Analog Devices manufactured the SSM2017 until about 2000 when it was discontinued. A year or two later, ADI released the 2019 which is available today.

The Burr Brown division of Texas Instruments offered the

INA163, which had similar performance to the SSM2017, but was not pin compatible with it. After the 2017 was discontinued, TI introduced its INA217 in the SSM2017 pinout. Today, TI produces a number of INA-family instrumentation amplifiers suitable for microphone preamps including the INA103, INA163, INA166, INA217, and the first digitally gain-controlled preamp: the PGA2500.

In 2005, THAT Corporation introduced a pair of microphone preamplifiers in pinouts to match the familiar SSM2019/INA217 as well as the INA163. The THAT 1510 and the performance-enhanced THAT 1512 use dielectric isolation to provide higher bandwidth than the junction-isolated INA and SSM series products. (Dielectric isolation is explained in section 16.3.4.1 *Voltage-Controlled Amplifiers*). In 2009, with the THAT 1570, the company introduced a "new" topology in microphone preamps which it dubbed a "de-integrated" approach.[32] This version, and the THAT 1583 (introduced in 2013) omitted the integrated feedback resistors in order to allow external control over their values, which is especially in digitally controlled applications. Fig. 16-58 shows a block diagram of this topology (compare to Fig. 16-59).

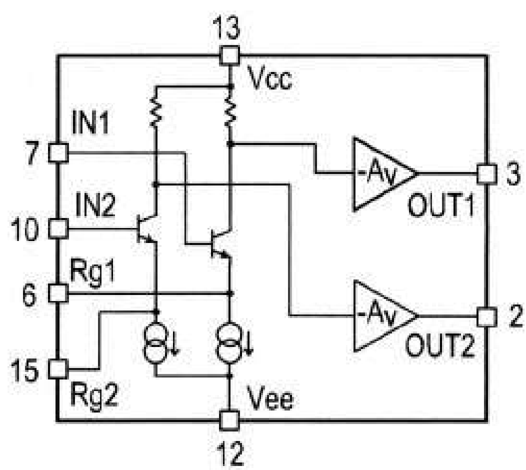


Figure 16-58. THAT 1570 block diagram. Courtesy THAT Corporation.

Because they can offer much more complex circuits at relatively low cost, integrated solutions for microphone preamplifiers generally offer better measurable performance than the prior, discrete approaches,[33] although there remains a certain "fondness" for the coloration introduced by different discrete circuits. Among the integrated offerings, today's several different families of parts each have different strengths and weaknesses. Differences exist in gain bandwidth, noise floor, distortion, gain structure, and supply consumption. The optimum part for any given application will depend on the exact requirements of the designer. A designer considering any one of these parts should compare their specs carefully before finalizing a new design.

16.3.5.4 Integrated Circuit Microphone Preamplifier Application Circuits

The THAT 1510 series block diagram is shown in Fig. 16-59. Its topology is similar to those of the TI and ADI parts. A typical application circuit is shown in Fig. 16-60. The balanced mic-level signal is applied to the input pins, In+ and In-. A single resistor (R_G), connected between pins R_{G1} and R_{G2}, sets the gain in conjunction with the internal resistors R_A and R_B. The input stage consists of two independent low-noise amplifiers in a balanced differential amplifier configuration with both ac and dc feedback returned to the emitters of the differential pair. This topology is essentially identical to the AD524 current feedback amplifier as described by Wurcer et al.

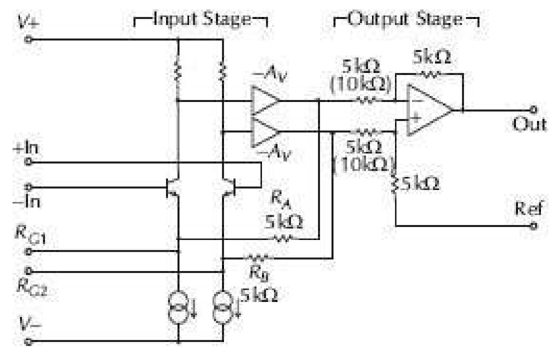


Figure 16-59. THAT 1510/1512 block diagram. Courtesy THAT Corporation.

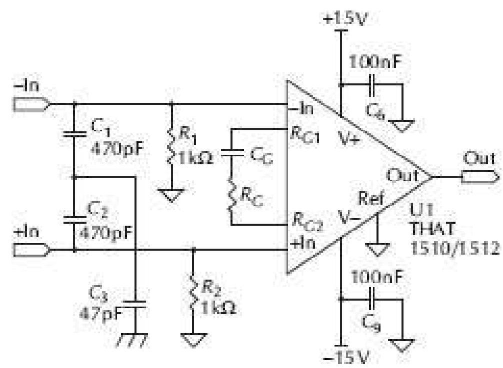


Figure 16-60. THAT 1510/1512 Basic Application. Courtesy THAT Corporation.

The output stage is a single op-amp differential amplifier that converts the balanced output of the gain stage into single-ended form. The THAT 1500 series offers a choice of gains in this stage: 0dB for the 1510, and −6dB for the 1512. Gain is controlled by the input-side resistor values: 5kΩ for the 1510 and 10kΩ for the 1512.

The gain equations for the THAT 1510 are identical to that of the SSM2017/2019 and the INA217. The INA163 and THAT 1512 have unique gain equations.

For the THAT 1510, SSM 2019, and INA217 the equation is

$$Av = 1 + \frac{10\,k\Omega}{R_G} \tag{16-65}$$

For the INA163 it is

$$Av = 1 + \frac{6\,k\Omega}{R_G} \tag{16-66}$$

For the THAT 1512 it is

$$Av = 0.5 + \frac{5\,k\Omega}{R_G} \tag{16-67}$$

where,

Av is the voltage gain of the circuit.

All these parts can reach unity gain but the value of R_G required varies considerably. For the 1510, 2017, 2019, 163, and 217, gain is 0dB (Av =1) when R_G is open: this is the minimum gain of all these ICs. For the 1512, gain is −6dB (Av = 0.5) with R_G open. To go from 60dB to 0dB gain, R_G must span a large range: 10Ω to open circuit for the 1510 and its equivalents.

R_G is typically a reverse log potentiometer (or set of switched resistors) to provide smooth rotational control of gain. In many applications, and, as shown in Fig. 16-60, a large value capacitor is placed in series with R_G to limit the dc gain of the device, thus preventing shifts in output dc-offset with gain changes. For 60dB of gain with the THAT 1512, R_G = 5Ω (6Ω in the case of the INA163). Because of this, C_G must be quite large, typically ranging from 1000 to 6800µF to preserve low frequency response. Fortunately, C_G does not have to support large voltages: 6.3V is acceptable.

Parts from all manufacturers exhibit excellent voltage noise performance of $\sim 1\,\text{nV}/\sqrt{\text{Hz}}$ at high gains. Differences in noise performance begin to show up at lower gains, with the THAT 1512 offering the best performance ~34 nV/$\sqrt{Hz}$ at 0dB gain) of the group. These parts are all generally optimized for the relatively low source impedances of dynamic microphones with typically a few hundred ohm output impedance.

Fig. 16-60 provides an application example for direct connection to a dynamic microphone. Capacitors C_1–C_3 filter out radio frequencies that might cause interference (forming an RFI filter). R_1 and R_2 provide a bias current path for the inputs and terminate the microphone output. R_G sets the gain as defined in the previous equation. C_G blocks dc in the input stage feedback loop, limiting the dc gain of this stage to unity and avoiding output offset change with gain. C_6 and C_9 provide power supply bypass.

Fig. 16-61 shows the THAT 1512 used as a preamp capable of being used with phantom power. C_1–C_3 provide RFI protection. R_5 and R_6 feed phantom power to the microphone. R_9 terminates the microphone. C_4 and C_5 block 48 Vdc phantom potential from the THAT 1512. R_3, R_4, and D_1–D_4 provide current limiting and overvoltage protection from phantom power faults. R_1 and R_2 are made larger than previously shown to reduce the loading on C_4 and C_5.

16.3.6 Digitally Controlled Microphone Preamplifiers

In the 2000s, with the increasing acceptance of digital techniques for many audio processing functions, including not only recording but mixing and effects, it became clear that remote control of microphone preamp gain was a highly desirable function. Lacking integrated solutions for digital control, designers were creative in using anything from relays to analog switches to change gain resistors. Invariably, though these circuits were generally large, complex, and power hungry. Starting in the late 1990s, designers began to ask for integrated alternatives.

In 2003, Texas Instruments (TI) introduced the first monolithic digitally controlled microphone preamplifier, the PGA2500. Fig. 16-62, shows its block diagram. This part offered gain of 10dB through 65dB in 1dB steps, with an additional unity gain setting. While considered expensive by many, its noise and distortion performance were excellent, with gain controlled via a serial digital interface. The PGA2500 included a dc servo, zero-crossing detector, and accepted differential inputs while delivering differential outputs suitable for connection to a high-quality A/D converter. It was adopted by many pro audio manufacturers.

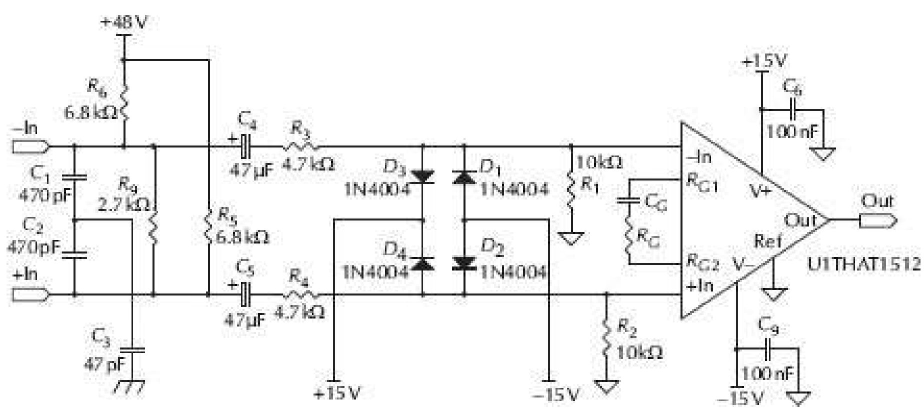


Figure 16-61. THAT preamp circuit with phantom power. Courtesy THAT Corporation.

In 2009, THAT Corporation announced a different approach to solving this problem: the THAT 5171 digital preamp controller to accompany a THAT 1570 analog gain block. Fig. 16-63 shows a simplified schematic of this approach. THAT's chip set offered gain of 13.6dB to 68.6dB in 1dB steps, and an additional 5.6dB gain setting. While similar in many respects to the PGA2500, including cost, features, and performance, a major difference was the supply range of the THAT offering (up to ±17V) compared to that of the TI part (±5V), with attendant implications to headroom.

A short while later, TI brought out the lower-cost PGA2505, similar to the PGA2500 but with 3dB steps. THAT followed up with a lower-cost (high voltage) pair: the THAT 5173 (3dB per step) and THAT 1583 analog gain block. The two-part approach, while perhaps less convenient, offers designers the choice of designing their own analog preamplifier to mate with the separate digital control section, which can also be used "in reverse" as a digitally controlled attenuator. At press time, it seems likely that one or both of these companies, if not others, will continue to offer more parts in this area; as audio processing moves more towards DSP, only the interfaces remain as analog.

Many variations are possible on the basic circuits presented above. The analog-controlled preamps can be extended with dc servos to reduce or eliminate some of the ac-coupling needed. Floating power supply arrangements can be used to eliminate input coupling capacitors. Microphone preamps with response to dc have also been proposed. The digitally controlled preamps can be modified similarly, particularly the ones based on THAT's building-block approach. For more information on possible configurations, see application notes published by Analog Devices, Texas Instruments, and THAT Corporation. (All available at their respective web sites: www.analog.com, www.ti.com, www.thatcorp.com.)

Modern IC microphone preamplifiers provide a simple building block with performance equaling or beating discrete solutions without a costly input transformer. However, one important weakness of the transformerless "active" approach is in susceptibility to damage via phantom power faults. All the manufacturers of IC preamps recommend schemes to protect the input devices from damage when phantom power is used, and designers are well advised to pay careful attention to this area. See especially references 25 and 36 for some startling information on the "phantom menace".

16.3.7 Balanced Line Interfaces

In professional audio, interconnections between devices frequently use *balanced lines*. These are especially important when analog audio signals are sent over long distances, where the ground references for the send and receive ends are different or where noise and interference may be picked up in the interconnection cables.

Differences in signal ground potentials arise as a result of current flowing into power-line safety grounds. These currents, flowing through finite ground impedances between equipment, can produce up to several volts potential difference between the ground references within a single building. These currents, usually at the power line frequency and its harmonics, produce the all-too-familiar hum and buzz known to every sound engineer.

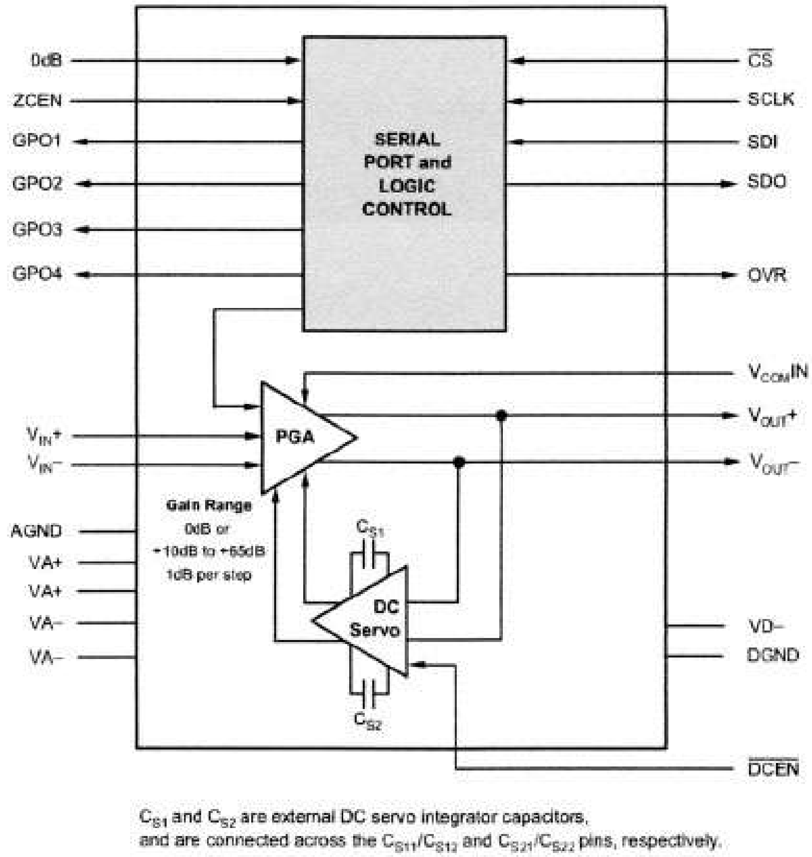


Figure 16-62. PGA2500 block diagram. Courtesy Texas Instruments Incorporated.

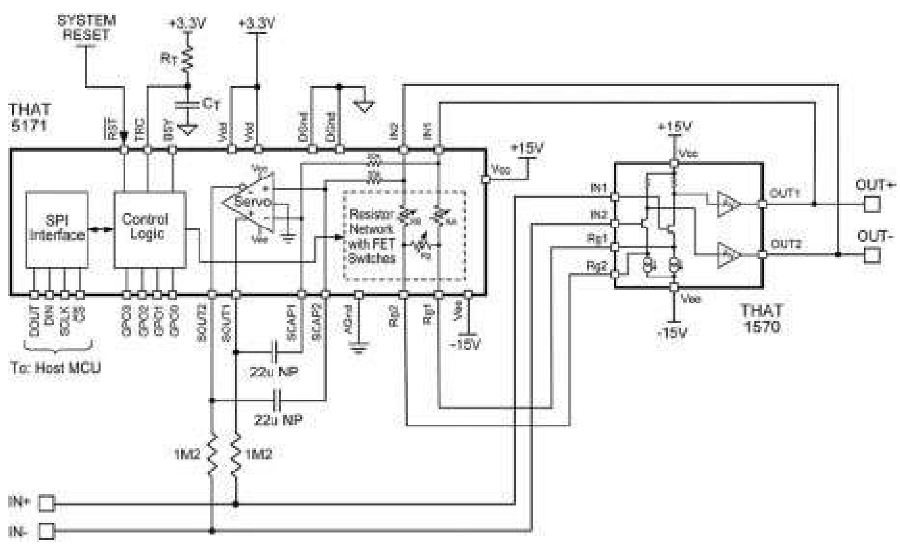


Figure 16-63. THAT 5171 and 1570 analog gain circuit. Courtesy THAT Corporation.

Two other forms of interference, electrostatic and magnetic, also create difficulty. Cable shielding reduces electrostatic interference from fields, typically using braided copper, foil wrap, or both. Magnetic interference from fields is much harder to prevent via shielding. The impact of magnetic fields in signal cables is reduced by balanced cable construction using twisted pair cable. Balanced circuits benefit from the pair's twist by ensuring that magnetic fields cut each conductor equally. This in turn ensures that the currents produced by these fields appear in common mode, wherein the voltages produced appear equally in both inputs.

The balanced line approach comes out of telephony, in which voice communications are transmitted over many miles of unshielded twisted pair cables with reasonable fidelity and freedom from hum and interference pickup. Two principles allow balanced lines to work. First, interference—whether magnetic or electrostatic—is induced equally in both wires in the twisted paired-conductor cable, and second, the circuits formed by the source and receiver, plus the two wires connecting them form a balanced bridge,[34] Fig. 16-64. Interfering signals appear identically (in common-mode) at the two (+ and –) inputs, while the desired audio signal appears as a difference (the differential signal) between the two inputs.

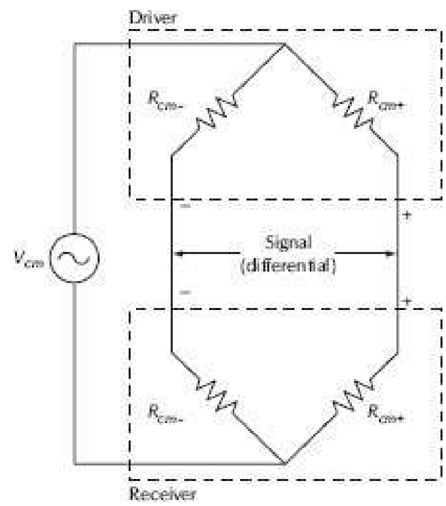


Figure 16-64. Balanced bridge. Courtesy THAT Corporation.

A common misconception in the design of balanced interfaces is that the audio signals must be transmitted as equal and opposite polarity on both lines. While this is desirable to maximize headroom in many situations, it is unnecessary to preserve fidelity and avoid noise pickup. It is enough if the bridge formed by the combination of the circuit's two common-mode source *impedances* (not the *signals*) working against the two common-mode load *impedances* remains balanced in all circumstances.

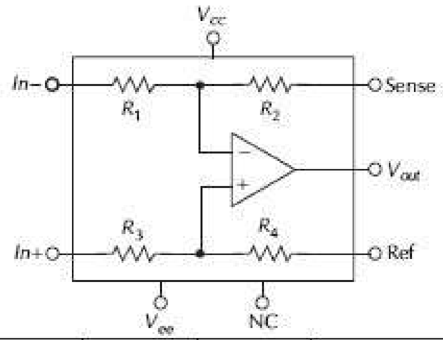
In telephony, and in early professional audio systems, transformers were used at both the inputs and outputs of audio gear to maintain bridge balance. Well-made output transformers have closely matched common-mode source impedances and very high common-mode impedance. (Common-mode impedance is the equivalent impedance from one or both conductors to ground.) The floating connections of most transformers—whether used for inputs or outputs—naturally offer very large common-mode impedance. Both of these factors, matched source impedances for output transformers, and high common-mode impedance (to ground) for both input and output transformers, work together to maintain the balance of the source/load impedance bridge across a wide range of circumstances. In addition, transformers offer *galvanic isolation*, which is sometimes helpful when faced with particularly difficult grounding situations.

On the other hand, as noted previously in the section on preamplifiers, transformers have drawbacks of high cost, limited bandwidth, distortion at high signal levels, and magnetic pickup.

16.3.7.1 Balanced Line Inputs

Transformers were used in early balanced line input stages, particularly in the days before inexpensive op-amps made it attractive to replace them. The advent of inexpensive op-amps, especially compared to the cost of transformers, motivated the development of active transformerless inputs. As the state of the art in op-amps improved, transformer-coupled inputs were replaced by less expensive, high-performance active stages based on general-purpose parts like the Texas Instruments TL070 and TL080 series, the National Semiconductor LF351 series, and the Signetics NE5534.

As with microphone preamplifiers, common-mode rejection is an important specification for line receiver inputs. The most common configuration for active balanced line input stages used in professional audio is the simple circuit shown in Fig. 16-65. To maintain high common-mode rejection (CMR), the four resistors used must match very closely. To maintain a 90dB CMR, for example, the resistor ratio R_1/R_2 must match that of R_3/R_4 within 0.005%. The requirement for precision-matched resistors to provide high CMR drove the development of specialized line receiver ICs.



Part no.	Gain	R_1, R_3	R_2, R_4
THAT1240	0dB	9kΩ	9kΩ
THAT1243	−3dB	10.5kΩ	7.5kΩ
THAT1246	−6dB	12kΩ	6kΩ

Figure 16-65. 1240 basic circuit. Courtesy THAT Corporation.

To maintain the high CMR potential of precision balanced line receivers, the interconnections between stages must be made through low-resistance connections, and the impedances in both lines of the circuit must be very nearly identical. A few ohms of contact resistance external to the line driver and receiver (due, for example, to oxidation or poor contact) or any imbalance in the driving circuit, can significantly reduce CMR by unbalancing the bridge circuit. The imbalance can be at the source, in the middle at a cable junction, or near the input of the receiving equipment. Although many balanced line receivers provide excellent CMR under ideal conditions, few provide the performance of a transformer under less-than-ideal real world circumstances.

16.3.7.2 Balanced Line Outputs

Transformers were also used in early balanced output stages, for the same reasons as they are used in inputs. However, to drive 600Ω loads, an output transformer must have more current capacity than an input transformer that supports the same voltage levels. This increased the cost of output transformers, requiring more copper and steel than input-side transformers and putting pressure on designers to find alternative outputs. Early active stages were either discrete or used discrete output transistors to boost the current available from op-amps. The NE5534, with its capability to directly drive a 600Ω load, made it possible to use op-amps without additional buffering as output stages.

One desirable property of transformer-coupled output stages was that the output voltage was the same regardless of whether the output was connected differentially or in single-ended fashion. While professional audio gear has traditionally used balanced input stages, sound engineers commonly must interface to consumer and semi-pro gear that use single-ended input connections referenced to ground. Transformers behave just as well when one terminal of their output winding is shorted to the ground of a subsequent single-ended input stage. On the other hand, an active-balanced output stage that provides equal and opposite drive to the positive and negative outputs will likely have trouble if one output is shorted to ground.

This led to the development of a *cross-coupled* topology by Thomas Hay of MCI that allowed an active balanced output stage to mimic this property of transformers.[35] When loaded equally by reasonable impedances (e.g., 600Ω or more) Hay's circuit delivers substantially equal—and opposite-polarity voltage signals at either output. However, because feedback is taken differentially, when one leg is shorted to ground, the feedback loop automatically produces twice the voltage at the opposing output terminal. This mimics the behavior of a transformer in the same situation.

While very clever, this circuit has at least two drawbacks. First, its resistors must be matched very precisely. A tolerance of 0.1% (or better) is often needed to ensure stability, minimize sensitivity to output loading, and maintain close matching of the voltages at either output. (Though, as noted earlier, this last requirement is unnecessary for good performance.) The second drawback is that the power supply voltage available to the two amplifiers limits the voltage swing at each output. When loaded differentially, the output stage can provide twice the voltage swing than it can when driving a single-ended load. But this means that headroom is reduced 6dB with single-ended loads.

One way to ensure the precise matching required by Hay's circuit is to use laser-trimmed thin-film resistors in an integrated circuit. SSM was the first to do just that when they introduced the SSM2142, a balanced line output driver with a cross-coupled topology.

16.3.7.3 Integrated Circuits for Balanced Line Interfaces

Instrumentation amplifier inputs have similar requirements to those of an audio line receiver. The INA105, originally produced by Burr Brown and now Texas Instruments, was an early instrumentation amplifier that featured laser-trimmed resistors to provide 86dB common-mode rejection. Although its application in professional audio was limited due to the performance of its internal op-amps, the INA105 served as the basis for the modern audio balanced line receiver.

In 1989, the SSM Audio Products Division of Precision Monolithics introduced the SSM2141 balanced line receiver and companion SSM2142 line driver. The SSM2141 was offered in the same pinout as the INA105 but provided low noise and a slew rate of almost 10V/μs. With a typical CMR of 90 dB, the pro-audio industry finally had a low-cost, high-performance replacement for the line input transformer. The SSM2142 line driver, with its cross-coupled outputs, became a low-cost replacement for the output transformer. Both parts have been quite successful.

Today, Analog Devices (which acquired Precision Monolithics) makes the SSM2141 line receiver and the SSM2142 line driver. The SSM2143 line receiver, designed for 6dB attenuation, was introduced later to offer increased input headroom. It also provides overall unity gain operation when used with an SSM2142 line driver, which has 6dB of gain.

The Burr Brown division of Texas Instruments now produces a similar family of balanced line drivers and receivers, including dual units. The INA134 audio differential line receiver is a second source to the SSM2141. The INA137 is similar to the SSM2143 and also permits gains of ±6dB. Both devices owe their pinouts to the original INA105. Dual versions of both parts are available as the INA2134 and 2137. TI also makes cross-coupled line drivers known as the DRV134 and DRV135.

THAT Corporation also makes balanced line drivers and receivers. THAT's 1240 series single and 1280 series dual balanced line receivers use laser-trimmed resistors to provide high common rejection in the familiar SSM2141 (single) and INA2134 (dual) pinouts. For lower cost applications, THAT offers the 1250- and 1290-series single and dual line receivers. These parts eliminate laser trimming, which sacrifices CMR to reduce cost. Notably, THAT offers both dual and single line receivers in the unique configuration of ±3dB gain, which can optimize dynamic range for many common applications.

THAT Corporation also offers a unique line receiver, the THAT

1200 series, based on technology licensed from William E. Whitlock of Jensen Transformers, Inc. (U.S. Patent 5,568,561).[36] This design, dubbed InGenius (a trademark of THAT Corporation), bootstraps the common-mode input impedance to raise it into the megohm range of transformers. This overcomes the loss of common-mode rejection when the impedances feeding the line receiver are slightly unbalanced and permits transformer-like operation. The InGenius circuit will be discussed in a following section.

THAT also offers the THAT 1646 balanced line driver, which has identical pinout to the SSM2142 and DRV134/135. THAT's 1606 balanced line driver is unique among these parts in that it provides not only a differential output, but also a differential input—enabling a more direct connection to digital to analog converters.

The THAT 1646 and 1606 use a unique output topology unlike conventional cross-coupled outputs which THAT calls "OutSmarts" (another trademark). OutSmarts is based on U.S. Patent 4,979,218 issued to Chris Strahm, then of Audio Teknology Incorporated, and U.S. Patent 6,316,970 issued to Gary Hebert of THAT Corporation.[37, 38] Strahm pioneered the use of a fully-differential op-amp as a line driver, using a common-mode feedback loop to raise the common-mode output impedance, thus producing the "floating" behavior that mimics an output transformer. Hebert observed that conventional cross-coupled outputs and Strahm's circuit both lose common-mode feedback when one output is shorted to ground to accommodate a single-ended load and the active output is driven into clipping. This allows large signal currents to flow into ground, increasing crosstalk and distortion. Hebert's circuit avoids this by implementing the common-mode feedback loop such that it remains operational even if the differential path is disabled due to clipping. Application circuits for the THAT 1646 will be described in the section 16.3.7.7 *Balanced Line Drivers*.

16.3.7.4 Balanced Line Input Application Circuits

Conventional balanced line receivers from Analog Devices, Texas Instruments, and THAT Corporation are substantially equivalent to the THAT 1240 circuit shown in Fig. 16-66. Some variations exist in the values of R_1–R_4 from one manufacturer to the other that will influence input impedance and noise. The ratio of R_1/R_3 to R_2/R_4 establishes the gain with $R_1 = R_3$ and $R_2 = R_4$. V_{out} is normally connected to the sense input resistor with the reference pin grounded.

Line receivers usually operate at either unity gain (SSM2141, INA134, THAT 1240, or THAT 1250) or in attenuation (SSM2143, INA137, THAT 1243, or THAT 1246, etc.). When a perfectly balanced signal (with each input line swinging 1/2 the differential voltage) is converted from differential to single-ended by a unity gain receiver, the output must swing twice the voltage of either input line for a net voltage gain from either input of +6dB. With only +21dBu output voltage available from a line receiver powered by bipolar 15V supplies, additional attenuation is often needed to provide headroom to accommodate pro audio signal levels of +24dBu or more. The ratios R_1/R_2 and R_3/R_4 are 2:1 in the SSM2143, INA137, and THAT 1246 to provide 6dB attenuation. These parts accommodate up to +27dBu inputs without clipping their outputs when running from bipolar 15V supplies. The THAT 1243, and THAT's other ±3dB parts (the 1253, 1283, and 1293) are unique with their 0.707 attenuation. This permits a line receiver that accommodates +24dBu inputs but avoids additional attenuation that increases noise. A −3dB line receiver is shown in Fig. 16-67.

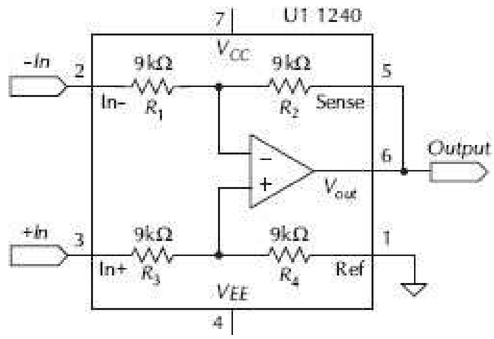


Figure 16-66. THAT 1240 with 0 dB gain. Courtesy THAT Corporation.

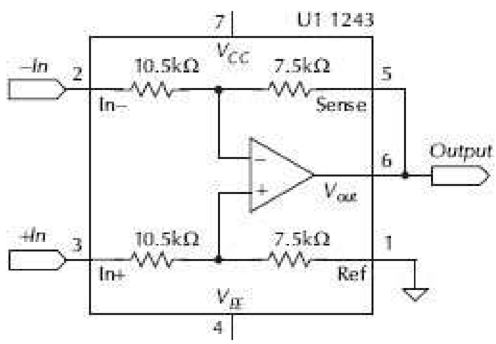


Figure 16-67. THAT 1243 with 3dB attenuation. Courtesy THAT Corporation.

The ±6dB parts from all three manufacturers (and the ±3dB parts from THAT) may be configured for gain instead of attenuation. To accomplish this, the reference and sense pins are used as inputs with the *In*−pin connected to V_{out} and the In+ pin connected to ground. A line receiver configured for 6dB gain is shown in Fig. 16-68.

Balanced line receivers may also be used to provide sum-difference networks for mid-side (M/S or M-S) encoding/decoding as well as general-purpose applications requiring precise difference amplifiers. Such applications take advantage of the precise matching of resistor ratios possible via monolithic, laser-trimmed resistors. In fact, while these parts are usually promoted as input stages, they have applications to many circuits where precise resistor ratios are required. The typical 90dB common-mode rejection advertised by many of these manufacturers requires ratio matching to within 0.005%.

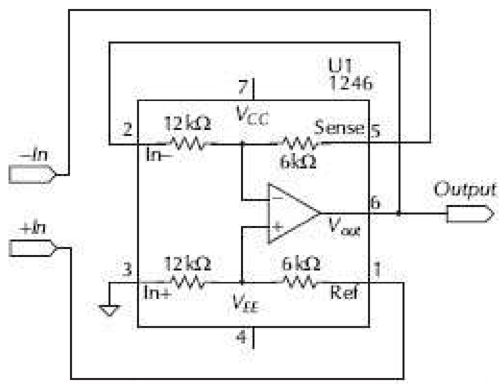


Figure 16-68. THAT 1246 with 6dB gain. Courtesy THAT Corporation.

Any resistance external to the line receiver input appears in series with the highly matched internal resistors. A basic line receiver connected to an imbalanced circuit is shown in Fig. 16-69. Even a slight imbalance, one as low as 10Ω from connector oxidation or poor contact, can degrade common-mode rejection. Fig. 16-70 compares the reduction in CMR for low common-mode impedance line receivers versus the THAT 1200 series or a transformer.

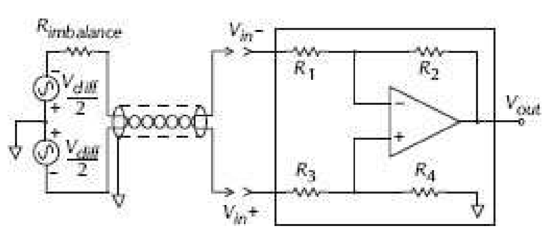


Figure 16-69. Balanced circuit with imbalance. Courtesy THAT Corporation.

The degradation of common-mode rejection from impedance imbalance comes from the relatively low-impedance load of simple line receivers interacting with external impedance imbalances. Since unwanted hum and noise appear in common-mode (as the same signal in both inputs), common-mode loading by common-mode input impedance is often a significant source of error. (The differential input impedance is the load seen by differential signals; the common-mode input impedances is the load seen by common-mode signals.) To reduce the effect of impedance imbalance, the common-mode input impedance, but not the differential impedance, must be made very high.

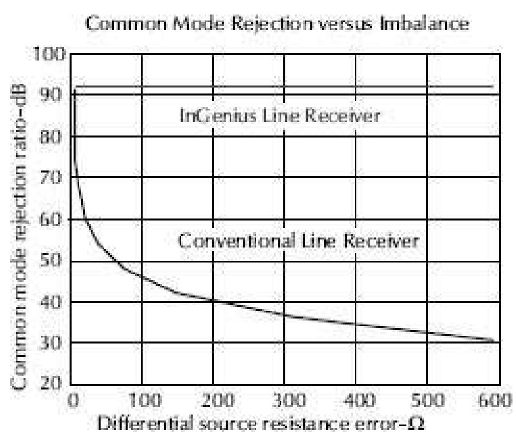


Figure 16-70. CMR imbalance versus source. Courtesy THAT Corporation.

16.3.7.5 Balanced Line Receivers with the Common-Mode Performance of a Transformer

The transformer input stage has one major advantage over most active input stages: its common-mode input impedance is extremely high regardless of its differential input impedance. This is because transformers offer floating connections without any connection to ground. Active stages, especially those made with the simple SSM2141-type IC have common-mode input impedances of approximately the same value as their differential input impedance. (Note that for simple differential stages such as these, the common-mode and differential input impedances are not always the same.) Op-amp input bias current considerations generally make it difficult to use very high impedances for these simple stages. A bigger problem is that the noise of these stages increases with the square root of the impedances chosen, so large input impedances inevitably cause higher noise.

Noise and op-amp requirements led designers to choose relatively low impedances (10k~25kΩ). Unfortunately, this means these stages have relatively low common-mode input impedance as well (20k~50kΩ). This interacts with the common-mode output impedance (also relative to ground) of the driving stage, and added cable or connector resistance. If the driver, cable, or connectors provide an unequal, nonzero common-mode output impedance, the input stage loading will upset the natural balance of any common-mode signal, converting it from common-mode to differential. No amount of precision in the input stage's resistors will reject this common-mode-turned-to-differential signal. This can completely spoil the apparently fine performance available from the precisely matched resistors in simple input stages.

An instrumentation amplifier, Fig. 16-71, may be used to increase common-mode input impedance. Input resistors R_{i1} and R_{i2} must be present to supply a bias current return path for buffer amplifiers OA1 and OA2. R_{i1} and R_{i2} can be made large—in the MΩ range—to minimize the effect of impedance imbalance. While it is possible to use this technique to make line receivers with very high common-mode input impedances, doing so requires specialized op-amps with bias-current compensation or FET input stages. In addition, this requires two more op-amps in addition to the basic differential stage (OA3).

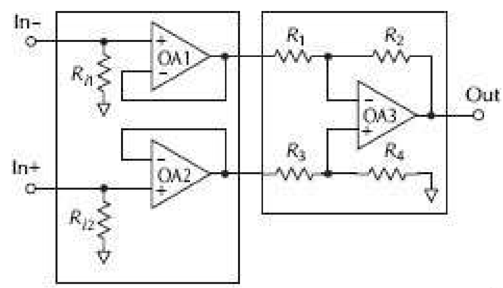


Figure 16-71. Instrumentation amplifier. Courtesy THAT Corporation.

With additional circuitry, even higher performance can be obtained by modifying the basic instrumentation amplifier circuit. Bill Whitlock of Jensen Transformers, Inc. developed and patented (U.S. Patent 5,568,561) a method of applying bootstrapping to the instrumentation amplifier in order to further raise common-mode input impedance.[36] THAT Corporation incorporated this technology in its InGenius series of input stage ICs.

16.3.7.6 InGenius High Common-Mode Rejection Line Receiver ICs

Fig. 16-72 shows the general principle behind ac bootstrapping in a single-ended connection. By feeding the ac component of the input into the junction of R_a and R_b, the effective value of R_a (at ac) can be made to appear quite large. The dc value of the input impedance (neglecting R_s being in parallel) is $R_a + R_b$. Because of bootstrapping, R_a and R_b can be made relatively small values to provide op-amp bias current, but the ac load on R_s (Z_{in}) can be made to appear to be extremely large relative to the actual value of R_a.

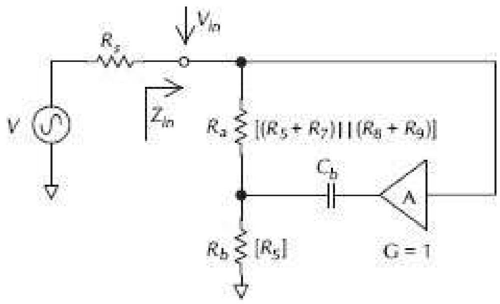


Figure 16-72. Single ended bootstrap. Courtesy THAT Corporation.

A circuit diagram of an InGenius balanced line receiver using the THAT 1200 is shown in Fig. 16-73. (All the op-amps and resistors are internal to the IC.) R_5–R_9 provides dc bias to internal op-amps OA1 and OA2. Op-amp OA4, along with R_{10} and R_{11} extract the common-mode component at the input and feed the ac common-mode component back through *Cb* to the junction of R_7 and R_8. Because of this positive feedback, the effective values of R_7 and R_8—at ac—are multiplied into the MΩ range. In its data sheet for the 1200 series ICs, THAT cautions that *Cb* should be at least 10µf to maintain common-mode input impedance (Z_{inCM}) of at least 1MΩ at 50Hz. Larger capacitors can increase Z_{inCM} at low power-line frequencies up to the IC's practical limit of ~10MΩ. This limitation is due to the precision of the gain of the internal amplifiers.

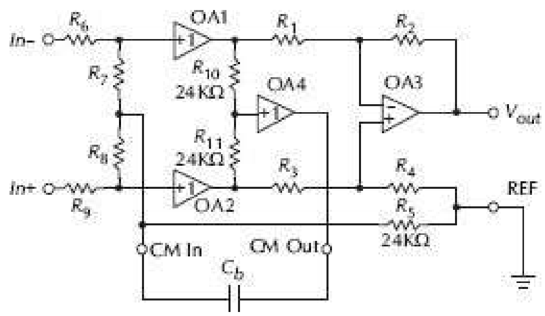


Figure 16-73. Balanced line receiver. Courtesy THAT Corporation.

The outputs of OA1 and OA2 contain replicas of the positive and negative input signals. These are converted to single-ended form by a precision differential amplifier OA3 and laser-trimmed resistors R_1–R_4. Because OA1 and OA2 isolate the differential amplifier, and the positive common-mode feedback ensures very high common-mode input impedance, a 1200-series input stage provides 90dB CMR even with high levels of imbalance.

It took Bill Whitlock and Jensen Transformers, Inc. to provide an active input as good as a transformer operating under conditions likely to be found in the real world.

A basic application circuit using the THAT 1200 series parts is shown in Fig. 16-74.

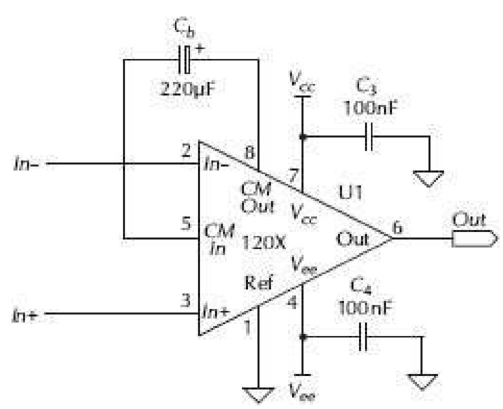


Figure 16-74. InGenius basic application. Courtesy THAT

Corporation.

16.3.7.7 Balanced Line Drivers

The Analog Devices SSM2142 and Texas Instruments DRV series balanced line drivers use a cross-coupled method to emulate a transformer's floating connection and provide constant level with both single-ended (grounded) terminations and fully balanced loads. A block diagram of a cross-coupled line driver is shown in Fig. 16-75. The force and sense lines are normally connected to each output either directly or through small electrolytic coupling capacitors. A typical application of the SSM2142 driving an SSM2141 (or SSM2143) line receiver is provided in Fig. 16-76. If one output of the cross-coupled line driver outputs is shorted to ground in order to provide a single-ended termination, the full short-circuit current of the device will flow into ground. Although this is not harmful to the device, and is in fact a recommended practice, large clipped signal currents will flow into ground, which can produce crosstalk within the product using the stage, as well as in the output signal line itself.

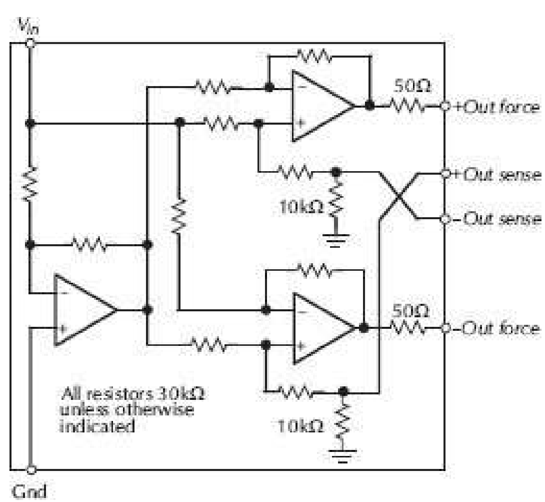


Figure 16-75. SSM2142 cross coupled output. Courtesy Analog

Devices, Inc.

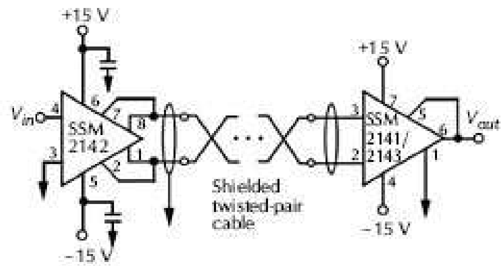


Figure 16-76. SSM2142 driving a SSM2141 line receiver. Courtesy Analog Devices, Inc.

THAT Corporation licensed a patented technology developed by Chris Strahm of Audio Teknology Incorporated. U.S. Patent 4,979,218, issued in December 1990, describes a balanced line driver that emulates a floating transformer output by providing a current-feedback system where the current from each output is equal and out of phase to the opposing output.[37] THAT trademarked this technology as OutSmarts and introduced its THAT 1646 line driver having identical pinout and functionality to the SSM2142. THAT also offers a version of the 1646 with differential inputs known as the THAT 1606. Fig. 16-77 is a simplified block diagram of the THAT 1646.

The THAT 1646 OutSmarts internal circuitry differs from other manufacturer's offerings. Outputs D_{out-} and D_{out+} supply current through 25Ω build-out resistors. Feedback from both sides of these resistors is returned into *two* internal common-mode feedback paths. The driven side of the build-out resistors is fed back into the common-mode C_{in-} input while the load side of the build-out resistors, through the sense– and sense+ pins, provides feedback into the C_{in+} input. A current feedback bridge circuit allows the 1646 to drive one output shorted to ground to allow a single-ended load to be connected. The output short increases apparent gain at the single output by 6dB (though, of course, the net gain from input to differential output has not changed), similarly to conventional cross-coupled topologies (and exactly how an output transformer would behave). However, it does so without loss of the common-mode feedback loop during clipping. This configuration prevents large, clipped signal currents flowing into ground, which in turn reduces the crosstalk and distortion produced by these currents.

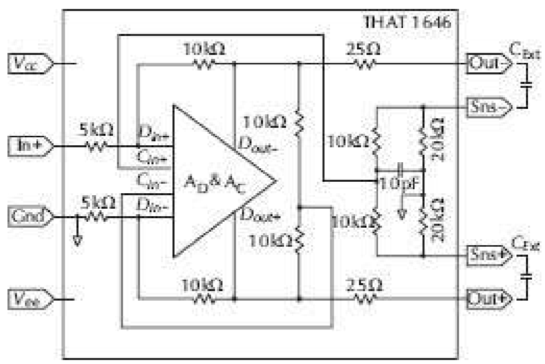


Figure 16-77. THAT 1646 block diagram. Courtesy THAT Corporation.

A typical application circuit for the THAT 1646 is shown in Fig. 16-78.

To reduce the amount of common-mode dc offset, the circuit in Fig. 16-79 is recommended. Capacitors C_1 and C_2, outside the primary signal path, minimize common-mode dc gain, which reduces common-mode output offset voltage and the effect of OutSmarts at low frequencies. Similar capacitors are used in the ADI and TI parts to the same effect.

THAT's 1606 version of OutSmarts provides a differential input for easier connection to a digital-to-analog converter's output. A typical application of the THAT 1606 is shown in Fig. 16-80. This circuit includes components D_1 through D_4 which protect against phantom power faults. Another advantage to the 1606 is that it requires only a single low-value capacitor (typically a film type) versus the two larger capacitors required by the THAT 1646, SSM2142, or DRV134.

Active balanced line drivers and receivers offer numerous advantages over transformers providing lower cost, weight, and distortion, along with greater bandwidth and freedom from magnetic pickup. When used properly, active devices perform as well as, and in many ways better than, the transformers they replace. With careful selection of modern IC building blocks from several IC makers, excellent performance is easy to achieve.

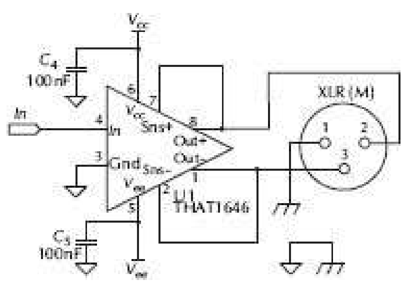


Figure 16-78. THAT 1646 application. Courtesy THAT Corporation.

16.3.8 Digital Integrated Circuits

Digital ICs produce an output of either 0 or 1. With digital circuits, when the input reaches a preset level, the output switches polarity. This makes digital circuitry relatively immune to noise.

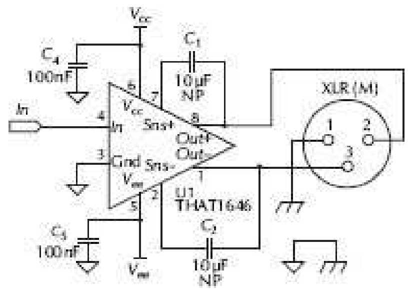


Figure 16-79. THAT 1646 CMR offset reduction circuit. Courtesy THAT Corporation.

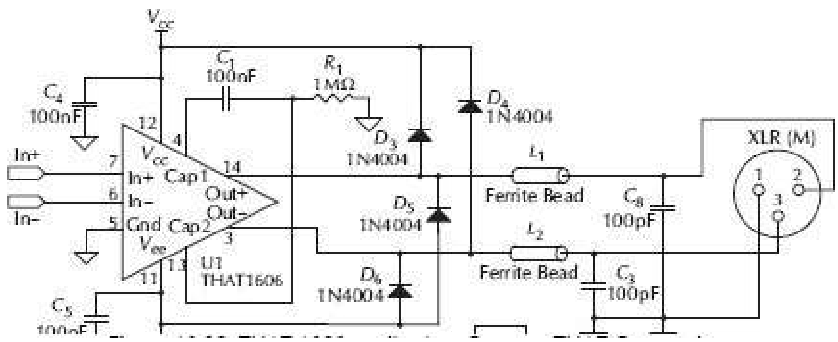


Figure 16-80. THAT 1606 application. Courtesy THAT Corporation.

Bipolar technology is characterized by very fast propagation time and high power consumption, while MOS technology has relatively slow propagation times, low power consumption, and high circuit density. Table 16-4 gives some of the terminology common to digital circuitry and digital ICs. Fig. 16-81 shows typical circuits and characteristics of the major bipolar logic families.

Table 16-4. Digital Circuit Terminology

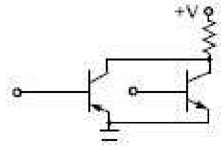
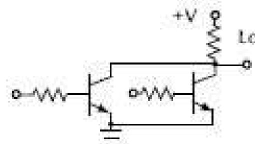
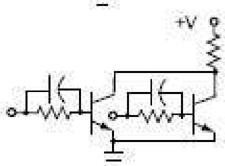
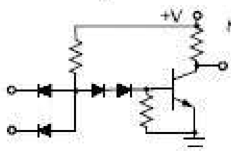
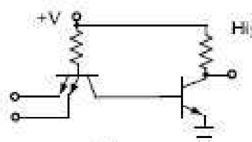
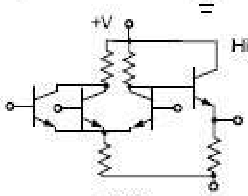
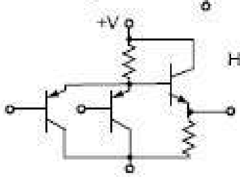
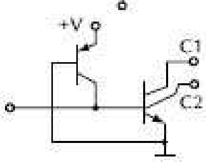
Adder	Switching circuits that generate sum and carry bits.
Address	A code that designates the location of information and instructions.
AND	A Boolean logic operation that performs multiplication. All inputs must be true for the output to be true.

Asynchronous	A free-running switching network that triggers successive instructions.
Bit	Abbreviation for binary digit; a unit of binary information.
Buffer	A noninverting circuit used to handle fan-out or convert input and output levels.
Byte	A fixed-length binary-bit pattern (word).
Clear	To restore a device to its standard state.
Clock	A pulse generator used to control timing of switching and memory circuits.
Clock rate	The frequency (speed) at which the clock operates. This is normally the major speed of the computer.
Counter	A device capable of changing states in a specified sequence or number of input signals.
Counter, binary	A single input flip-flop. Whenever a pulse appears at the input, the flip-flop changes state (called a *T flip-flop*).
Counter, ring	A loop or circuit of interconnected flip-flops connected so that only one is on at any given time. As input signals are received, the position of the on state moves in sequence from one flip-flop to another around the loop.
Fan-in	The number of inputs available on a gate.
Fan-out	The number of gates that a given gate can drive. The term is applicable only within a given logic family.
Flip-flop	A circuit having two stable states and the ability to change from one state to the other on application of a signal in a specified manner.
Flip-flop, D	*D* stands for delay. A flip-flop whose output is a function of the input that appeared one pulse earlier; that is, if a 1 appears at its input, the output will be a 1 a pulse later.
Flip-flop, JK	A flip-flop having two inputs designated *J* and K. At the application of a clock pulse, a 1 on the *J* input will set the flip-flop to the 1 or on state; a 1 on the K input will reset it to the 0 or off state; and 1s simultaneously on both inputs will cause it to change

	state regardless of the state it had been in.
Flip-flop, RS	A flip-flop having two inputs designated *R* and S. At the application of a clock pulse, a 1 on the S input will set the flip-flop to the 1 or on state, and a 1 on the *R* input will reset it to the 0 or off state. It is assumed that 1s will never appear simultaneously at both inputs.
Flip-flop, R, S, T	A flip-flop having three inputs, *R, S*, and *T*. The *R* and *S* inputs produce states as described for the *RS* flip-flop above; the *T* input causes the flip-flop to change states.
Flip-flop, T	A flip-flop having only one input. A pulse appearing on the input will cause the flip-flop to change states.
Gate	A circuit having two or more inputs and one output, the output depending on the combination of logic signals at the inputs. There are four gates: AND, OR, NAND, NOR. The definitions below assume positive logic is used.
Gate, AND	All inputs must have 1-state signals to produce a 0-state output.
Gate, NAND	All inputs must have 1-state signals to produce a 1-state output.
Gate, NOR	Any one or more inputs having a 1-state signal will yield a 0-state output.
Gate, OR	Any one or more inputs having a 1-state signal is sufficient to produce a 1-state output.
Inverter	The output is always in the opposite logic state as the input. Also called a NOT circuit.
Memory	A storage device into which information can be inserted and held for use at a later time.
NAND gate (D = ABC for positive inputs)	The simultaneous presence of all inputs in the positive state generates an inverted output.
Negative logic	The more negative voltage (or current) level represents the 1-state; the less negative level represents the 0-state.

NOR gate (D = A + B + C for positive inputs)	The presence of one or more positive inputs generates an inverted output.
NOT	A boolean logic operator indicating negation. A variable designated NOT will be the opposite of its AND or OR function. A switching function for only one variable.
OR	A boolean operator analogous to addition (except that two truths will only add up to one truth). Of two variables, only one need be true for the output to be true.
Parallel operator	Pertaining to the manipulation of information within computer circuits in which the digits of a word are transmitted simultaneously on separate lines. It is faster than serial operation but requires more equipment.
Positive logic	The more positive voltage (or current) level represents the 1-state; the less positive level represents the 0-state.
Propagation delay	A measure of the time required for a change in logic level to spread through a chain of circuit elements.
Pulse	A change of voltage or current of some finite duration and magnitude. The duration is called the *pulse width* or *pulse length;* the magnitude of the change is called the *pulse amplitude* or *pulse height*.
Register	A device used to store a certain number of digits in the computer circuits, often one word. Certain registers may also include provisions for shifting, circulating, or other operations.
Rise time	A measure of the time required for a circuit to change its output from a low level (zero) to a high level (one).
Serial operation	The handling of information within computer circuits in which the digits of a word are transmitted one at a time along a single line. Though slower than parallel operation, its circuits are much less complex.

Shift register	An element in the digital family that uses flip-flops to perform a displacement or movement of a set of digits one or more places to the right or left. If the digits are those of a numerical expression, a shift may be the equivalent of multiplying the number by a power of the base.
Skew	Time delay or offset between any two signals.
Synchronous timing	Operation of a switching network by a clock pulse generator. Slower and more critical than asynchronous timing but requires fewer and simpler circuits.
Word	An assemblage of bits considered as an entity in a computer. Typical digital circuits and their characteristics for the major logic families. (Adapted from Reference 4.)

Symbol	Circuit Diagram	Speed*	Power*	Fan-Out*	Noise Immunity*	Trade Name	Remarks
DCTL		Medium	Medium	Low	Low	Series 53	Variations in input characteristics result in base current "hogging" problems. Proper operation not always guaranteed. More susceptibile to noise because of low operating and signal voltages.
RTL		Low	Low	Low	Low	RTL	Very similar to DCTL. Resistors resolve current "hogging" problem and reduce power dissipation. However, operating speed is reduced.
RCTL		Low	Low	Low	Low	Series 51	Though capacitors can increase speed capability, noise immunity is affected by capacitive coupling of noise signals.
DTL		Medium	Medium	Medium	Medium to high	930 DTL	Use of pull-up resistor and charge-control technique improves speed capabilities. Many variations of this circuit exist, each having specific advantages.
TTL		High	Medium	Medium	Medium to high	SUHL Series 54/74	Very similar to DTL. Has lower parasitic capacity at inputs. With the many existing variations, this has become very
CML (ECL)		High	High	High	Medium to high	MECL ECCSL	Similar to a differential amplifier, the reference voltage sets the threshold voltage. High-speed, high-fan-out operation is possible with associated high power dissipation. Also known as *emitter-coupled logic* (ECL).
CTL		High	High	Medium	Medium	CTML	More difficult manufacturing process results in compromises of active device characteristics and higher cost.
I^2L		High	Low	High	Medium	I^2L	Provides smallest and most dense bipolar gate. Simple manufacturing process and higher component packing density than the MOS process. Also known as merged-transistor logic (MTL).

	Speed	Power	Fan-Out	Noise Immunity
*Low =	<5MHz	<5mW	<5	<300mV
Medium =	5–15MHz	5–15mW	5–10	300–500mV
High =	>15MHz	>15mW	>10	>500mV

Figure 16-81. Typical digital circuits and their characteristics for the major logic families. (Adapted from Reference 4.)

References

1. Bruce Carter, "A Differential Op-Amp Circuit Collection", Texas Instruments Application Report, July 2001.

2. B. Gilbert, "A precise Four Quadrant Multiplier with Subnanosecond Response," IEEE *J. of Solid State Circuits*, Vol. SC-3, no. 4, December 1968.

3. David Blackmer, "RMS Circuits with Bipolar Logarithmic Converter," United States Patent 3,681,618, August 1, 1972.

4. David Blackmer, "Multiplier Circuits," United States Patent 3,714,462, January 30, 1973.

5. "Solid State Music Dual Linear-Antilog Voltage Controlled Amplifier," SSM2000 Data Sheet, September, 1976.

6. Douglas Frey, "An Integrated Generalized Voltage Controlled Building Block", *J. Audio Eng Soc.*, Preprint No. 2403, November 1986.

7. Curtis Electro Music Website, http://curtiselectromusic.com/Doug_Curtis.html.

8. David Baskind, Harvey Rubens et al, "The Design and Integration of A High Performance Voltage Controlled Attenuator," *J. Audio Eng Soc.*, Preprint No. 1555, November 1979.

9. David Welland, "Compensation for VCA Op-Amp Errors," United States Patent 4,434,380, February 28, 1984

10. Analog Devices, "SSM2018T Trimless Voltage Controlled Amplifiers," SSM2018T Datasheet, July 2002.

11. Douglas Frey, "An Integrated Generalized Voltage Controlled Building

Block," *J. Audio Eng Soc.*, Preprint No. 2403, November 1986.

12. Douglas Frey, "Voltage-controlled element," United States Patent 4,471,320, September 11, 1984.

13. Douglas Frey, "Monolithic Voltage-Controlled Element," United States Patent 4,560,947, December 24, 1985.

14. Harvey Rubens, and David Baskind, "Voltage Controlled Attenuator," United States Patent 4,155,047, May 15, 1979.

15. "Solid State Music-Dual Linear-Antilog Voltage Controlled Amplifier," SSM2000 Data Sheet, September 1976.

16. IEEE Std 152-1991-IEEE Standard for Audio Program Level Measurement, June 22, 1992.

17. Roy Chapel, Jr. and Macit Gurol, "Thermally Isolated Monolithic Semiconductor Die," United States Patent 4,346,291, August 24, 1982.

18. Analog Devices, AD636 Datasheet.

19. Robert Adams, "Networks for the Log Domain," United States Patent 4,430,626, February 7, 1984.

20. THAT Corporation, "Adaptive Attack and Release Rates Using THAT Corporation RMS Detectors," Design Note DN-114, 2000.

21. Dean Jensen, "JE-990 Discrete Operational Amplifier," *J. Audio Eng Soc.*, January/February 1980.

22. Waltert Jung, *IC Op-Amp Cookbook*, No. 20969, Howard W. Sams, 1974.

23. www.jensentransformers.com, May 2008.

24. Gary Hebert and Frank Thomas, “The 48 Volt Phantom Menace,” Audio Engineering Society Convention Paper 5335, May 2001.

25. R. Bortoni & W. Kirkwood, “The 48V Phantom Menace Returns,” Audio Engineering Society Convention Paper 7909, 2009.

26. Robert Demrow, “Evolution From Operational Amplifier To Data Amplifier,” Analog Devices, 1968.

27. S. Wurcer, “A Programmable Instrumentation Amplifier for 12b Resolution Systems,” *International Solid State Circuits Conference*, February 1982. See also Analog Devices, Analog Dialog, Vol. 30 No. 2, 1996.

28. Harrison Systems, PC1041 Microphone Preamp Schematic, 1978.

29. G. J. Cohen, “Double Balanced Microphone Amplifier,” Audio Engineering Society Preprint 2106, September 1984.

30. Solid State Music, SSM2011 Datasheet, 1982.

31. D. Bowers, “An Ultra-Low-Noise Monolithic Microphone Preamplifier,” Audio Engineering Society Preprint 2495, 1987.

32. Les Tyler, “De-Integrating IC Preamps,” Presentation at the 131st Audio Engineering Society convention, available from THAT Corporation.

33. Gary Hebert, “Designing Microphone Preamplifiers,” Presentation at the 129th Audio Engineering Society convention, available from THAT Corporation.

34. Bill Whitlock, “A New Balanced Audio Input Circuit For Maximum Common-Mode Rejection In Real-World Environments,” www.jensen-transformers.com. 1996.

35. Thomas Hay, "Differential Technology In Recording Consoles And The Impact Of Transformerless Circuitry On Grounding Technique," Audio Engineering Society Preprint 1723, October 1980.

36. Bill Whitlock, "Differential Line Receiver With Common-Mode Bootstrapping," U.S. patent 5,568,561, October 22, 1996.

37. Chris Strahm, "Balanced Line Output Circuit," United States Patent 4,979,218, December, 18, 1990.

38. Gary Hebert, "Floating, Balanced Output Circuit" United States Patent 6,316,970, November 13, 2001.

Chapter 17

Heatsinks and Relays

by Glen Ballou

17.1 Heatsinks

*17.1.1 Thermal Management of Today's Audio Systems**

Today's audio systems, like all electronic systems are being powered by smaller devices, packaged in smaller systems that are generating more heat. We need to increase our level of understanding on all of the latest techniques for the management of this added heat in as effective a means as possible. Let's first start with the understanding of the three methods of heat transfer—specifically, convection, conduction, and radiation as all three methods of heat transfer contribute to the complete thermal management provided by the heat-sinks installed in an audio system.

17.1.1.1 Convection

Convection is the transfer of heat from a solid surface to the surrounding gas, which is always air in the case of a typical audio system. This method of heat transfer drives the amount of required fin surface area that is accessible to the surrounding air so that it may heat up the surrounding air, allow it to move away, and make room for the process to repeat itself. This process can be greatly accelerated with the use of a fan to provide more energy to the moving of the air than just the natural buoyant force of the heated air.

Natural convection is when there is no external fan and the heat transfer occurs with very low air flow rates, typically as low as 35 linear feet per minute (lfm) for obstructed natural convection to 75 lfm for optimum unobstructed vertical natural convection. Natural convection is never zero air flow rate because without air movement there would be no heat transfer. Think of the closed cell plastic foam insulation. It works as an insulator because the closed cell prevents the air from moving away.

Forced convection is when a system fan imparts a velocity to the air surrounding the heatsink fins. The fan may be physically attached to the convective fin surface area of the heatsink to increase the air velocity over the fin surfaces. There is impingement flow—fan blows down from on top of the fins—and through flow—fan blows from the side across the fin set.

Forced convection thermal systems are most generally significantly smaller (50% or more) than their natural convection equivalents. The penalties for the smaller size are the added power to operate the fan, an added failure mechanism, the added cost, and the noise from the fan. Fan noise is probably the most important consideration when applying them in audio systems.

17.1.1.2 Conduction

Conduction is the transfer of heat from one solid to the next adjacent solid. The amount and thermal gradient of heat transfer are dependent on the surface finishes—flatness and roughness—and the interfacial pressure generated by the attachment system. This mechanically generated force is accomplished by screws, springs, snap assemblies, etc. The thermal effectiveness of a conductive interface is measured by the resultant temperature gradient in °C. This may be calculated from the interface thermal resistance at the mounted pressure times the watts of energy moving across the joint divided by the cross-sectional area. These temperature gradients are most significant for high wattage components in small packages —divisors less than 1.0 are, in actual effect, multipliers. Good thermal solutions have attachment systems that generate pressures of 25–50psi.

Table 17-1 compares the thermal performance of most of the common interface material groups with a dry joint—this makes amply clear why it is never acceptable to specify or default through design inaction to a dry joint.

Table 17-1. Thermal Performance of Common Interface Materials

Interface Material Group	Thermal Performance Range in °C in^2/W	Comments
Dry Mating Surfaces	3.0-12.0	Too much uncertainty to use. Too big a thermal gradient
Gap Fillers	0.4-4.0	Minimize thickness required Spring mechanical load
Electrically Insulating	0.2- 1.5	Maximize mechanical loads
High Performance Pads	0.09-0.35	Minimize thickness Maximize mechanical loads
Phase Change Pads	0.02-0.14	Must follow application method Spring mechanical load
Low	0.04-0.16	Screen apply Spring mechanical

Performance Grease		load
High Performance Grease	0.009-0.04	Must Screen apply Spring mechanical load Best at high loads (>50psi)

The primary heat transfer driving force is the temperature difference between $T_{maxcase}$ and $T_{maxamblent}$ modified by the conductive ΔT losses in the interface and any extraordinary hot spot offsets and spreading losses. If heatsinks are mounted over spread hot spots these last conductive losses are not sufficiently large to consider. They really only become significant when considering very unusual arrangements—high-watt density loads such as typical LEDs.

17.1.1.3 Radiation

Radiation is the third and least important method of heat transfer for audio system heatsinks. Radiation has a maximum 20–25% impact in natural convection applications with a negligible impact after 200lfm applications. Radiation is a function of the fourth power of the absolute temperature difference between the hot side and surrounding cooler surfaces that look at each other and their respective emissivities. In the real world in which we live, these are not significant enough to suggest a lot of effort to understand and optimize.

Aluminum extruded heatsinks were typically made black in an anodizing process at a significant cost to get an emissivity of 0.95 (dimensionless). The typical aluminum surface forms an oxide film in less than a second after machining with an emissivity of about 0.30–0.40. Nominally, almost half the benefit will come free so that our advice is “Leave the radiation effects alone.” What you get beneficially you were going to largely get anyway, free from Mother Nature. In review, heatsinks use all three methods of heat transfer to produce the desired effect of cooling the typical electronic component in the typical audio system.

17.1.1.4 Summary

Convection is usually the most significant method, and it depends on having sufficient fin surface area in direct contact with the surrounding air and design features to minimize the insulating effects of boundary films. Aerodynamic shapes and adequate open fin spacing that allows free air movement are critical design issues.

Conduction is the first step in the heat transfer chain in that conduction transfers the heat from the device into the heatsink, then through the heatsink to the fin surface where convection takes over. Some heatsinks need conduction enhancements such as heat pipes to keep the conduction temperature gradients to a value that is low enough to allow the convection to complete the heat transfer without exceeding the application temperature limits.

Radiation is a secondary level effect that is always present, marginally significant in natural convection, but not economical to control.

17.1.2 New Technologies to Make Things Fit More Easily

The range of technologies, materials, and fabrication processes available to the thermal designer today is quite impressive. The primary goal when employing these advanced technologies, materials, and fabrication processes is to increase the effective density of the of the resultant heat transfer system. Technically, we are increasing the volumetric efficiency of the thermal solution proposed for a given application. In “man speak” the required heatsink gets much smaller in size and therefore fits more easily into the ever-shrinking product envelope. A smaller heatsink has a decreased conductive thermal spreading resistance and therefore a smaller conductive temperature gradient. As an example, we have a convective solution defined for a baseline heatsink. The baseline heatsink is fabricated from an extruded aluminum alloy (6063-T5). The following describes a technology, material or fabrication process and give a volume ratio or range of volume ratios that can be applied to the existing solution to quickly see the benefit of applying this technology, material, or fabrication process to the audio application at hand. Ratios that are less than 1.00 are indicating a reduction in heatsink volume.

Thermal solution problem solving is an iterative process balancing the application boundary specifications against the affordable technologies/materials/fabrication processes until a system compromise solution is defined.[1] For example; marketing has directed that only a natural convection solution is acceptable but the heatsink is too big. One solution might require the $T_{maxambient}$ be reduced by 5°C and the heatsink be fabricated from copper, C110 soldered together. This could reduce the size of the heatsink by 25–35%. The penalties would be the weight would increase between two and three times and the unit cost of the heatsink increase by three to four times. There are software systems[2] that specialize in defining these trade-offs rapidly, allowing a real-time compromise to be made, even during the design review meeting with marketing.

Table 17-2 summarizes the thermal solution benefits possible with the proper application of new technologies, materials and fabrication processes.

Table 17-2. Thermal Solutions with New Technology, Materials, and Fabrication Processes

Technology/ Material/ Fabrication	Title	Volumetric Ratio Range	Cost Range	Comments
M	Copper C110	0.8	3.5 ×	Volumetric ratios are even lower for conduction limited applications. Weight almost triples (3 ×)
MF	Molded Plastic Conductive Dielectric Elastomeric	0.97 (<200LFM) 1.03 (>200LFM) 1.07 (>500LFM)	0.5–0.7 After tooling if not a standard	Saves weight and finishing[3] Hybrids—molded base metal fins
TM	Base-Mounted Heat Pipes[3]	1 Heat Pipe 0.79 2 Heat Pipe 0.73	1.5–2.0	Aluminum base
TM	Base-Mounted Heat Pipes[3]	1 Heat Pipe 0.71 2 Heat Pipe 0.66	4.0–4.5	Copper base
TM	Base-Mounted Vapor Chamber[3]	0.69 Al 0.58 Cu	2.5 Al 4.8 Cu	Achieves optimum spreading
M	Graphite	0.72	6–8 ×	Relatively fragile 35% reduction in weight
TM	Solid-State Heat Pipes (TPG)[4]	0.75	3.6 Al 5.4 Cu	Eliminates burnout as a failure mode
FM	Bonded Fin and Folded Fin	0.90 Al 0.76 Cu	2 × 3.6 ×	More convective fin surface per unit volume fin shapes break up boundary film layers for performance gains

Extruded heatsinks have fin thicknesses that are much greater, thicker than required thermally. They are thicker to accommodate the strength requirements of the die, which is close to the melting point of aluminum during the extrusion process.

Bonded fin and folded fin heatsink designs use sheet stock for the fins so that they may be optimally sized as required to carry the thermal load without regard to the mechanical requirements of the extrusion process. These heatsinks can, therefore, without compromising the required open fin spacing, have a greater number of fins and be convectively much more volumetrically effective. These sheet metal fins are attached to the heatsink bases with either thermal epoxy adhesives or soldered. Since this joint only represents ~3% of the total thermal resistance of the heatsink, the adhesive choice is never critical.

Air flow management is the most critical parameter to control in optimizing the convective heat transfer for any thermal solution. Baffles, shrouds, and fan sizing are all very critical in making the most of the convective portion of the heat transfer thermal solution. An example is an audio amplifier with two rows of very hot components. With two facing extrusions that formed a box shape, a fan was mounted at the end and blew the air down the chute with great success. The air flow was fully contained and no leakage occurred. And so the audio cooling tube was born.

There should always be a space, 0.5–0.8in along the axis of the fan, between the fan outlet and the fin set that is fully shrouded to force the air to pass over the convective fin surfaces. This is called the *plenum*. Its function is to allow the upstream pressure generated by the fan to reach an equilibrium and thereby equalize the air flow through each fin opening.

Audio systems that require fans need to be carefully designed to have an air flow path that is well defined so that the fan may be operated at a minimal speed. This results in the fan generating a minimum of noise. High-velocity fans are noisy. Noise abatement is very expensive and seldom truly satisfactory, therefore, the best solution is to minimize the fan generated noise.

17.1.3 How Heatsinks Work

Heatsinks are used to remove heat from a device, often the semiconductor junction. To remove heat, there must be a temperature differential (ΔT) between the junction and the air. For this reason, heat removal is always after the fact. Unfortunately, there is also resistance to heat transfer between the junction and its case, any insulating material, the heatsink, and the air, Fig. 17-1.

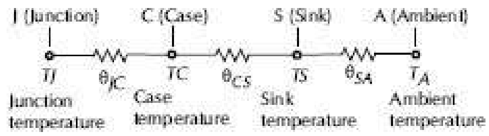


Figure 17-1. Series thermal resistance/temperature circuit.

17.1.3.1 Thermal Resistance

The total thermal resistance between the junction and the air is the sum of the individual thermal resistances

$$\Sigma\theta = \theta_{JC} + \theta_{CI} + \theta_{IS} + \theta_{SA} \qquad (17\text{-}1)$$

where,

θ is the thermal resistance in degrees Celsius per watt (°C/W),

JC is the junction to case,

CI is the case to insulator,

IS is the insulator to heatsink,

SA is the heatsink to air.

The temperature at the junction can be determined from the ambient temperature, the thermal resistance between the air and the junction, and the power dissipated at the junction.

$$T_J = T_A + \theta_{JA} P_D \qquad (17\text{-}2)$$

where,

T_A is the temperature of the air,

θ_{JA} is the thermal resistance from the air to the junction,

P_D is the power dissipated.

If the junction temperature was known, then the power dissipated at the junction can be determined

$$P_D = \frac{\Delta T}{\Sigma\theta} \qquad (17\text{-}3)$$

where,

ΔT is $T_J - T_A$.

17.1.3.2 Heatsink Materials and Design

Heatsinks[5] are generally made from extruded aluminum or copper and are painted black, except for the areas in which the heat-producing device is mounted. The size of heatsinks will vary with the amount of heat to be radiated and the ambient temperature and the maximum average forward current through the element. Several different types of heatsinks are pictured in Fig. 17-2. The rate of heat flow from an object is

$$Q = \frac{KA\Delta T}{L} \qquad (17\text{-}4)$$

where,

Q is the rate of heat flow,

K is the thermal conductivity of material,

A is the cross-sectional area,

ΔT is the temperature difference,

L is the length of heat flow.

A. Small heat sinks used with diodes and transistors. Their diameter is less than a dime.

B. Large heat sink for use with heavy current rectifiers. The stud of the rectifier is screwed into the center fin of the sink.

Figure 17-2. Conduction type heatsinks used for cooling diodes and transistors. Courtesy Wakefield Engineering Co.

For best conduction of heat, the material should have a high thermal conductivity and have a large cross-sectional area. The ambient or material temperature should be maintained as low as possible, and the thermal path should be short. The amount of heat (energy) radiated by a body is dependent on its surface area, temperature, and emissivity. For best results, the heatsink should:

- Have maximum surface area/volume (hence the use of vertical fins).
- Be made of a high thermal conductivity material.
- Have material of high emissivity (painted aluminum or copper).
- Have proper ventilation and location (should be below, not above, other heat radiators).

- Be placed so that the lowest power device is below the higher power devices, and all devices should be as low as possible on the heatsink.

The overall effectiveness of a heatsink is dependent to a great extent on the intimacy of the contact between the device to be cooled and the surface of the heatsink. Intimacy between these two is a function of the degree of conformity between the two surfaces and the amount of pressure that holds them together. The application of a silicone oil to the two surfaces will help to minimize air gaps between the surfaces, improving conduction. The use of a mica washer between the base of the device to be cooled and the heatsink will add as much as 0.5°C/W to the thermal resistance of the combination. Therefore, it is recommended that (whenever possible) an insulating washer be used to insulate the entire heat-sink from the chassis to which it is to be mounted. This permits the solid-state device to be mounted directly to the surface of the heatsink (without the mica washer). In this way, the thermal resistance of the mica washer is avoided.

A typical heatsink is shown in Fig. 17-3. The heat-sink has 165 in^2 of radiating surface. The graph in Fig. 17-4 shows the thermal characteristics of a heatsink with a transistor mounted directly on its surface. A silicone oil is used to increase the heat transfer. This graph was made with the heatsink fins in a vertical plane, with air flowing from convection only. Fig. 17-5 shows the effect of thermal resistance with forced air blown along the length of the fin.

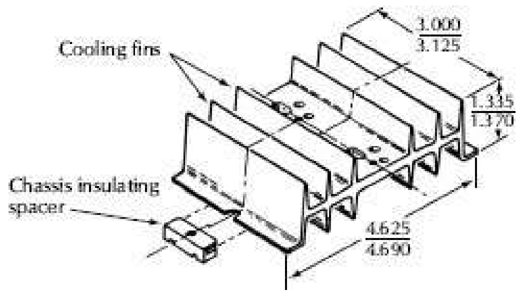


Figure 17-3. Typical heatsink for mounting two transistors. Courtesy Delco Electronics Corp.

Thermally conductive adhesives can also be used. These adhesives offer a high heat transfer, low shrinkage, and a coefficient of thermal expansion comparable to copper and aluminum.

The thermal capacity of a cooling fin or heatsink must be large compared to the thermal capacity of the device and have good thermal conductivity across its entire area. The specific thermal resistance ρ of interface materials used for heatsinks and insulating devices is shown in Table 17-3.

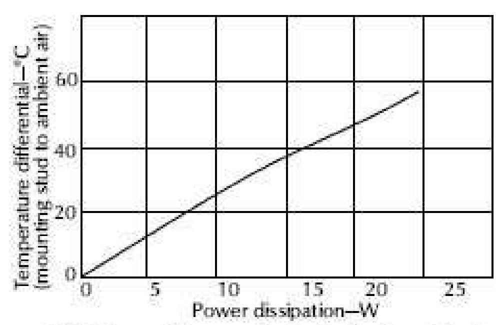


Figure 17-4. Thermal characteristics for the heatsink shown in Fig. 17-3, with conventional air cooling. Courtesy of Delco Electronics Corp.

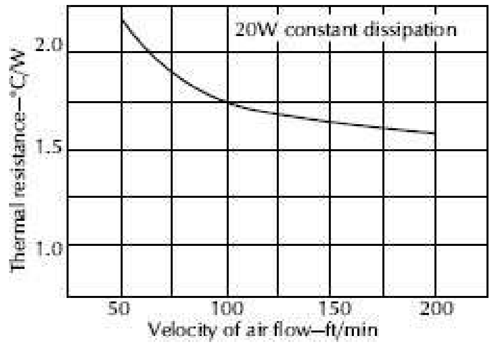


Figure 17-5. Thermal characteristics for the heatsink shown in Fig. 17-3, with forced air cooling. Courtesy of Delco Electronics Corp.

Table 17-3. Specific Thermal Resistance *p* of Interface Materials, °C in/W

Material	ρ
Still air	1200
Mylar film	236
Silicone grease	204
Mica	66
Wakefield Type 120 Compound	56
Wakefield Delta Bond 152	47
Anodize	5.6
Aluminum	0.19
Copper	0.10

Courtesy of Wakefield Engineering, Inc.

The thermal resistance θ for these materials can be determined by the equation

$$\theta = \frac{\rho t}{A} \qquad (17\text{-}5)$$

where,

ρ is the specific thermal resistance,

t is the material thickness in in,

A is the area in in^2.

For instance, a square copper plate 4in (10cm) per side and 0.125in (3.2mm) thick would have a θ of 0.00078°C/W, while a mica insulator 0.003 in (0.076mm) thick with a diameter of 1in (25.4mm) would have a θ of 0.25°C/W. If the semiconductor dissipates 100W, the temperature drop across the copper plate would be 0.07°C (0.13°F) and across the mica washer it would be 25°C (45°F). In transistor replacement in older equipment, it would be best to replace the mica insulator with a new type of insulator.

In the selection of a heatsink material, the thermal conductivity of the material must be considered. This determines the thickness required to eliminate thermal gradients and the resultant reduction in emissivity. An aluminum fin must be twice as thick as a comparable copper fin, and steel must be eight times as thick as copper.

Except for the smallest low-current solid-state devices, most devices must use a heatsink, either built in or external.

Space for heatsinks is generally limited, so the minimum surface area permissible may be approximately calculated for a flat aluminum heat plate by

$$A = 133\frac{W}{\Delta T}\text{in}^2 \qquad (17\text{-}6)$$

where,

W is the power dissipated by the device,

ΔT is the temperature differences between the ambient and case temperature in °C.

The approximate wattage dissipated by the device can be calculated from the load current and the voltage drop across it

$$W = I_L V_D \qquad (17\text{-}7)$$

where,

I_L is the load current,

V_D is the voltage drop across the device.

For a triac, V_D is about 1.5V; for SCRs, about 0.7V. For transistors it could be from 0.7V to more than 100V.

For example, to determine the minimum surface area required for a flat aluminum heatsink to keep the case temperature of 75°C (167°F) for a triac while delivering a load current of 15 A, at 25°C (77°F) ambient and a voltage drop across the triac of 1.5V use the equation

$$\begin{aligned} \Delta T &= T_{case} - T_{ambient} \\ &= 75°C - 25°C \\ &= 50°C \end{aligned}$$

Using Eq. 17-7

$$\begin{aligned} W &= V_D I_L \\ &= 1.5 \times 15 \\ &= 22.5\ \text{W} \end{aligned}$$

Using Eq. 17-6

$$\begin{aligned} A &= 133 \frac{W}{\Delta T} \text{in}^2 \\ &= 133 \times \frac{22.5}{50} \\ &= 59.85\ \text{in}^2 \end{aligned}$$

It is important that the case temperature, T_{case}, does not exceed the maximum allowed for a given load current, I_L (see typical derating curves in Fig. 17-6).

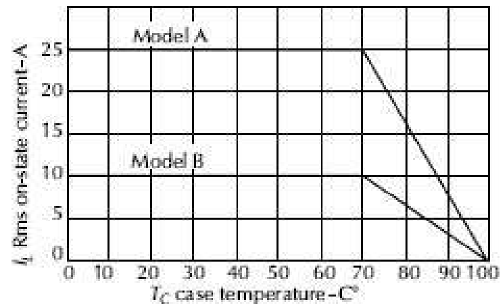


Figure 17-6. A typical derating curve for solid state devices.

Eq. 17-6 gives the surface area needed for a vertically mounted heatsink. With free air convection, a vertically mounted heatsink, Fig. 17-7, has a thermal resistance approximately 30% lower than with horizontal mounting.

In restricted areas, forced-convection cooling may be necessary to reduce the effective thermal resistance of the heatsink. When forced air cooling is used to cool the component, the cubic feet per minute (cfm or ft^3/min) required is determined by

$$\begin{aligned} cmf &= \frac{Btu/h}{60} \times 0.02 \text{ temperature rise} \\ &= \frac{1.76Q}{\Delta TK} \end{aligned} \qquad (17\text{-}8)$$

where,

1W is 3.4Btu,

temperature rise is in °C,

Q is the heat dissipated in W,

ΔT is the heatsink mounting temperature minus the ambient temperature,

K is the coupling efficiency (0.2 for wide spaced fins, 0.6 for close spaced fins).

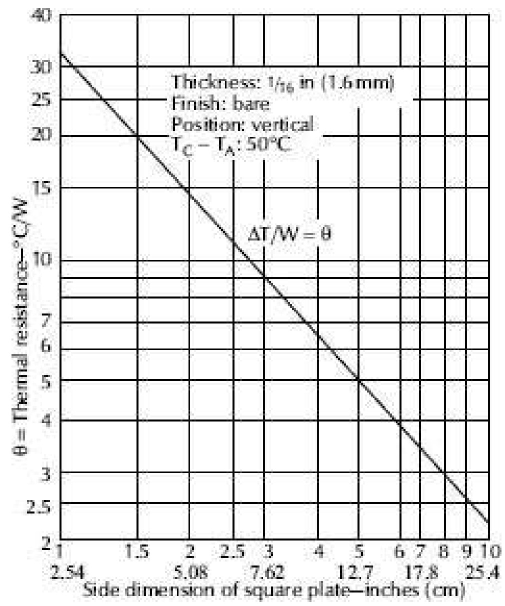


Figure 17-7. Thermal resistance for a vertically mounted 1⁄16 in (1.6mm) aluminum plate of various dimensions.

17.1.3.3 Monitoring Heatsink and Pipe Temperature

High-power electronic components require effective heat removal and thermal management to ensure their performance and reliability by using heatsinks of different materials and various shapes to conduct heat away from the components. Often the heatsinks and pipes are over-designed to account for unexpected environmental inputs. For demanding applications, monitoring the heatsink temperature and implementing a feedback control is required.

Usually a thermistor or another type of direct-contact temperature sensor is mounted to the heatsink and wires are run to connect it to the printed circuit board (PCB) where the measurement and control units are located. This complicates manufacturing, and the process often involves manual assembly.

Measuring temperatures using infrared (IR) sensors without direct contact with an object is a way to optimize thermal monitoring and control. The thermal management of high-power electronic components is an area that can benefit significantly. New thermopile sensor technology has a significant reduction in size and power consumption. Digital processing allows the sensor to be placed in space-constrained locations and operate independently of an embedded controller or microprocessor in calculating the object temperature from the IR radiation. The sensor can be mounted directly on the circuit board with standard surface-mount technology, allowing it to provide information for both the local temperature of the PCB on which it is mounted through its on-die temperature sensor, and monitor the temperature of the heat sink directly in front of it by measuring its IR radiation and computing its corresponding temperature.

Fig. 17-8 illustrates sensor placement. The sensor's field of view (straight dash lines) is completely encompassed by the heat sink, an important requirement to ensure that only the heatsink's IR radiation (and temperature) is picked up by the sensor and not the surrounding components or case walls. Some sensors allow temperature limits to be stored in their built-in memory to compare them with the heatsink's temperature to control a shutdown. In this way, the system takes action only in the case of an over- or under-temperature event. The controller or processor is offloaded from having to poll the sensor for measurement data while the temperature is in the acceptable range.

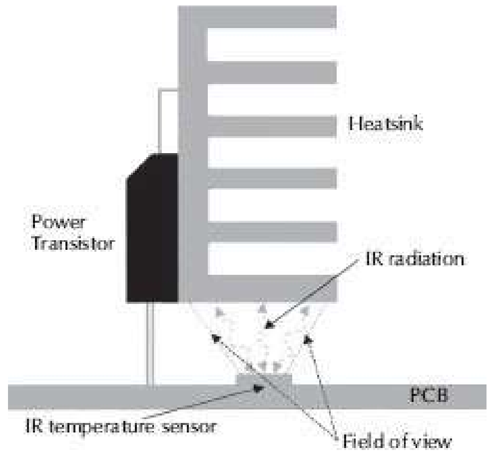


Figure 17-8. IR temperature measurement setup.

17.2 Relays

A *relay* is an electrically operated switch connected to or actuated by a remote circuit. The relay causes a second circuit or group of circuits to operate. The relay may control many different types of circuits connected to it. These circuits may consist of motors, bells, lights, audio circuits, power supplies, and so on, or the relay may be used to switch a number of other circuits at the same time or in sequence from one input.

Relays may be electromechanical or solid state. Both have advantages and disadvantages. Only a few years ago relays were big and cumbersome and required either an octal-type socket or were externally wired. Today relays are very compact and come in many layouts. A few are given below.

Solder Connectors. Connectors vary in size and spacing, depending on the current carrying capacity.

Octal Sockets. Plug into standard 8 pin and 11 pin sockets.

Rectangular Sockets. Plug into a 10 pin, 11 pin, 12 pin, 14 pin, 16 pin, 22 pin, or 28 pin socket.

DIP Relays. Designed to mount directly on a printed circuit board on a 0.1in (2.54mm) spacing. Sockets can be 8 pin or 16 pin.

SIP 4 Pin Relay. Plug into a SIP socket or mount on a printed circuit board on a 0.2 in (5.08 mm) in-line spacing.

17.2.1 Glossary of Terms

This glossary was compiled from NARM Standard RS-436, MIL STD 202, and MIL STD R5757.

Actuate Time. The time measured from coil energization to the stable contact closure (Form A) or stable contact opening (Form B) of the contact under test. (See also Operate Time.)

Ampere Turns (AT). The product of the number of turns in an electromagnetic coil winding and the current in amperes passing through the winding.

Bandwidth. The frequency at which the RF power insertion loss of a relay is 50%, or −3dB.

Bias, Magnetic. A steady magnetic field applied to the magnetic circuit of a switch to aid or impede its operation in relation to the coil's magnetic field.

Bounce, Contact. Intermittent and undesired opening of closed contacts or closing of opened contacts usually occurring during operate or release transition.

Breakdown Voltage. The maximum voltage that can be applied across the open switch contacts before electrical breakdown occurs. In reed relays it is primarily dependent on the gap between the reed switch contacts and the type of gas fill used. High AT switches within a given switch family have larger gaps and higher breakdown voltage. It is also affected by the shape of the contacts, since pitting or whiskering of the contact surfaces can develop regions of high electric field gradient that promote electron emission and avalanche breakdown. Since such pitting can be asymmetric, breakdown voltage tests should be performed with forward and reverse polarity. When testing bare switches, ambient light can affect the point of avalanche and should be controlled or eliminated for consistent testing. Breakdown voltage measurements can be used to detect reed switch capsule damage. See Paschen Test.

Carry Current. The maximum continuous current that can be carried by a closed relay without exceeding its rating.

Coaxial Shield. Copper alloy material that is terminated to two pins of a reed relay within the relay on each side of the switch. Used to simulate the outer conductor of a coaxial cable for high-frequency transmission.

Coil. An assembly consisting of one or more turns of wire around a common form. In reed relays, current applied to this winding generates a magnetic field that operates the reed switch.

Coil AT. The coil *ampere turns* (AT) is the product of the current flowing through the coil (and therefore directly related to coil power) and the number of turns. The coil AT exceeds the switch AT by an appropriate design margin to ensure reliable switch closure and adequate switch overdrive. Sometimes abbreviated as *NI*, where N is the number of turns and I is the coil current.

Coil Power. The product, in watts, of the relay's nominal voltage and current drawn at that voltage.

Cold Switching. A circuit design that ensures the relay contacts are fully closed before the switched load is applied. It must take into account bounce, operate and release time. If technically feasible, cold switching is the best method for maximizing contact life at higher loads.

Contact. The ferromagnetic blades of a switch often plated with rhodium, ruthenium, or tungsten material.

Contact Resistance, Dynamic. Variation in contact resistance during the period in which contacts are in motion after closing.

Contact Resistance, Static. The dc resistance of closed contacts as measured at their associated contact terminals. Measurement is made after stable contact closure is achieved.

Crosstalk (Crosstalk Coupling). When applied to multichannel relays, the ratio, expressed in dB, of the signal power being emitted from a relay output contact to the power being applied to an adjacent input channel at a specified frequency.

Duty Cycle. A ratio of energized to de-energized time.

Electrostatic Shield. Copper alloy material terminated to one pin within the reed relay. Used to minimize coupling and electrostatic noise between the coil and contacts.

Form-A. Contact configuration that has one single pole–single throw normally open (SPST n.o.) contact.

Form-B. Contact configuration that has one single pole–single throw normally closed (SPST n.c.) contact.

Form-C. Contact configuration that has one single pole–double throw (SPDT) contact. (One common point connected to one normally open and one normally closed contact.) Sometimes referred to as a transfer contact.

Hard Failure. Permanent failure of the contact being tested.

Hermetic Seal. An enclosure that is sealed by fusion to ensure a low rate of gas leakage. In a reed switch, a glass-to-metal seal is employed.

Hot Switching. A circuit design that applies the switched load to the switch contacts at the time of opening and closure.

Hysteresis. When applied to reed relays, the difference between the electrical power required to initially close the relay and the power required to just maintain it in a closed state. (Usually expressed in terms of the relay's pull-in voltage and drop-out voltage.) Some degree of hysteresis is desirable to prevent chatter and is also an indicator of adequate switch contact force.

Impedance (Z). The combined dc resistance and ac reactance of a relay, at a specified frequency and found with the equation

$$Z = R + jX \qquad (17\text{-}9)$$

where,

R is the dc resistance,

X is $2\pi fL - \dfrac{1}{2\pi fC}$,

f is the frequency.

Because of the small residual capacitance across the open contacts of a reed relay, the impedance decreases at higher frequencies, resulting in lower isolation at higher frequencies. Conversely, increasing inductive reactance at higher frequencies causes the impedance of a closed relay to rise, increasing the insertion loss at higher frequencies.

Impedance Discontinuity. A deviation from the nominal RF impedance of 50Ω at a point inside a reed relay. Impedance discontinuities cause signal absorption and reflectance problems resulting in higher signal losses. They are minimized by designing the relay to have ideal transmission line characteristics.

Insertion Loss. The ratio of the power delivered from an ac source to a load via a relay with closed contacts, compared to the power delivered directly, at a specified frequency, and is found with the equation

$$Insertion\ Loss = -20\log\frac{V_t}{V_i} \qquad (17\text{-}10)$$

where,

V_t is the transmitted voltage,

V_i is the incident voltage.

Insertion loss, isolation and return loss are often expressed with the sign reversed; for example, the frequency at which 50% power loss occurs may be quoted as the −3dB point. Since relays are passive and always produce net losses, this does not normally cause confusion.

Inrush Current. Generally, the current waveform immediately after a load is connected to a source. Inrush current can form a surge flowing through a relay that is switching a low-impedance source load that is typically a highly reactive circuit or one with a nonlinear load characteristic such as a tungsten lamp load. Such abusive load surges are sometimes encountered when relays are inadvertently connected to test loads containing undischarged capacitors or to long transmission lines with appreciable amounts of stored capacitive energy. Excessive inrush currents can cause switch contact welding or premature contact failure.

Insulation Resistance. The dc resistance between two specified test points.

Isolation. The ratio of the power delivered from a source to a load via a relay with open contacts, compared to the power delivered directly, at a specified frequency. If V_i is the incident voltage and V_t is the transmitted voltage, then isolation can be expressed in decibel format as

$$Isolation = -20\log\frac{V_t}{V_i} \tag{17-11}$$

where,

V_t is the transmitted voltage,

V_i is the incident voltage.

Latching Relay. A bistable relay, typically with two coils, that requires a voltage pulse to change state. When pulse is removed from the coil, the relay stays in the state in which it was last set.

Life Expectancy. The average number of cycles that a relay will achieve under specified load conditions before the contacts fail due to sticking, missing or excessive contact resistance. Expressed as mean cycles before failure (MCBF).

Low Thermal Emf Relay. A relay designed specifically for switching low-voltage level signals such as thermocouples. These types of relays use a thermally compensating ceramic chip to minimize the thermal offset voltage generated by the relay.

Magnetic Interaction. The tendency of a relay to be influenced by the magnetic field from an adjacent energized relay. This influence can result in depression or elevation of the pull-in and dropout voltage of the affected relay, possibly causing them to fall outside their specification. Magnetic interaction can be minimized by alternating the polarity of adjacent relay coils, by magnetic shielding, or by placing two relays at right angles to each other.

Magnetic Shield. A ferromagnetic material used to minimize magnetic coupling between a relay and external magnetic fields.

Mercury Wetted Contact. A form of reed switch in which the reeds and contacts are wetted by a film of mercury obtained by a capillary action from a mercury pool encapsulated within the reed switch. The switch in this type of relay must be mounted vertically to ensure proper operation.

Missing (Contacts). A reed switch failure mechanism, whereby an open contact fails to close by a specified time after relay energization.

Nominal Voltage. The normal operating voltage of the relay.

Operate Time. The time value measured from the energization of the coil to the first contact closure, Form A, or the first contact open, Form B.

Operate Voltage. The coil voltage measured at which a contact changes state from its unenergized state.

Overdrive. The fraction or percentage by which the voltage applied to the coil of a relay exceeds its pull-in voltage. An overdrive of at least 25% ensures adequate closed contact force and well-controlled bounce times, which result in optimum contact life. For instance, Coto Technology's relays are designed for a minimum of 36% overdrive so a relay with a nominal coil voltage of 5V will pull in at no greater than 3.75V.

When using reed relays, the overdrive applied to the relay should not drop below 25% under field conditions. Issues such as power supply droop and voltage drops across relay drivers can cause a nominally acceptable power supply voltage to drop to a level where adequate overdrive is not maintained.

Release Time. The time value measured from coil de-energization to the time of the contact opening, Form-A or first contact closure, Form-B.

Release Voltage. The coil voltage measured at which the contact returns to its de-energized state.

Return Loss. The ratio of the power reflected from a relay to that incident on the relay, at a specified frequency and can be found with the equation

$$Return\ loss = -20\log\frac{V_r}{V_i} \qquad (17\text{-}12)$$

where,

V_r is the reflected voltage,

V_i is the incident voltage.

Signal Rise Time. the time required for its output signal to rise from 10–90% of its final value, when the input is changed abruptly by a step function signal.

Shield, Coaxial. A conductive metallic sheath surrounding a reed relay's reed switch, appropriately connected to external pins by multiple internal connections, and designed to preserve a 50Ω impedance environment within the relay. Used in relays designed for high-frequency service to minimize impedance discontinuities.

Shield, Electrostatic. A conductive metallic sheath surrounding a reed relay's reed switch, connected to at least one external relay pin, and designed to minimize capacitive coupling between the switch and other relay components, thus reducing high-frequency noise pickup. It is similar to a coaxial shield, but not designed to maintain a 50Ω RF impedance environment.

Shield, Magnetic. An optional plate or shell constructed of magnetically permeable material such as nickel-iron or mu-metal, fitted external to the relay's coil. Its function is to reduce the effects of magnetic interaction between adjacent relays and to improve the efficiency of the relay coil. A magnetic shell also reduces the influence of external magnetic fields, which is useful in security applications. Magnetic shields can be fitted externally or may be buried inside the relay housing.

Soft Failure. Intermittent self-recovering failure of a contact.

Sticking (Contacts). A switch failure mechanism, whereby a closed contact fails to open by a specified time after relay de-energization. Can be subclassified as hard or soft failures.

Switch AT. The ampere turns required to close a reed switch, pull-in AT, or just to maintain it closed, drop-out AT, and is specified with a specific type and design of coil. Switch AT depends on the length of the switch leads and increases when the reed switch leads are cropped. This must be taken into account when specifying a switch for a particular application.

Switching Current. The maximum current that can be hot-switched by a relay at a specified voltage without exceeding its rating.

Switching Voltage. The maximum voltage that can be hot-switched by a relay at a specified current without exceeding its rating. Generally lower than breakdown voltage, since it has to allow for any possible arcing at the time of contact breaking.

Transmission Line. In relay terms an interruptible waveguide consisting of two or more conductors, designed to have a well-controlled characteristic RF impedance and to efficiently transmit RF power from source to load with minimum losses, or to block RF energy with minimum leakage. Structures useful within RF relays include microstrips, coplanar waveguides, and coaxial transmission line elements.

VSWR (Voltage Standing Wave Ratio). The ratio of the maximum RF voltage in a relay to the minimum voltage at a specified frequency and calculated from

$$VSWR = (1+\rho)/(1-\rho) \tag{17-13}$$

where,

ρ is the voltage reflected back from a closed relay terminated at its output with a standard reference impedance, normally 50Ω.

17.2.2 Contact Characteristics

Contacts may switch either power or dry circuits. A power circuit always has current flowing, while a dry circuit has minimal or no current flowing, such as an audio circuit. A dry or low-level circuit typically is less than 100mV or 1mA.[6,7,8]

The mechanical design of the contact springs is such that when the contacts are closed, they slide for a short distance over the surfaces of each other before coming to rest. This is called a *wiping contact*, and it ensures good electrical contact.

Contacts are made of silver, palladium, rhodium, or gold and may be smooth or bifurcated. Bifurcated contacts have better wiping and cleaning action than smooth contacts and, therefore, are used on dry circuits.

There are various combinations of contact springs making up the circuits that are operated by the action of the relay. Typical spring piles are shown in Fig. 17-9.

As contacts close, the initial resistance is relatively high, and any films, oxides, and so on further increase the contact resistance. Upon closing, current begins to flow across the rough surface of the contacts, heating and softening them until the entire contact is mating, which reduces the contact resistance to milliohms. When the current through the circuit is too low to heat and soften the contacts, gold contacts should be used since the contacts do not oxidize and, therefore, have low contact resistance. On the other hand, gold should not be used in power circuits where current is flowing.

The contact current specified is the maximum current, often the make-or-break current. For instance, the make current of a motor or capacitor may be 10–15 times as high as its steady-state operation. Silver cadmium oxide contacts are very common for this type of load. The contact voltage specified is the maximum voltage allowed during arcing during break. The break voltage of an inductor can be 50 times the steady-state voltage of the circuit.

To protect the relay contacts from high transient voltages, arc suppression should be used. For dc loads, this may be in the form of a reverse-biased diode (rectifier), variable resistor (varistor), or RC network, as shown in Fig. 17-10. Determine *R* and *C* in an RC circuit with

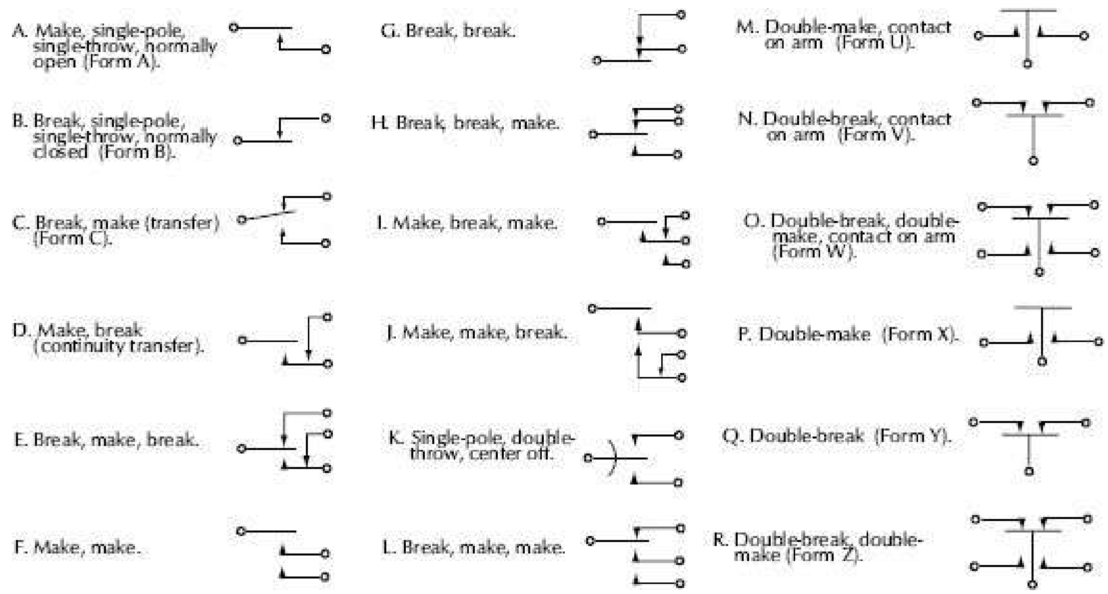


Figure 17-9. Various contact arrangements of relays. (From *American National Standard Definitions and Terminology for Relays for Electronics Equipment C83.16-1971*.)

$$C = \frac{I^2}{10}\,\mu\text{F} \qquad (17\text{-}14)$$

$$R = \frac{0.01\,V}{I^{\left(1+\frac{50}{V}\right)}} \qquad (17\text{-}15)$$

where,

C is in µF,

I is in A,

R is in Ω,

V is in V.

When using a rectifier, the rectifier is an open circuit to the power source because it is reverse biased; however, when the circuit breaks, the diode conducts. This technique depends on a reverse path for the diode to conduct; otherwise, it will flow through some other part of the circuit. It is important that the rectifier have a voltage rating equal to the transient voltage.

Contact bounce occurs in all mechanical-type relays except the mercury-wetted types that, because of the thin film of mercury on the contacts, do not break during make. Bounce creates noise in the circuit, particularly when switching audio where it acts as a dropout.

17.2.3 Relay Loads

Never assume that a relay contact can switch its rated current no matter what type of load it sees. High in-rush currents or high induced back electromotive force (emf) like those of Fig. 17-11 can quickly erode or weld electromechanical relay contacts and destroy solid-state relays.[8]

17.2.3.1 The Effects of Various Loads

Incandescent Lamps. The cold resistance of a tungsten-filament lamp is extremely low, resulting in in-rush currents as much as 15 times the steady-state current. This is why lamp burnout almost always occurs when turning on.

Capacitive Loads. The initial charging current to a capacitive circuit can be extremely high, since the capacitor acts as a short circuit, and current is limited only by the circuit resistance. Capacitive loads may be long transmission lines, filters for electromagnetic interference (emi) elimination, and power supplies.

Motor Loads. High in-rush current is drawn by most motors, because at standstill their input impedance is very low. This is particularly bad when aggravated by contact bounce causing several high-current makes and breaks before final closure. When the motor rotates, it develops an internal back emf that reduces the current. Depending on the mechanical load, the starting time may be very long and produce a relay-damaging in-rush current.

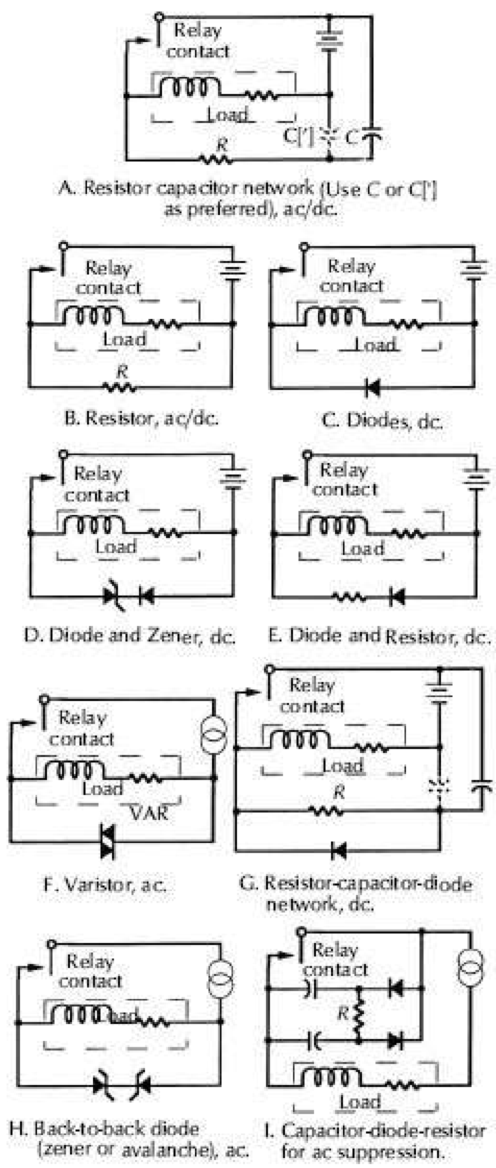


Figure 17-10. Methods of suppressing transients across contacts. Courtesy Magnecraft Electric Co.

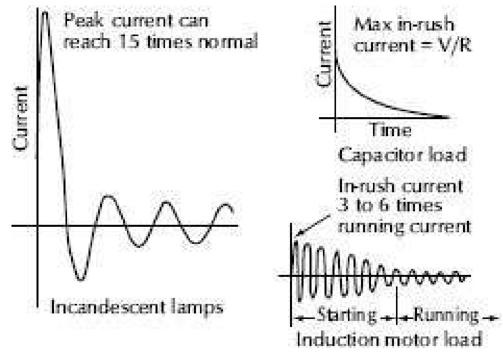


Figure 17-11. High in-rush current on turn-on can damage relays.

Inductive Loads. In-rush current is limited by inductance; however, when turned off, energy stored in magnetic fields must be dissipated.

Dc Loads. These are harder to turn off than ac loads because the voltage never passes through zero. When electromagnetic radiation (emr) contacts open, an arc is struck that may be sustained by the applied voltage, burning contacts.

17.2.4 Electromechanical Relays

Regardless of whether the relay operates on ac or dc, it will consist of an actuating coil, a core, an armature, and a group of contact springs that are connected to the circuit or circuits to be controlled. Associated with the armature are mechanical adjustments and springs. The mechanical arrangement of the contacts may be such that when the relay is at rest, certain circuits are either open or closed. If the contacts are open when the relay is at rest (not energized) they are called *normally open* contacts.

Relays are wound in many different manners, Fig. 17-12. Among them are the single wound, double wound, trifilar wound, bifilar wound, and two coil, which are nonelectromagnetic.

17.2.4.1 dc Relays

Direct current (dc) relays are designed to operate at various voltages and currents by varying the dc resistance of the actuating coils which may vary from a few ohms to several thousand ohms. Dc relays may operate as marginal, quick-operate, slow-operate, or polarized.

A *marginal relay* operates when the current through its winding reaches a specified value, and it releases when the current falls to a given value.

In the *quick-operate* type, the armature is attracted immediately to the pole piece of the electromagnet when the control circuit is closed.

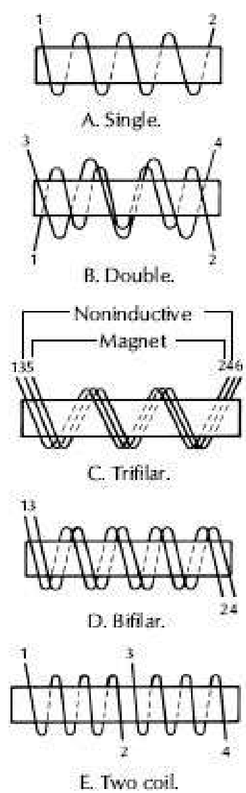


Figure 17-12. Types of relay coil windings.

Slow-operate relays have a time-delay characteristic; that is, the armature is not immediately attracted to the pole piece of the electromagnet when the control circuit is closed. To accomplish this a copper collar is placed around the armature end of the pole piece. They differ from the slow-release variety in that the latter type has the copper collar around the end of the pole piece opposite from the armature.

A *polarized relay* is designed to react to a given direction of current and magnitude. Polarized relays use a permanent magnet core. Current in a given direction increases the magnetic field, and in the opposite direction it decreases the field. Thus, the relay will operate only for a given direction of current through the coil.

A *latching relay* is stable in both positions. One type of latching relay contains two separate actuating coils. Actuating one coil latches the relay in one position where it remains until it is unlatched by energizing the other coil.

A second and more modern type is a bistable magnetic latching relay. This type is available in single-or dual-coil latching configurations. Both are bistable and will remain in either state indefinitely. The coils are designed for intermittent duty: 10s maximum on-time. The relay sets or resets on a pulse of 100ms or greater. Fig. 17-13 shows the various contact and coil forms.

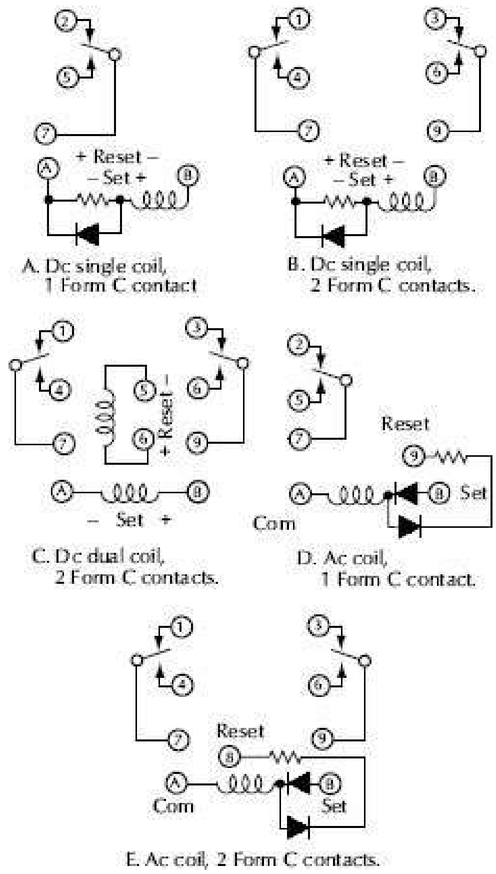


Figure 17-13. Various types and pin connections for latching relays. Courtesy Magnecraft Electric Co.

17.2.4.2 Ac Relays

Alternating-current (ac) relays are similar in construction to the dc relays. Since ac has a zero value every half cycle, the magnetic field of an ac-operated relay will have corresponding zero values in the magnetic field every half cycle.

At and near the instants of zero current, the armature will leave the core, unless some provision is made to hold it in position. One method consists of using an armature of such mass that its inertia will hold it in position. Another method makes use of two windings on separate cores. These windings are connected so that their respective currents are out of phase with each other. Both coils effect a pull on the armature when current flows in both windings.

A third type employs a split pole piece of which one part is surrounded by a copper ring acting as a shorted turn. Alternating current in the actuating coil winding induces a current in the copper coil. This current is out of phase with the current in the actuating coil and does not reach the zero value at the same instant as the current in the actuating coil. As a result, there is always enough pull on the armature to hold it in the operating position.

An *ac differential relay* employs two windings exactly alike, except they are wound in opposite directions. Such relays operate only when one winding is energized. When both windings are energized in opposite directions, they produce an aiding magnetic field, since the windings are in opposite directions. When the current through the actuating coils is going in the same direction, the coils produce opposite magnetic fields. If the current through the two coils is equal, the magnetic fields neutralize each other and the relay is nonoperative.

A differential polar relay employs a split magnetic circuit consisting of two windings on a permanent magnet core. A differential polar relay is a combination of a differential and a polarized relay.

17.2.5 Reed Relays

Reed relays were developed by the Bell Telephone Laboratories in 1960 for use in the Bell System central offices. The glass envelope is surrounded by an electromagnetic coil connected to a control circuit. Although originally developed for the telephone company, such devices have found many uses in the electronics industry.[6,7,8,9]

The term *reed relay* covers dry reed relays and mercury-wetted contact relays, all of which use hermetically sealed reed switches. In both types, the reeds (thin, flat blades) serve multiple functions, as conductor, contacts, springs, and magnetic armatures. Reed relays are usually soldered directly onto a circuit board or plugged into a socket that is mounted onto a circuit board.

17.2.5.1 Contact Resistance and Dynamics

Reed relays have much better switching speed than electromechanical relays. The fastest Coto Technology switching reed relay is the 9800 series, with a typical actuate time of 100μs. Release time is approximately 50μs. Actuate time is defined as the period from coil energization until the contact is closed and has stopped bouncing. After the contacts have stopped bouncing, they continue to vibrate while in contact with one another for a period of about 1ms. This vibration creates a wiping action and variable contact pressure.

Static contact resistance (SCR) is the resistance across the contact terminals of the relay after it has been closed for a sufficient period of time to allow for complete settling. For most reed relays, a few milliseconds is more than adequate, but the relay industry uses 50ms to define the measurement.

Another contact resistance measurement that has provided great insight into the overall quality of the relay is contact resistance stability (CRS). CRS measures the repeatability of successive static contact resistance measurements.

17.2.5.2 Magnetic Interaction

Reed relays are subject to external magnetic effects including the earth's magnetic field (equivalent to approximately 0.5AT and generally negligible), electric motors, transformers external magnets, etc., which may change performance characteristics. Such magnetic sources include one common source of an external magnetic field acting on a relay or another relay operating in close proximity. The potential for magnetic coupling must be taken into account when installing densely packed single- or multichannel relays.

An example of magnetic interaction is shown in Fig. 17-14 where two relays, K1 and K2, with identical coil polarities are mounted adjacent to each other. When K2 is “off”, relay K1 operates at its designed voltage. When K2 is activated, the magnetic fields oppose so the effective magnetic flux within K1 is reduced, requiring an increase in coil voltage to operate the reed switch. For closely packed relays without magnetic shields, a 10–20% increase in operate voltage is typical, which can drive the relays above their specified limits. The opposite effect occurs if K1 and K2 are polarized in opposite directions making the operating voltage for K1 less.

There are several ways to reduce magnetic interaction between relays:

- Specify relays that incorporate an internal or external magnetic shield.
- Apply an external magnetic shield to the area where the relays are mounted. A sheet of mu-metal or other high-magnetic-permeability ferrous alloy 2–5mils thick is effective.

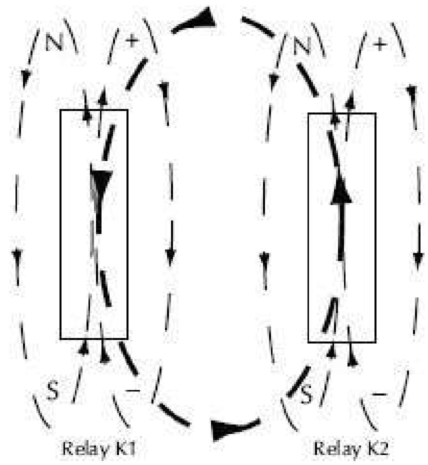


Figure 17-14. Adverse magnetic interaction. Courtesy Coto Technology.

- Provide increased on-center spacing between relays. Each doubling of this distance reduces the interaction effect by a factor of approximately four.
- Avoid simultaneous operation of adjacent relays.
- Provide alternating coil polarities for relays used in a matrix.

17.2.5.3 Environmental Temperature Effects

The resistance of the copper wire used in reed relay coils increases by 0.4%/1°C rise in temperature. Reed relays are current-sensitive devices so their operate and release levels are based on the current input to the coil. If a voltage source is used to drive the relays, an increase in coil resistance causes less current to flow through the coil, so the voltage must be increased to compensate and maintain current flow. Industry standards define that relays are typically specified at 25°C ambient. If the relay is used in higher ambient conditions or near external sources of heat, this must be carefully considered.

For example, a standard relay nominally rated at 5Vdc has a

3.8Vdc maximum operate value at 25°C as allowed by the specifications. If the relay is used in a 75°C environment, the 50°C temperature rise increases the operate voltage by 50 × 0.4%, or 20%. The relay now will operate at 3.8Vdc + (3.8Vdc × 20%), or 4.56Vdc. If there is more than a 0.5Vdc drop in supply voltage due to a device driver or sagging power supply, the relay may not operate. Under these conditions there will be increases in operate and release timing to approximately the same 20%.

17.2.5.4 Dry Reed Relays

Because of the tremendous increases in low-level logic switching, computer applications, and other business machine and communication applications, dry reed relays have become an important factor in the relay field. They have the great advantage of being hermetically sealed, making them impervious to atmospheric contamination. They are very fast in operation and when operated within their rated contact loads, they have a very long life. They can be manufactured automatically and therefore are relatively inexpensive. A typical dry reed switch capsule is shown in Fig. 17-15.

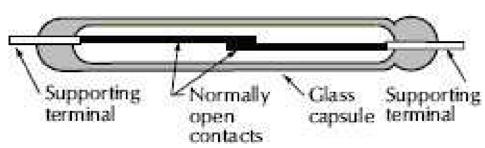


Figure 17-15. Construction of a switch capsule of a typical dry reed relay—Form A. Courtesy Magnecraft Electric Co.

In this basic design, two opposing reeds are sealed into a narrow glass capsule and overlap at their free ends. At the contact area, they are plated with rhodium over gold to produce a low contact resistance when they meet. The capsule, surrounded by an electromagnetic coil, is made of glass and filled with a dry inert gas. When the coil is energized in the basic Form A contact combination, the normally open contacts are brought together; when the field is removed the reeds separate by their own spring tension.

Some may contain permanent magnets for magnetic biasing to achieve normally closed contacts (Form B). Single-pole, double-throw contact combinations (Form C) are also available. Current rating, which is dependent on the size of the reed and the type and amount of plating, may range from low level to 1A. Effective contact protection is essential in most applications unless switching is done dry.

Relay packages using up to four Form C and six Form A dry reed switches are common, providing multiple switching arrangements. The reed relay may be built for a large variety of operational modes such as pulse relay, latch relay, crosspoint relay, and logic relay. These relays may also be supplied with electrostatic or magnetic shields. The relay in Fig. 17-16 has two Form C contacts.

Reed switches have the following characteristics:

- A high degree of reliability stemming from their controlled contact environment.

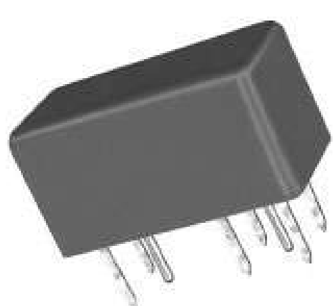


Figure 17-16. Coto Technology 2342 multipole relay, Courtesy

Coto Technology.

- Consistency of performance resulting from a minimum number of parts.
- Long operational life.
- Ease of packaging as a relay.
- High-speed operation.
- Small size.
- Low cost.

Number of Switches. There appears to be no limit to the number of switches that can be actuated by a common coil. However, as the number increases, coil efficiency decreases and power input increases. This can lead to a practical limitation. On the other hand, the increase in power required to operate one more switch capsule is usually less than the total required if the assembly were split in two. The single contact relay is the most frequently used but relays with four or more switches in a single coil are quite common.

Sensitivity. The power input required to operate dry reed relays is determined by the sensitivity of the particular reed switch used, by the number of switches operated by the coil, by the permanent magnet biasing (if used), and by the efficiency of the coil and the effectiveness of its coupling to the reeds. The minimum input required to effect closure ranges from milliwatts for a single capsule sensitive unit to several watts for a multipole relay.

Operate Time. Coil time constant, overdrive, and the characteristics of the reed switch determine operate time. With maximum overdrive, reed relays will operate in approximately 200μs or less. Drive at rated voltage usually results in a 1ms operate time.

Release Time. With the relay coil unsuppressed, dry reed switch contacts release in a fraction of a millisecond. Form A contacts open in as little as 50μs. Magnetically biased Form B contacts and normally closed contacts of Form C switches reclose from 100μs to 1ms, respectively.

If the relay coil is suppressed, release times are increased. Diode suppression can delay release for several milliseconds, depending on coil characteristics, drive level, and reed release characteristics.

Bounce. As with the other hard contact switches, dry reed contacts bounce on closure. The duration of bounce is typically quite short and is in part dependent on drive level. In some of the faster devices, the sum of operate time and bounce is relatively constant so as drive is increased, the operate time decreases and bounce increases.

While normally closed contacts of a Form C switch bounce more than normally open contacts, magnetically biased Form B contacts exhibit essentially the same bounce as Form A.

Contact Resistance. Because the reeds in a dry reed switch are made of a magnetic material that has a high volume resistivity, terminal-to-terminal resistance is somewhat higher than in some other types of relays. Typical specification limit for initial maximum resistance of a Form A reed relay is 0.200Ω.

Switch Lead Modifications. It is recommended that the lead bend of switches be no less than 1.0mm from the glass seal to prevent cracking of the seal or destruction of the residual stress within the seal.

The glass seal has a very high resistance to pressure, but a low resistance to pull forces. The ability of the seal to properly withstand a switch lead modification is dependent on several factors: the relation of the wire to glass dimensions, the length of the seal, the pull force or type of modification to be performed, and the distance and direction of the mechanical force in relation to the seal itself.

It is important to have the proper support and clamping of the leads to avoid damage to the seal. Even then, the plastic deformation strain of the NiFe lead wires can be transmitted through the clamping area and into the seal.

The risk of plastic deformation strain in the NiFe wire leads is very high so supporting or clamping the lead wire between the glass seal and the cut or bend is necessary to avoid damage of the glass seal.

Any form of bending, cutting or modification of the switch leads requires appropriate tooling and fixtures to minimize the introduction of mechanical stress on the glass seal area. Many years of experience has proven that switch damage can be avoided when lead modification is required by proper fixtures and the avoidance of any modification less than 1mm from the glass seal.

17.2.5.5 Mercury-Wetted Contact Relays

Mercury-wetted contact relays are a form of reed relays consisting of a glass-encapsulated reed with its base immersed in a pool of mercury and the other end capable of moving between one or two stationary contacts. The mercury flows up to the reed by capillary action and wets the contact surface of the moving end of the reed as well as the contact surfaces of the stationary contacts. Thus a mercury-to-mercury contact is maintained in a closed position. The mercury-wetted relay is usually actuated by a coil around the capsule.

Aside from being extremely fast in operation and having relatively good load-carrying capacity, mercury-wetted contact relays have extremely long life since the mercury films are reestablished at each contact closure and contact erosion is eliminated. Since the films are “stretchable,” there is no contact bounce. Contact interface resistance is extremely low.

Three disadvantages of this type of reed relays are:

1. The freezing point of mercury is (−38.8°C or −37.8°F).
2. They have poor resistance to shock and vibration.
3. Some type need to mount in a near vertical position.

These relays are available in a compact form for printed-circuit board mounting. Multipole versions can be provided by putting additional capsules inside the coil. They are used for a great variety of switching applications such as are found in computers, business machines, machine tool control systems, and laboratory instruments.

Mercury-wetted switches can also come as a non-position sensitive, mercury-wetted, reed relay that combines the desirable features of both dry reed and mercury-wetted capsules. This allows the user to place the reed relay in any position and is capable of withstanding shock and vibration limits usually associated with dry reed capsules. On the other hand, they retain the principal advantages of other mercury-wetted switches—no contact bounce and low stable contact resistance.

Operation of the nonposition-sensitive switch is made possible by the elimination of the pool of mercury at the bottom of the capsule. Its design captures and retains the mercury on contact and blade surfaces only. Due to the limited amount of mercury film, this switch should be restricted for use at low-level loads.

Mercury-wetted reed relays are a distinct segment of the reed relay family. They are different from the dry reed relays in the fact that contact between switch elements is made via a thin film of mercury. Thus, the most important special characteristics of mercury-wetted relays are:

- Contact resistance is essentially constant from operation to operation throughout life.
- Contacts do not exhibit bounce. The amount of mercury at the contacts is great enough to both cushion the impact of the underlying members and to electrically bridge any mechanical bounce that remains.
- Life is measured in billions of operations, due to constant contact surface renewal.
- Contacts are versatile. The same contacts, properly applied, can handle relatively high-power and low-level signals.
- Electrical parameters are constant. With contact wear eliminated, operating characteristics remain the same through billions of operations.

To preserve these characteristics, the rate of change of voltage across the contacts as they open must be limited to preclude damage to the contact surface under the mercury. For this reason, suppression should be specified for all but low-level applications.

Mounting Position. To ensure that distribution of mercury to the relay contacts is proper, position sensitive types should be mounted with switches oriented vertically. It is generally agreed that deviation from vertical by as much as 30° will have some effect on performance. The nonposition-sensitive mercury-wetted relay, which is the most common type today, is not affected by these limitations.

Bounce. Mercury-wetted relays do not bounce if operated within appropriate limits. However, if drive rates are increased, resonant effects in the switch may cause rebound to exceed the level that can be bridged by the mercury, and electrical bounce will result. Altered distribution of mercury to the contacts, caused by the high rate of operation, may also contribute to this effect.

Contact Resistance. Mercury-wetted relays have a terminal-to-terminal contact resistance that is somewhat lower than dry reed relays. Typical specification limit for maximum contact resistance is 0.150Ω.

17.2.5.6 RF Relays

RF relays are used in high-frequency applications, usually in a 50Ω circuit. The RF coaxial shielded relay in Fig. 17-17 can switch up to 200Vdc at 0.5A.

Figure 17-17. Coto Technology 9290 RF reed relay. Courtesy Coto Technology.

Insertion and Other Losses. In the past, the typical parameters used to quantify RF performance of reed relays were Insertion loss, isolation, and return loss (sometimes called reflection loss). These are frequency-related vector quantities describing the relative amount of RF power entering the relay and either being transmitted to the output or being reflected back to the source. For example, with the relay's reed switch closed and 50% power being transmitted through the relay, the insertion loss would be 0.5 or −3dB. The frequency at which a −3dB roll-off occurs is a convenient scalar (single-valued) quantity for describing insertion loss performance.

Isolation. The RF isolation of the reed relay can be determined by injecting an RF signal of known power amplitude with the reed switch open (coil unactivated). Sweeping the RF frequency and plotting the amount of RF energy exiting the relay allows the isolation curve to be plotted on a dB scale. At lower frequencies, the isolation may be −40dB or greater, indicating that less than 0.01% of the incident power is leaking through the relay. The isolation decreases at higher frequencies, because of capacitive leakage across reed switch contacts.

Voltage Standing Wave Ratio (VSWR). VSWR is a measurement of how much incident signal power is reflected back to the source when an RF signal is injected into a closed relay terminated with a 50Ω impedance. It represents the ratio of the maximum amplitude of the reflected signal envelope amplitude divided by the minimum at a specified frequency. A VSWR of 1 indicates a perfect match between the source, relay, and output load impedance and is not achievable. VSWR at any particular frequency can be converted from y-axis return loss using Table 17-4.

Table 17-4. Return Loss Versus VSWR

Return Loss VSWR (dB)	VSWR
−50	1.01
−40	1.02
−30	1.07
−20	1.22
−10	1.93
−3	5.85

Return Loss. Return loss represents the amount of RF power being reflected back to the source with the reed switch closed and the output terminated with a standard impedance, normally 50Ω. If the relay was closely matched to 50Ω at all frequencies, the reflected energy would be a very small fraction of the incident energy from low to high frequencies. In practice, return loss increases (more power is reflected) as frequency increases. High return loss (low reflective energy) is desirable for high-speed pulse transmission, since there is less risk of echoing signal collisions that can cause binary data corruption and increased bit error rates. Return loss is calculated from the reflection coefficient (ρ), which is the ratio of the magnitude of signal power being reflected from a closed relay to the power input at a specified frequency

$$Return\ loss = -20\log\rho \qquad (17\text{-}16)$$

To determine the RF performance of a reed relay involves injecting a swept frequency RF signal of known power into the relay and measuring the amount of RF energy transmitted through or reflected back from it. These measurements can be conveniently made using a Vector Network Analyzer (VNA). These test instruments comprise a unified RF sweep frequency generator and quantitative receiver/detector. In the case of a Form A relay, the device is treated as a network with one input and one output port, and the amount of RF energy entering and being reflected from each port is recorded as a function of frequency. Thus a complete characterization of a Form A relay comprises four data vectors, designated as follows:

S_{11}	power reflected from input port.
S_{12}	power transmitted to input port from output port.
S_{21}	power transmitted to output port from input port.
S_{22}	power reflected from output port.

Rise Time. The rise time of a reed relay is the time required for its output signal to rise from 10% to 90% of its final value, when the input is changed abruptly by a step function signal. The relay can be approximated by a simple first-order low-pass filter. The rise time is approximately

$$\begin{aligned} T_r &= RC \times \ln\frac{90\%}{10\%} \\ &= 2.3RC \end{aligned} \qquad (17\text{-}17)$$

where,

T_r is in ps,

RC is a figure of merit expressed in pF•Ω where R is the closed contact resistance in Ω and C is the open contact capacitance in pF.

Substituting into the equation for the 50% roll-off frequency $f_{-3\,\mathrm{dB}} = 1/2\pi RC$ yields the relationship

$$T_r = \frac{0.35}{f_{-3\,\mathrm{dB}}} \qquad (17\text{-}18)$$

Therefore the relay's rise time can be simply estimated from the S_{21} insertion loss curve by dividing the −3dB roll-off frequency into 0.35. For example, the Coto Technology B40 ball grid relay has $f_{-3\mathrm{dB}}$ = 11.5GHz, from which the rise time can be estimated as 30ps.

Effect of Lead Form on High Frequency Performance. Surface mount (SMD) relays give better RF performance than those with through hole leads. SMD lead-forms comprise gullwing, J-bend, and axial forms. Each has its advantages and disadvantages, but axial relays generally have the best RF performance in terms of signal losses, followed by J-bend and gullwing. The straight-through signal path of axial relays minimizes capacitive and inductive reactance in the leads and minimizes impedance discontinuities in the relay, resulting in the highest bandwidth. J-bend relays provide the next-best RF performance and have the advantages of requiring slightly less area on the PCB. The gullwing form is the most common type of SMD relay. Initial pick-and-place soldering is simple, as is rework, resulting in a broad preference for this lead type unless RF performance is critical.

Coto Technology's leadless relays have greatly enhanced RF performance. They do not have traditional exposed metal leads; instead, the connection to the user's circuit board is made with ball-grid-array (BGA) attachment, so that the devices are essentially leadless. In the BGA relays, the signal path between the BGA signal input and output is designed as an RF transmission line, with an RF impedance close to 50Ω throughout the relay. This is achieved using a matched combination of coplanar waveguide and coaxial structures with very little impedance discontinuity through the relays. The Coto B10 and B40 reed relays, Fig. 17-18 achieve bandwidths greater than 10GHz and rise times of 35ps or less.

Figure 17-18. Coto Technology B40 Ball Grid surface mount 4-channel reed relay. Courtesy Coto Technology.

Skin Effect in Reed Relays. At high frequencies, RF signals tend to travel near the surface of conductors. The skin effect is exaggerated in metals with high magnetic permeability, such as the nickel-iron alloy used for reed switch blades which has to carry the switched current and also respond to a magnetic closure field. Skin effect does not appreciably affect the operation of reed relays at RF frequencies because the increase in ac resistance due to skin effect is proportional to the square root of frequency, whereas the losses due to increasing reactance are directly proportional to *L* and inversely proportional to *C*. Also the external lead surfaces are coated with tin or solder alloys for enhanced solder-ability which helps to reduce skin effect losses.

Selecting Reed Relays for High Frequency Service. High-speed switching circuits can be accomplished with reed relays, electromechanical relays (EMRs) specifically designed for high-frequency service, solid-state relays (SSRs), PIN diodes, and microelectromechanical systems (MEMS) relays. In many cases, reed relays are an excellent choice, particularly with respect to their unrivalled RC product. RC is a figure of merit expressed in pF•Ω, where *R* is the closed contact resistance and *C* is the open contact capacitance. The lower this figure is, the better the high-frequency performance. The RC product of a Coto Technology B40 relay for example, is approximately 0.02pF•Ω. SSRs have pF•Ω products equal to about 6, almost 300 times higher, plus, the breakdown voltage at these pF•Ω levels is much lower than that of a reed switch. The turn-off time for SSRs is also longer than the 50μs needed by a reed relay to reach its typical $10^{12}\Omega$ off resistance. Many reed relays have demonstrated MCBF values of several hundred million to several billion closure cycles at typical signal switching levels.

17.2.5.7 Dry Reed Switches

A dry reed switch is an assembly containing ferromagnetic contact blades that are hermetically sealed in a glass envelope and are operated by an externally generated magnetic field. The field can be a coil or a permanent magnet. The switches in Figs. 17-19A and 17-19B can switch up to 175Vdc at 350mA or 140Vac at 250ma. The switch in Fig. 17-19C can switch 200Vdc at 1A or 140Vac at 1A.

Fig. 17-20 shows three methods of operating a reed switch using a coil.

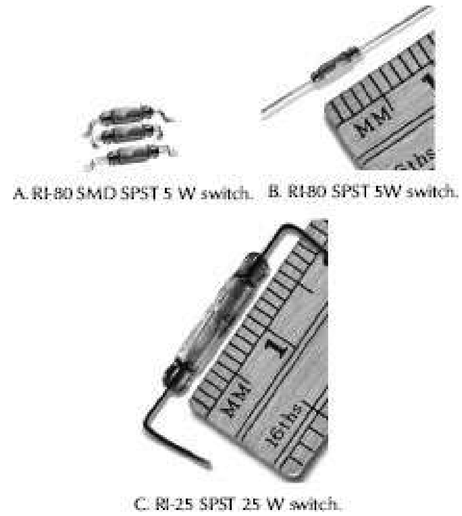


Figure 17-19. Coto Technology dry reed switches. Courtesy Coto Technology.

17.2.6 Solid-State Relays

Solid-state relays (SSRs) utilize the on–off switching properties of transistors and SCRs for opening and closing dc circuits and triacs for switching ac circuits.[10]

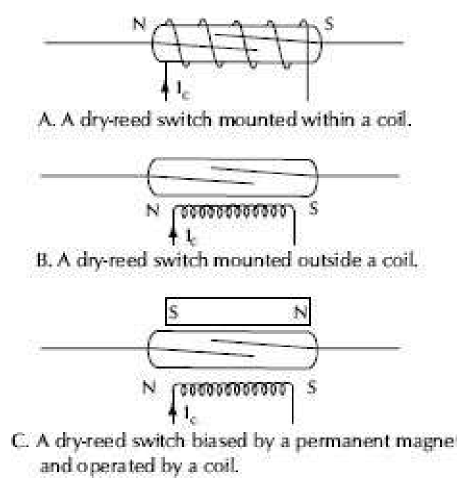


Figure 17-20. Energizing a dry-reed switch with a coil. Courtesy Coto Technology.

17.2.6.1 Advantages

SSRs have several advantages over their electromechanical counterparts: no moving parts, arcing, burning, or wearing of contacts; and the capacity for high-speed, bounceless, noiseless operation. Many SSRs are available that feature optical coupling; thus, the signal circuit includes a lamp or light-emitting diode that shines on a phototransistor serving as the actuating device. In other types of SSRs, a small reed relay or transformer may serve as the actuating device. A third type is direct coupled and therefore not actually an SSR because there is no isolation between input and output. These are better called an amplifier. All three types are shown in Fig. 17-21.

Ac relays turn on and off at zero crossing; therefore, they have reduced *dv/dt*. However, this does slow down the action to the operating frequency.

17.2.6.2 Disadvantages and Protection

Solid-state relays also have some inherent problems as they are easily destroyed by short circuits, high surge current, high *dv/dt*, and high peak voltage across the power circuit.

Short-circuit and high-surge current protection is performed with fast blow fuses or series resistors. A standard fuse normally will not blow before the SCR or triac is destroyed since the fuses are designed to withstand surge currents. Fast blow fuses will act on high in-rush currents and usually protect solid-state devices.

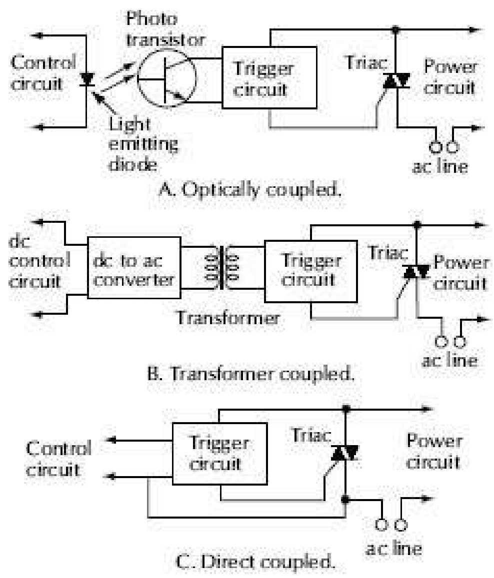


Figure 17-21. Various types of solid-state relays.

Using a current-limiting resistor will protect the SSR; however, it creates a voltage drop that is current dependent and, at high current, dissipates high power.

A common technique for protecting solid-state switching elements against high *dv/dt* transients is by shunting the switching element with an RC network (snubber), as shown in Fig. 17-22. The following equations provide effective results:

$$R_1 = \frac{L}{V} \times \frac{dv}{dt} \qquad (17\text{-}19)$$

$$R_2 = \frac{\sqrt{1-(PF)^2}}{2\pi f} \times \frac{dv}{dt} \qquad (17\text{-}20)$$

$$C = \frac{4L}{R_2^2} \qquad (17\text{-}21)$$

$$C = \frac{4}{R_2^{\,2}} \times \frac{V}{I} \times \frac{\sqrt{1-PF^2}}{2\pi F} \qquad (17\text{-}22)$$

where,

L is the inductance in H,

V is the line voltage,

dv/dt is the maximum permissible rate of change of voltage in V/μs,

I is the load current,

PF is the load power factor,

C is the capacitance in μF,

R_1, R_2 are the resistance in Ω,

f is the line frequency.

RC networks are often internal to SSRs.

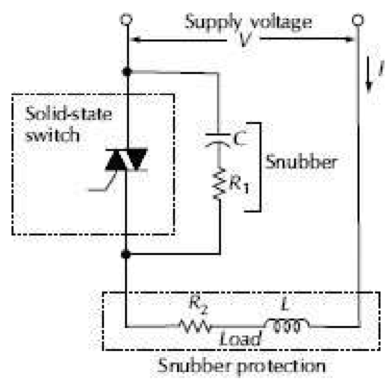


Figure 17-22. Snubber circuit for solid-state relay protection.

17.2.6.3 High-Peak-Transient-Voltage Protection

Where high-peak-voltage transients occur, effective protection can be obtained by using metal-oxide varistors (MOVs). The MOV is a bidirectional voltage-sensitive device that becomes low impedance when its design voltage threshold is exceeded.

Fig. 17-23 shows how the proper MOV can be chosen. The peak nonrepetitive voltage (V_{DSM}) of the selected relay is transposed to the MOV plot of peak voltage versus peak amperes. The corresponding current for that peak voltage is read off the chart.

Using this value of current (I) in

$$V_{DSM} = V_p - IR \tag{17-23}$$

where,

I is the current,

V_p is the peak instantaneous voltage transient,

R is the load plus source resistance.

It is important that the V_{DSM} peak nonrepetitive voltage of the SSR is not exceeded.

The energy rating of the MOV must not be exceeded by the value of

$$E = V_{DSM} \times I \times t \tag{17-24}$$

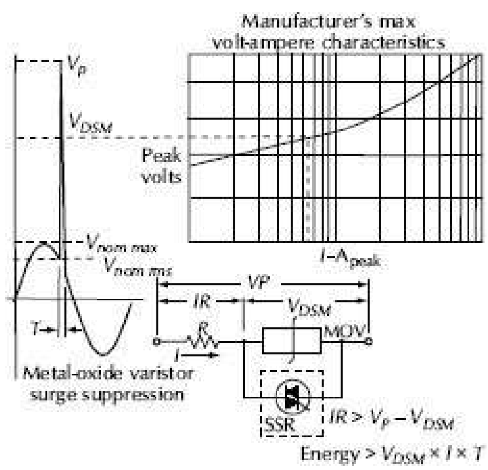


Figure 17-23. Metal-oxide varistor peak transient protector.

17.2.6.4 Low Load Current Protection

If the load current is low, it may be necessary to take special precautions to ensure proper operation. Solid-state relays have a finite off-state leakage current. SSRs also need a minimum operating current to latch the output device.

If the off-state voltage across the load is very high, it could cause problems with circuit dropout and component overheating. In these applications a low-wattage incandescent lamp in parallel with the load offers a simple remedy. The nonlinear characteristics of the lamp allow it to be of lower resistance in the off state while conserving power in the on state. It must be remembered to size the SSR for the combined load.

17.2.6.5 Optically Coupled Solid-State Relays

The optically coupled solid-state relay arrangement (SSR) shown in Fig. 17-21A is capable of providing the highest control/power-circuit isolation—many thousands of volts in compact, convenient form. The triac trigger circuit is energized by a phototransistor, a semiconductor device (encapsulated in transparent plastic) whose collector-emitter current is controlled by the amount of light falling on its base region.

A phototransistor is mounted in a light-tight chamber with a light-emitting diode, the separation between them being enough to give high isolation (thousands of volts) between the control and power circuit.

The light-emitting diode requires only 1.5V to energize and has very rapid response time. The power circuit consists of a high-speed phototransistor and an SCR for dc power source, and a triac for ac application.

The relay not only responds with high speed but is also capable of very fast repetitious operation and provides very brief delays in turnoff. In some applications, the photocoupler housing provides a slotted opening between the continuously lit light-emitting diode and the phototransistor. On–off control is provided by a moving arm, vane, or other mechanical device that rides in the slot and interrupts the light beam in accordance with some external mechanical motion. Typical optically coupled SSRs have the following characteristics:

Turn-on control voltage	3–30Vdc
Isolation	1500Vac
dv/dt	100V/μs
Pickup control voltage	3Vdc
Dropout control voltage	1Vdc
One-cycle surge (rms)	7–10 times nominal
1 second overload	2.3 times nominal
Maximum contact voltage drop	1.5–4V

17.2.6.6 Transformer-Coupled Solid-State Relays

In Fig. 17-21B, the dc control signal is changed to ac in a converter circuit, the output of which is magnetically coupled to the triac trigger circuit by means of a transformer. Since there is no direct electrical connection between the primary and secondary of the transformer, control/power-circuit isolation is provided up to the voltage withstanding limit of the primary/secondary insulation.

17.2.6.7 Direct-Coupled Solid-State Relays

The circuit shown in Fig. 17-21C cannot truly be called a solid-state relay because it does not have isolation between input and output. It is the simplest configuration; no coupling device is interposed between the control and actuating circuits, so no isolation of the control circuit is provided. This circuit would be better called an amplifier.

One other variation of these solid-state circuits is occasionally encountered—the *Darlington* circuit. A typical arrangement is shown in Fig. 17-24. Actually a pair of cascaded power transistors, this circuit is used in many solid-state systems to achieve very high power gain—1000 to 10,000 or more. Marketed in single-transistor cases, it can be obtained as what appears to be a single transistor with high operating voltage ratings that control high amperage loads with only a few volts at the base connection and draw only a few milliamperes from the control circuit. It can be used for relay purposes in a dc circuit the same way, either by direct control signal coupling or with intermediate isolation devices like those described. It is not usable in ac power circuits.

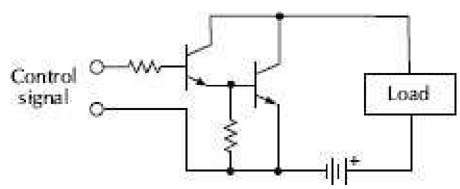


Figure 17-24. Darlington direct-coupled solid-state relay.

17.2.6.8 Solid-State Time-Delay Relays

Solid-state time-delay relays, Fig. 17-25, can operate in many different modes since they do not rely on heaters or pneumatics.[11] Simple ICs allow the relays to do standard functions plus totaling, intervals, and momentary action as described in the following.

On-Delay. Upon application of control power, the time-delay period begins. At the end of time delay, the output switch operates. When control power is removed, the output switch returns to normal, Fig. 17-25A.

Nontotalizer. Upon the opening of the control switch, the time-delay period begins. However, any control switch closure prior to the end of the time delay will immediately recycle the timer. At the end of the time-delay period, the output switch operates and remains operated until the required continuous power is interrupted, as shown in Fig. 17-25B.

Totalizer/Preset Counter. The output switch will operate when the sum of the individual control switch closure durations equal the preset time-delay period. There may be interruptions between the control switch closures without substantially altering the cumulative timing accuracy. The output switch returns to normal when the continuous power is interrupted, as shown in Fig. 17-25C.

Off-Delay. Upon closure of the control switch, the output switch operates. Upon opening of the control switch, the time-delay period begins. However, any control switch closure prior to the end of the time-delay period will immediately recycle the timer. At the end of the time-delay period, the output switch returns to normal. Continuous power must be furnished to this timer, as shown in Fig. 17-25D.

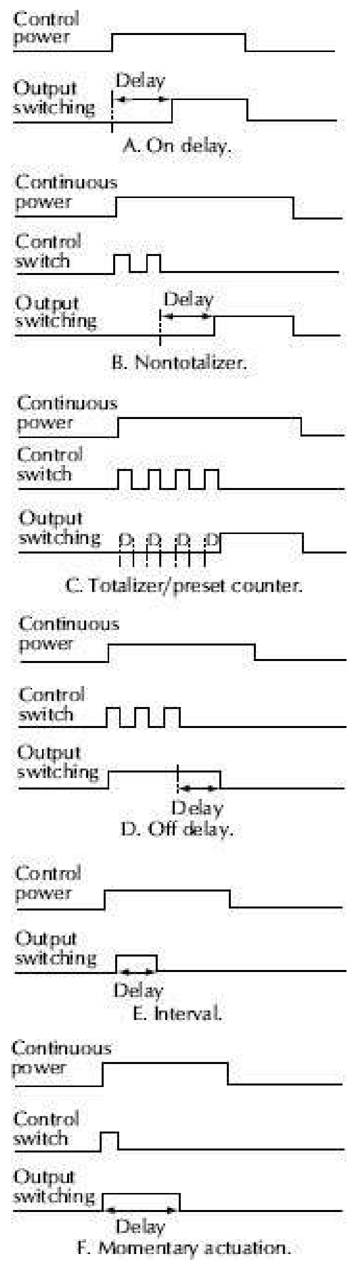


Figure 17-25. Types of time-delay relays.

Interval. Upon application of the control power, the output switch operates. At the end of the time-delay period, the output switch returns to normal. Control power must be interrupted in order to recycle, as shown in Fig. 17-25E.

Momentary Actuation. Upon closure of the control switch, the output switch operates, and the time-delay period begins. The time-delay period is not affected by duration of the control switch closure. At the end of the time-delay period, the output switch returns to normal. Continuous power must be furnished to this timer, as shown in Fig. 17-25F.

Programmable Time-Delay Relay. Programmable time-delay relays are available where the time and functions can be programmed by the user. The Magnecraft W211PROGX-1 relay in Fig. 17-26 is an example of this type. It plugs into an octal socket, has ±0.1% repeatability and four input voltage ranges. It has four programmable functions, On Delay, Off Delay, One Shot, and On Delay and Off Delay. There are 62 programmable timing ranges from 0.1s to 120min and the relay has 10A DPDT contacts. An eight position DIP switch is used to program the timing function and a calibrated knob is used to set the timing.

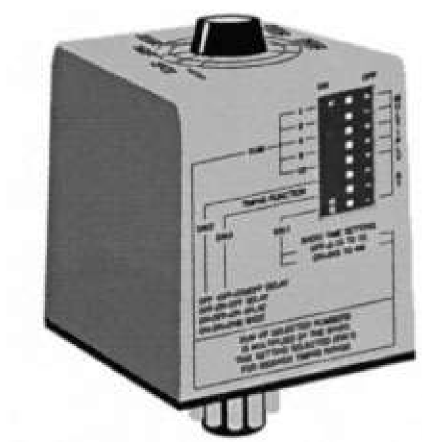


Figure 17-26. A programmable time delay relay. Courtesy Magnecraft Electric Co.

References

1. Logan Hedin and Gary Arnold, *Thermal Design of an LED Commercial Florescent Tube*, Presented at IMAPs New England 2008 35th Symposium.

2. Solver-Calc©, An Excel Spreadsheet Program for Rapid Thermal Management Calculations, by Villaume Associates, LLC, www.thermalseminars.com.

3. Heat pipes have become increasingly useful in spreading heat loads and moving heat to an area where the convective fin surface area may be more effectively positioned in the available air flow path. "Google" for vendors as there are quite a few suppliers of this component in a thermal solution.

4. TPG is a product by Momentive Performance Materials, 180 East Broad Street, Columbus, Ohio, 43215.

5. Two major suppliers; Wakefield Thermal Solutions, Pelham, N.H., and Cool Options Inc., Warwick, R.I.

6. *Designer's Handbook and Catalog of Reed and Mercury Wetted Contact Relays*, Magnecraft Electric Co.

7. "Relays, Solenoids, Timers, and Motors," *Electronic Buyer's Handbook*, 1978. Copyright 1978 by CMP Publications. Reprinted with permission.

8. "Relay Loads" *Electronic Design* 26, December 20, 1978.

9. *Coto Technology Reed Relays and Dry Reed Switches*, Coto Technology.

10. *Solid-State Relays*, Magnecraft Electric Co.

11. *Designer's Handbook and Catalog of Time-Delay Relays*, Magnecraft Electric Co.

* Section 17.1.1–17.1.2 supplied in part by Henry Villaume, Villaume Associates, LLC.

Chapter 18

Transmission Techniques: Wire and Cable

by Steve Lampen and Glen Ballou

18.1 Introduction

It was not long ago that wire was the only method to inexpensively and reliably transmit sound or pictures from one place to another. Today we not only have wire, but we also have fiber optics, and wireless radio frequency (RF) transmission from Blu-tooth to wireless routers, cell phones, and microwave and satellite delivery. RF transmission is discussed briefly in Chapter 20.10, Wireless Microphones. This chapter will discuss the various forms of wire and cable used in audio and video.

Wire is a single conductive element. Wire can be insulated or uninsulated. Cable, on the other hand, is two or more conductive elements. While they theoretically could be uninsulated, the chance of them touching each other and creating a short circuit requires that they are usually both insulated. A cable can be multiple insulated wires, called a *multiconductor cable*, or wires that are twisted together, called a *twisted pair cable*, or cables with one wire in the center, surrounded by insulation and then a covering of metal used as another signal path, called *coaxial cable*.

18.2 Conductors

Wire and cable are used to connect one circuit or component to another. They can be internal, connecting one circuit to another inside a box, or externally connecting one box to another.

18.2.1 Resistance and Wire Size

Wire is made of metal, or other conductive compounds. All wire has resistance which dissipates power through heat. While this is not apparent on cables with small signals, such as audio or video signals, it is very apparent where high power or high current travels down a cable, such as a power cord. Resistance is related to the size of the wire. The smaller the wire, the greater the resistance.

18.2.2 Calculating Wire Resistance

The resistance for a given length of wire is determined by:

$$R = \frac{KL}{d^2} \qquad (18\text{-}1)$$

where,

R is the resistance of the length of wire in Ω,

K is the resistance of the material in ohms per circular mil foot,

L is the length of the wire in ft, d is the diameter of the wire in mils.

The resistance, in ohms per circular mil foot (Ω/cir mil ft), of many of the materials used for conductors is given in Table 18-1. The resistance shown is at 20°C (68°F), commonly called *room temperature*.

Table 18-1. Resistance of Metals and Alloys

Material	Symbol	Resistance (Ω/cir mil ft)
Silver	Ag	9.71
Copper	Cu	10.37
Gold	Au	14.55
Chromium	Cr	15.87
Aluminum	Al	16.06
Tungsten	W	33.22
Molybdenum	Mo	34.27
High-brass	Cu-Zn	50.00
Phosphor-bronze	Sn-P-Cu	57.38
Nickel, pure	Ni	60.00
Iron	Fe	60.14
Platinum	Pt	63.80
Palladium	Pd	65.90
Tin	Sn	69.50
Tantalum	Ta	79.90
Manganese-nickel	Ni-Mn	85.00
Steel	C-Fe	103.00
Lead	Pb	134.00
Nickel-silver	Cu-Zn-Ni	171.00
Alumel	Ni-Al-Mn-Si	203.00
Arsenic	As	214.00
Monel	Ni-Cu-Fe-Mn	256.00
Manganin	Cu-Mn-Ni	268.00
Constantan	Cu-Ni	270.00
Titanium	Ti	292.00
Chromel	Ni-Cr	427.00
Steel, manganese	Mn-C-Fe	427.00
Steel, stainless	C-Cr-Ni-Fe	549.00
Chromax	Cr-Ni-Fe	610.00

Nichrome V	Ni-Cr	650.00
Tophet A	Ni-Cr	659.00
Nichrome	Ni-Fe-Cr	675.00
Kovar A	Ni-Co-Mn-Fe	1732.00

When determining the resistance of a twisted pair, remember that the length of wire in a pair is twice the length of a single wire. Resistance in other constructions, such as coaxial cables, can be difficult to determine from just knowing the constituent parts. The center conductor might be easy to determine but a braid or braid + foil shield can be difficult. In those cases, consult the manufacturer.

Table 18-1 show the resistance in ohms (O) per foot per circular mil area for various metals, and combinations of metals (alloys). Of the common metals, silver is the lowest resistance. But silver is expensive and hard to work with. The next material, copper, is significantly less expensive, readily available, and lends itself to being annealed, which is discussed in Section 18.2.4. Copper is therefore the most common material used in the manufacture of wire and cable. However, where price is paramount and performance not as critical, aluminum is often used. The use of aluminum as the conducting element in a cable should be an indication to the user that this cable is intended to be lower cost and possibly lower performance.

One exception to this rule might be the use of aluminum foil which is often used in the foil shielding of even expensive high-performance cables. Another exception is emerging for automobile design, where the weight of the cable is a major factor. Aluminum is significantly less weight than copper, and the short distances required in cars means that resistance is less of a factor.

Table 18-1 may surprise many who believe, in error, that gold is the best conductor. The advantage of gold is its inability to oxidize. This makes it an ideal covering for articles that are exposed to the atmosphere, pollution, or moisture such as the pins in connectors or the connection points on insertable circuit boards. As a conductor, gold does not require annealing, and is often used in integrated circuits since it can be made into very fine wire. But, in normal applications, gold would make a poor conductive material, closer to aluminum in performance than copper.

One other material on the list commonly found in cable is steel. As can be seen, this material is almost ten times the resistance of copper, so many are puzzled by its use. In fact, in the cables that use steel wires, they are coated with a layer of copper, called copper-clad steel and signal passes only on the copper layer, an effect called skin effect that will be discussed in Section 18.2.8. Therefore, the steel wire is used for strength and is not intended to carry signals.

Copper-clad steel is also found in cables where cable pulling strength (pulling tension) is paramount. Then a stranded conductor can be made up of many copper-clad steel strands to maximize strength. Such a cable would compromise basic resistive performance. As is often the case, one can trade a specific attribute for another. In this case, better strength at the cost of higher resistance.

18.2.3 Resistance and Gage Size

In the United States, wire is sized by the American Wire Gage (AWG) method. AWG was based on the previous Brown and Sharpe (B & S) system of wire sizes which dates from 1856. AWG numbers are most common in the United States, and will be referred to throughout this book. The wire most often used in audio ranges from approximately 10AWG to 30AWG, although larger and smaller gage sizes exist. Wire with a small AWG number, such as 4AWG, is very heavy, physically strong but cumbersome, and has very low resistance, while wire of larger numbers, such as 30AWG can be very light weight and fragile, and has high resistance. Resistance is an important factor in determining the appropriate wire size in any circuit. For instance, if an 8ohm loudspeaker is being connected to an amplifier 500ft away through a 19AWG wire, 50% of the power would be dropped in the wire in the form of heat. This is discussed in Section 18.25 regarding loudspeaker cable.

Each time the wire size changes three numbers, such as from 16AWG to 19AWG the resistance doubles. The reverse is also true. With a wire changed from 16AWG to 13AWG, the resistance halves. This also means that combining two identical wires of any given gage decreases the total gage of the combined wires by three units, and reduces the resistance. Two 24AWG wires combined (twisted together) would be 21AWG, for instance. If wires are combined of different gages, the resulting gage can be easily calculated by adding the circular mil area (CMA) shown in Tables 18-2 and 18-3. For instance, if three wires were combined, one 16AWG (2583CMA), one 20AWG (1022CMA) and one 24AWG (404CMA), the total CMA would be 2583 + 1022 + 404 = 4009CMA. Looking in Table 18-1, this numbers falls just under 14AWG. While even number gages are the most common, odd number gages (e.g., 23AWG) can sometimes be found. There are many Category 6 (Cat 6) premise/data cables that are 23AWG, for instance. When required, manufacturers can even produce partial gages. There are coaxial cables with 28.5AWG center conductors. Such specialized gage sizes might require equally special connectors.

There are two basic forms of wire, solid and stranded. A solid conductor is one continuous piece of metal. A stranded conductor is made of multiple smaller wires combined to make a single conductor. Solid wire has slightly lower resistance, with less flexibility and less flex-life (flexes to failure) than stranded wire.

18.2.4 Drawing and Annealing

Copper conductors start life as copper ore in the ground. This ore is mined, refined, and made into bars or rod. Five sixteenth inch copper rod is the most common form used for the making of wire and cable. Copper can be purchased at various purities. These commonly follow the ASTM (American Society for Testing and Materials) standards. Most of the high-purity copper is known as ETP, electrolytic tough pitch. For example, many cable products are manufactured with ASTM B115 ETP. This copper is 99.95% pure. Copper of higher purity can be purchased should the requirement arise. Many consumer audiophiles consider these to be oxygen free, when this term is really a discussion of copper purity and is determined by the number of nines of purity. The cost of the copper rises dramatically with each “9” that is added.

To turn $^5/_{16}$ inch rod into usable wire, the copper rod is drawn through a series of dies. Each time it makes the rod slightly smaller. Eventually you can work the rod down to a very long length of very small wire. To take $^5/_{16}$ inch rod down to a 12AWG wire requires drawing the conductor through eleven different dies. Down to 20AWG requires fifteen dies. To take that wire down to 36AWG requires twenty-eight dies.

The act of drawing the copper work hardens the material making it brittle. The wire is run through an in-line annealing oven, at speeds up to 7000 feet per minute (ft/min), and a temperature of 900 to 1000°F (482 to 537°C). This temperature is not enough to melt the wire, but it is enough to let the copper lose its brittleness and become flexible again, to reverse the work hardening. Annealing is commonly done at the end of the drawing process. However, if the next step requires more flexibility, it can be annealed partway through the drawing process. Some manufacturers draw down the wire and then put the entire roll in an annealing oven. In order to reduce oxygen content, some annealing ovens have inert atmospheres, such as nitrogen. This increases the purity of the copper by reducing the oxygen content. But in-line annealing is more consistent than a whole roll in an oven.

Lack of annealing, or insufficient annealing time or temperature, can produce a conductor which is stiff, brittle, and prone to failure. With batch annealing, the inner windings in a roll may not be heated as effectively as the outer windings. Cables made in other countries may not have sufficient purity for high-performance applications. Poor quality copper, or poor annealing, are very hard to tell from initial visual inspection but often shows up during or after installation.

18.2.5 Plating and Tinning

Much of the wire manufactured is plated with a layer of tin. This can also be done in-line with the drawing and annealing by electroplating a layer on the wire. Tinning makes the wire especially resistant to pollutants, chemicals, salt (as in marine applications). But such a plated conductor is not appropriate for high-frequency applications where the signal travels on the skin of the conductor, called *skin effect*. In that case, bare copper conductors are used. The surface of a conductor used for high frequencies is a major factor in good performance and should have a mirror finish on that surface. Wires are occasionally plated with silver. While silver is slightly more conductive, its real advantage is that silver oxide is the same resistance as bare silver. This is not true with copper, where copper oxide is a semiconductor. Therefore, where reactions with a copper wire are predicted, silver plating may help preserve performance. So silver plating is sometimes used for marine cables, or cables used in similar outdoor environments.

Some plastics, when extruded (melted) onto wires, can chemically affect the copper. This is common, for instance, with an insulation of extruded TFE (tetrafluoroethylene), a form of Teflon™. Wires used inside these cables are often silver plated or silver-clad. Any oxidizing caused by the extrusion process therefore has no effect on performance. Of course, just the cost of silver alone makes any silver-plated conductor significantly more expensive than bare copper.

18.2.6 Conductor Parameters

Table 18-2 shows various parameters for solid wire from 4AWG to 40AWG. Table 18-3 shows the same parameters for stranded wire. Note that the resistance of a specific gage of solid wire is lower than stranded wire of the same gage. This is because the stranded wire is not completely conductive; there are spaces (interstices) between the strands. It takes a larger stranded wire to equal the resistance of a solid wire.

Table 18-2 Parameters for Solid Wire from 4AWG to 40AWG

AWG	Nominal Diameter	CMA (×1000)	Bare lbs/ft	Ω/ 1000ft	Current A	MM² Equivalent
4	0.2043	41.7	0.12636	0.25	59.57	21.1
5	0.1819	33.1	0.10020	0.31	47.29	16.8
6	0.162	26.3	0.07949	0.4	37.57	13.3
7	0.1443	20.8	0.06301	0.5	29.71	10.6
8	0.1285	16.5	0.04998	0.63	23.57	8.37
9	0.1144	13.1	0.03964	0.8	18.71	6.63
10	0.1019	10.4	0.03143	1	14.86	5.26
11	0.0907	8.23	0.02493	1.26	11.76	4.17
12	0.0808	6.53	0.01977	1.6	9.33	3.31
13	0.075	5.18	0.01567	2.01	7.40	2.62
14	0.0641	4.11	0.01243	2.54	5.87	2.08
15	0.0571	3.26	0.00986	3.2	4.66	1.65
16	0.0508	2.58	0.00782	4.03	3.69	1.31
17	0.0453	2.05	0.00620	5.1	2.93	1.04
18	0.0403	1.62	0.00492	6.4	2.31	0.823
19	0.0359	1.29	0.00390	8.1	1.84	0.653
20	0.032	1.02	0.00309	10.1	1.46	0.519
21	0.0285	0.81	0.00245	12.8	1.16	0.412
22	0.0254	0.642	0.00195	16.2	0.92	0.324
23	0.0226	0.51	0.00154	20.3	0.73	0.259
24	0.0201	0.404	0.00122	25.7	0.58	0.205
25	0.0179	0.32	0.00097	32.4	0.46	0.162
26	0.0159	0.253	0.00077	41	0.36	0.128
27	0.0142	0.202	0.00061	51.4	0.29	0.102
28	0.0126	0.159	0.00048	65.3	0.23	0.08
29	0.0113	0.127	0.00038	81.2	0.18	0.0643
30	0.01	0.1	0.00030	104	0.14	0.0507
31	0.0089	0.0797	0.00024	131	0.11	0.0401
32	0.008	0.064	0.00019	162	0.09	0.0324
33	0.0071	0.0504	0.00015	206	0.07	0.0255
34	0.0063	0.0398	0.00012	261	0.06	0.0201
35	0.0056	0.0315	0.00010	331	0.05	0.0159
36	0.005	0.025	0.00008	415	0.04	0.0127
37	0.0045	0.0203	0.00006	512	0.03	0.0103
38	0.004	0.016	0.00005	648	0.02	0.0081
39	0.0035	0.0123	0.00004	847	0.02	0.0062
40	0.003	0.0096	0.00003	1080	0.01	0.0049

Table 18-3 Parameters for ASTM Class B Stranded Wires from 4AWG to 40AWG

AWG	Nominal Diameter	CMA (×1000)	Bare lbs/ft	Ω/ 1000ft	Current A*	MM² Equivalent
4	0.232	53.824	0.12936	0.253	59.63	27.273
5	0.206	42.436	0.10320	0.323	47.27	21.503
6	0.184	33.856	0.08249	0.408	37.49	17.155
7	0.164	26.896	0.06601	0.514	29.75	13.628
8	0.146	21.316	0.05298	0.648	23.59	10.801
9	0.13	16.9	0.04264	0.816	18.70	8.563
10	0.116	13.456	0.03316	1.03	14.83	6.818
11	0.103	10.609	0.02867	1.297	11.75	5.376
12	0.0915	8.372	0.02085	1.635	9.33	4.242
13	0.0816	6.659	0.01808	2.063	8.04	3.374
14	0.0727	5.285	0.01313	2.73	5.87	2.678
15	0.0647	4.186	0.01139	3.29	4.66	2.121
16	0.0576	3.318	0.00824	4.35	3.69	1.681
17	0.0513	2.632	0.00713	5.25	2.93	1.334
18	0.0456	2.079	0.00518	6.92	2.32	1.053
19	0.0407	1.656	0.00484	8.25	1.84	0.839
20	0.0362	1.31	0.00326	10.9	1.46	0.664
21	0.0323	1.043	0.00284	13.19	1.16	0.528
22	0.0287	0.824	0.00204	17.5	0.92	0.418
23	0.0256	0.655	0.00176	20.99	0.73	0.332
24	0.0228	0.52	0.00129	27.7	0.58	0.263
25	0.0203	0.412	0.01125	33.01	0.46	0.209
26	0.018	0.324	0.00081	44.4	0.36	0.164
27	0.0161	0.259	0.00064	55.6	0.29	0.131
28	0.0143	0.204	0.00051	70.7	0.23	0.103
29	0.0128	0.164	0.00045	83.99	0.18	0.083
30	0.0113	0.128	0.00032	112	0.14	0.0649
31	0.011	0.121	0.00020	136.1	0.11	0.0613
32	0.009	0.081	0.00020	164.1	0.09	0.041
33	0.00825	0.068	0.00017	219.17	0.07	0.0345
34	0.0075	0.056	0.00013	260.9	0.06	0.0284
35	0.00675	0.046	0.00011	335.96	0.04	0.0233
36	0.006	0.036	0.00008	414.8	0.04	0.0182
37	0.00525	0.028	0.00006	578.7	0.03	0.0142
38	0.0045	0.02	0.00005	658.5	0.02	0.0101
39	0.00375	0.014	0.00004	876.7	0.02	0.0071
40	0.003	0.009	0.00003	1028.8	0.01	0.0046

* For both solid and stranded wire, amperage is calculated at 1A for each 700CMA. See also Section 18.2.9.

Stranded Cables.

Stranded cables are more flexible, and have greater flex-life (flexes to failure) than solid wire. Table 18-4 shows some suggested construction values. The two numbers (65 × 34, for example) show the number of strands (65) and the gage size of each strand (34) for each variation in flexing.

Table 18-4 Suggested Conductor Strandings for Various Degrees of Flexing Severity

Typical Applications	AWG	mm	AWG	mm
	12 AWG		14 AWG	
Fixed Service (Hook-Up Wire Cable in Raceway)	19 × 25	19 × 0.455	Solid 19 × 27	19 × 0.361
Moderate Flexing (Frequently Disturbed for Maintenance)	65 × 30	65 × 0.254	19 × 27 41 × 30	19 × 0.361 41 × 0.254
Severe Flexing (Microphones and Test Prods)	165 × 34	165 × 0.160	104 × 34	104 × 0.160
	16 AWG		18 AWG	
Fixed Service (Hook-Up Wire Cable in Raceway)	Solid 19 × 29	19 × 0.287	Solid 7 × 26 16 × 30	7 × 0.404 16 × 0.254
Moderate Flexing (Frequently Disturbed for Maintenance)	19 × 29 26 × 30	19 × 0.287 26 × 0.254	16 × 30 41 × 34	16 × 0.254 41 × 0.160
Severe Flexing (Microphones, Test Prods)	65 × 34 104 × 36	65 × 0.160 104 × 0.127	41 × 34 65 × 36	41 × 0.160 65 × 0.127
	20 AWG		22 AWG	
Fixed Service (Hook-Up Wire Cable in Raceway)	Solid 7 × 28 10 × 30	7 × 0.320 10 × 0.254	Solid 7 × 30	7 × 0.254
Moderate Flexing (Frequently Distributed for Maintenance)	7 × 28 10 × 30 19 × 32 26 × 34	7 × 0.320 10 × 0.254 19 × 0.203 26 × 0.160	7 × 30 19 × 34	7 × 0.254 19 × 0.160
Severe Flexing (Microphones, Test Prods)	26 × 34 42 × 36	26 × 0.160 42 × 0.127	19 × 34 26 × 36	19 × 0.160 26 × 0.127
	24 AWG		26 AWG	
Fixed Service (Hook-Up Wire Cable in Raceway)	Solid 7 × 32	7 × 0.203	Solid 7 × 34	7 × 0.160
Moderate Flexing (Frequently Disturbed for Maintenance)	7 × 32 10 × 34	7 × 0.203 10 × 0.160	7 × 34	7 × 0.160
Severe Flexing (Microphones, Test Prods)	19 × 36 45 × 40	19 × 0.127 45 × 0.079	7 × 34 10 × 36	7 × 0.160 10 × 0.127

Courtesy Belden.

18.2.7 Pulling Tension

Pulling tension must be adhered to so the cable will not be permanently elongated. The pulling tension for annealed copper conductors is shown in Table 18-5.

Multiconductor cable pulling tension can be determined by multiplying the total number of conductors by the appropriate value. For twisted pair cables, there are two wires per pair. For shielded twisted pair cables, with foil shields, there is a drain wire that must be included in the calculations. Be cautious: the drain wire can sometimes be smaller gage than the conductors in the pair. The pulling tension of coaxial cables or other cables that are not multiple conductors is much harder to calculate. Consult the manufacturer for the required pulling tension.

Table 18-5. Pulling Tension for Annealed Copper Conductors

24AWG	5.0lbs
22AWG	7.5lbs
20AWG	12.0lbs
18AWG	19.5lbs
16AWG	31.0lbs
14 AWG	49.0lbs
12 AWG	79.0lbs

18.2.8 Skin Effect

As the frequency of the signal increases on a wire, the signal travels closer to the surface of the conductor. Since very little of the area of the center conductor is used at high frequencies, some cable is made with a copper clad-steel core center conductor. These are known as copper-clad, copper-covered, or Copper-weld™ and are usually used by CATV/broadband service providers.

Copper-clad steel is stronger than copper cable so it can more easily withstand pulling during installation, or wind, ice, and other outside elements after installation. For instance, a copper-clad 18AWG coaxial cable has a pull strength of 102lbs while a solid copper 18AWG coax would have a pull strength of 69lbs. The main disadvantage is that steel is not a good conductor below 50MHz, between four and seven times the resistance of copper, depending on the thickness of the copper layer.

This is a problem where signals are below 50MHz such as DOCSIS data delivery, or VOD (video-on-demand) signals which are coming from the home to the provider. When installing cable in a system, it is better to use solid copper cable so it can be used at low frequencies as well as high frequencies.

This is also why copper-clad conductors are not appropriate for any application below 50MHz, such as baseband video, CCTV, analog, or digital audio. Copper-clad is also not appropriate for applications such as SDI or HD-SDI video, and similar signals where a significant portion of the data is below 50MHz.

Copper-clad is often specified incorrectly for satellite dish installations. In this application, dc power is fed to the antenna, and the broadband intermediate frequency signal comes from the dish the other way. While the cable might have the appropriate loss characteristics for satellite frequencies, the steel center makes it a bad choice for the dc power and such a cable can have dramatic distance limitations.

Copper-clad is often specified incorrectly for satellite dish installations. In this application, dc power is fed to the antenna, and the broadband intermediate frequency signal comes from the dish the other way. While the cable might have the appropriate loss characteristics for satellite frequencies, the steel center makes it a

bad choice for the dc power and such a cable can have dramatic distance limitations.

The skin depth for copper conductors can be calculated with the equation

$$D = \frac{2.61}{\sqrt{f}} \qquad (18\text{-}2)$$

where,

D is the skin depth in in,

f is the frequency in Hz.

Table 18-6 compares the actual skin depth and percent of the center conductor actually used in an RG-6 cable. The skin depth always remains the same no matter what the thickness of the wire is. The only thing that changes is the percent of the conductor utilized. Determining the percent of the conductor utilized requires using two times the skin depth because we are comparing the diameter of the conductor to its depth.

Table 18-6. Skin Depths at Various Frequencies

Frequency	Skin Depth in Inches	% Used of #18AWG Conductor
1kHz	0.082500	100
10 kHz	0.026100	100
100kHz	0.008280	41
1 MHz	0.002610	13
10MHz	0.000825	4.1
100MHz	0.000261	1.3
1 GHz	0.0000825	0.41
10 GHz	0.0000261	0.13

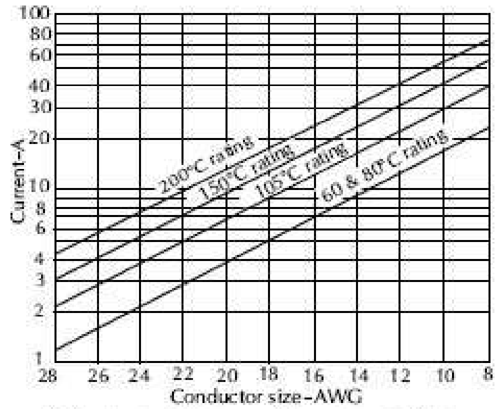
As can be seen, by the time the frequencies are high, the depth of the signal on the skin can easily be micro-inches. For signals in that range, such as high-definition video signals, for example, this means that the surface of the wire is as critical as the wire itself. Therefore, conductors intended to carry high frequencies should have a mirror finish.

Since the resistance of the wire at these high frequencies is of no consequence, it is sometimes asked why larger conductors go farther. The reason is that the *surface area*, the *skin*, on a wire is greater as the wire gets larger in size.

Further, some conductors have a tin layer to help prevent corrosion. These cables are obviously not intended for use at frequencies above just a few megahertz, or a significant portion of the signal would be traveling in the tin layer. Tin is not an especially good conductor as can be seen in Table 18-1.

18.2.9 Current Capacity

For conductors that will carry large amounts of electrical flow, large amperage or current from point to point, a general chart has been made to simplify the current carrying capacity of each conductor. To use the current capacity chart in Fig. 18-1, first determine conductor gage, insulation and jacket temperature rating, and number of conductors from the applicable product description for the cable of interest. These can usually be obtained from a manufacturer's website or catalog.



# Conductors*	Factor	# Conductors*	Factor
1	1.6	6–15	0.7
2–3	1.0	16–30	0.5
4–5	0.8		

* do not count shields unless used as a conductor

Figure 18-1. Current ratings for electronic cable. Courtesy Belden.

Next, find the current value on the chart for the proper temperature rating and conductor size. To calculate the maximum current rating/conductor multiply the chart value by the appropriate conductor factor. The chart assumes the cable is surrounded by still air at an ambient temperature of 25°C (77°F). Current values are in amperes (rms) and are valid for copper conductors only. The maximum continuous current rating for an electronic cable is limited by conductor size, number of conductors contained within the cable, maximum temperature rating of the insulation on the conductors, and environment conditions such as ambient temperature and air flow. The four lines marked with temperatures apply to different insulation plastics and their melting point. Consult the manufacturer's Web site or catalog for the maximum insulation or jacket temperature.

The current ratings of Fig. 18-1 are intended as general guidelines for low-power electronic communications and control applications. Current ratings for high-power applications generally are set by regulatory agencies such as Underwriters Laboratories (UL), Canadian Standards Association (CSA), National Electrical Code

(NEC), and others and should be used before final installation.

Table 310-15(b)(2)(a) of the NEC contains amperage adjustment factors for whenever more than three current carrying conductors are in a conduit or raceway.

Section 240-3 of the NEC provides requirements for overload protection for conductors other than flexible cords and fixture wires. Section 240-3(d), Small Conductors, states that 14 AWG to 10 AWG conductors require a maximum protective overcurrent device with a rating no higher than the current rating listed in the 60°C column. These currents are 15A for 14 AWG copper wire, 20A for 12 AWG copper wire, and 30A for 10 AWG copper wire. These values are familiar as the breaker ratings for commercial installations.

When connecting wire to a terminal strip or another wire etc., the temperature rise in the connections must also be taken into account. Often the circuit is not limited by the current carrying capacity of the wire but of the termination point.

Wire Current Ratings

Current carrying capacity of wire is controlled by the NEC, particularly in Table 310-16, Table 310-15(b)(2)(a), and Section 240-3.

Table 310-16 of the NEC shows the maximum current carrying capacity for insulated conductors rated from 0 to 2000V, including copper and aluminum conductors. Each conductor amperage is given for three temperatures: 60°C, 75°C, and 90°C. Copper doesn't melt until almost 2000°, so the current limit on a copper wire is not the melting point of the wire but the melting point of the insulation. This number is listed by most manufacturers in their catalog or on their Web site. For instance, PVC (polyvinyl chloride) can be formulated to withstand temperatures from 60°C to as high as 105°C. The materials won't melt right at the specified temperature, but may begin to fail certain tests, such as cracking when bent.

18.3 Insulation

Wire can be bare, often called *bus bar* or *bus wire*, but is most often insulated. It is covered with a non-conducting material. Early insulations included cotton or silk woven around the conductor, or even paper. Cotton-covered house wiring can still be found in perfect operating condition in old houses. Today, most insulation materials are either some kind of rubber or some kind of plastic. The material chosen should be listed in the manufacturer's catalog with each cable type. Table 18-7 lists some of the rubber-based materials with their properties. Table 18-8 lists the properties of various plastics. The ratings in both tables are based on average performance of general-purpose compounds. Any given property can usually be improved by the use of selective compounding.

Table 18-7. Comparative Properties of Rubber Insulation

Properties	Rubber	Neoprene	Hypalon (Chlorosulfonated Polyethylene)	EPDM (Ethylene Propylene Diene Monomer)	Silicone
Oxidation Resistance	F	G	E	G	E
Heat Resistance	F	G	E	E	O
Oil Resistance	P	G	G	F	F-G
Low Temperature Flexibility	G	F-G	F	G-E	O
Weather, Sun Resistance	F	G	E	E	O
Ozone Resistance	P	G	E	E	O
Abrasion Resistance	E	G-E	G	G	P
Electrical Properties	E	P	G	E	O
Flame Resistance	P	G	G	P	F-G
Nuclear Radiation Resistance	F	F-G	G	G	E
Water Resistance	G	E	G-E	G-E	G-E
Acid Resistance	F-G	G	E	G-E	F-G
Alkali Resistance	F-G	G	E	G-E	F-G
Gasoline, Kerosene, etc. (Aliphatic Hydrocarbons) Resistance	P	G	F	P	P-F
Benzol, Toluol, etc.(Aromatic Hydrocarbons) Resistance	P	P-F	F	F	P
Degreaser Solvents (Halogenated Hydrocarbons) Resistance	P	P	P-F	P	P-G
Alcohol Resistance	G	F	G	P	G

P = poor, F = fair, G = good, E = excellent, O = outstanding

Courtesy Belden

Table 18-8. Dielectric Constant

Dielectric Constant	Material	Note
1	Vacuum	By definition
1.0167	Air	Very close to 1
1.35	Foam, Air-filled Plastic	Current technological limit
2.1	Solid Teflon™	Best solid plastic
2.3	Solid Polyethylene	Most common plastic
3.5-6.5	Solid Polyvinyl Chloride	Low price, easy to work with

18.3.1 Plastics and Dielectric Constant

Table 18-9 is a list of various insulation materials with details on performance, requirements, and special advantages. Insulation, when used on a cable intended to carry a signal, is often referred to as a *dielectric*. The performance of any material, its ability to insulate with minimal effect to the signal running on the cable is called the *dielectric constant* and can be measured in a laboratory. Table 18-9 shows some standard numbers as a point of reference.

18.3.2 Wire Insulation Characteristics

The key difference between rubber compounds and plastic compounds is their recyclability. Plastic materials can be ground up, and re-melted into other objects. Polyethylene, for instance, can be recycled into plastic bottles, grocery bags, or even park benches. And, should the need arise, these objects could themselves be ground up and turned back into wire insulation, or many other uses. The term *thermoplastic* means changed by heat and is the source of the common term *plastic*.

Rubber compounds, on the other hand, are thermoset. That is, once they are made, they are set, and the process cannot be reversed. Rubber, and its family, is cured in a process sometimes called *vulcanizing*. These compounds cannot be ground up and recycled into new products. There are natural rubber compounds (such as latex-based rubber) and artificial, chemical-based rubber compounds such as EPDM (ethylene-propylene-diene monomer).

Table 18-9. Comparative Properties of Plastic Insulation

Properties	PVC	Low-Density Polyethylene	Cellular Polyethylene	High-Density Polyethylene	Polyethylene	Polyurethane	Nylon	Teflon®
Oxidation Resistance	E	E	E	E	E	E	E	O
Heat Resistance	G-E	G	G	E	E	G	E	O
Oil Resistance	F	G	G	G-E	F	E	E	O
Low Temperature Flexibility	P-G	G-E	E	E	P	G	G	O
Weather, Sun Resistance	G-E	E	E	E	E	G	E	O
Ozone Resistance	E	E	E	E	E	E	E	E
Abrasion Resistance	F-G	F-G	F	E	F-G	O	E	E
Electrical Properties	F-G	E	E	E	E	P	P	E
Flame Resistance	E	P	P	P	P	P	P	O
Nuclear Radiation Resistance	G	G	G	G	F	G	F-G	P
Water Resistance	E	E	E	E	E	P-G	P-F	E
Acid Resistance	G-E	G-E	G-E	G-E	E	F	P-F	E
Alkali Resistance	G-E	G-E	G-E	G-E	E	F	E	E
Gasoline, Kerosene, etc. (Aliphatic Hydrocarbons) Resistance	P	P-F	P-F	P-F	P-F	G	G	E
Benzol, Toluol, etc. (Aromatic Hydrocarbons) Resistance	P-F	P	P	P	P-F	P	G	E
Degreaser Solvents (Halogenated Hydrocarbons) Resistance	P-F	P	P	P	P	P	G	E
Alcohol Resistance	G-E	E	E	E	E	P	P	E

P = poor, F = fair, G = good, E = excellent, O = outstanding

Courtesy Belden.

The vast majority of wire and cable insulations are plastic-based compounds. Rubber, while it is extremely rugged, is considerably more expensive that most plastics, so there are fewer and fewer manufacturers offering rubber-based products. These materials, both rubber and plastic, are used in two applications with cable. The first application is insulation of the conductor(s) inside the cable. The second is as a jacket material to protect the contents of the cable.

18.4 Jackets

The jacket characteristics of cable have a large effect on its ruggedness and the effect of environment. A key consideration is often flexibility, especially at low temperatures. Audio and broadcast cables are manufactured in a wide selection of standard jacketing materials. Special compounds and variations of standard compounds are used to meet critical audio and broadcast application requirements and unusual environmental conditions. Proper matching of cable jackets to their working environment can prevent deterioration due to intense heat and cold, sunlight, mechanical abuse, impact, and crowd or vehicle traffic.

18.5 Plastics

Plastic is a shortened version of the term *thermoplastic*. *Thermo* means heat, *plastic* means change. Thermoplastic materials can be changed by heat. They can be melted and extruded into other shapes. They can be extruded around wires, for instance, forming an insulative (nonconductive) layer. There are many forms of plastic. Below is a list of the most common varieties used in the manufacture of wire and cable.

18.5.1 Vinyl

Vinyl is sometimes referred to as PVC or *polyvinyl chloride*, and is a chemical compound invented in 1928 by Dr. Waldo Semon (USA). Extremely high or low temperature properties cannot be found in one formulation, therefore, formulations may have −55°C to +105°C (−67°F to +221°F) rating while other common vinyls may have −20°C to +60°C (−4°F to +140°F). The many varieties of vinyl also differ in pliability and electrical properties fitting a multitude of applications. The price range can vary accordingly. Typical dielectric constant values can vary from 3.5 at 1000Hz to 6.5 at 60Hz, making it a poor choice if high performance is required. PVC is one of the least expensive compounds, and one of the easiest to work with. Therefore, PVC is used with many cables that do not require high performance, or where cost of materials is a major factor. PVC is easy to color, and can be quite flexible, although it is not very rugged. In high-performance cables, PVC is often used as the jacket material, but not inside the cable.

18.5.2 Polyethylene

Polyethylene, invented by accident in 1933 by E.W. Fawcett and R.O. Gibson (Great Britain), is a very good insulation in terms of electrical properties. It has a low dielectric constant value over all frequencies and very high insulation resistance. In terms of flexibility, polyethylene can be rated stiff to very hard depending on molecular weight and density. Low density is the most flexible and high density high molecular weight formulations are very hard. Moisture resistance is rated excellent. Correct brown and black formulations have excellent sunlight resistance. The dielectric constant is 2.3 for solid insulation and as low as 1.35 for gas-injected foam cellular designs. Polyethylene is the most common plastic worldwide.

18.5.3 Teflon®

Invented in 1937 by Roy Plunkett (USA) at DuPont, Teflon has excellent electrical properties, temperature range, and chemical resistance. It is not suitable where subjected to nuclear radiation, and it does not have good high voltage characteristics. FEP (fluorinated ethylene-propylene) Teflon is extrudable in a manner similar to vinyl and polyethylene, therefore, long wire and cable lengths are available. TFE (tetrafluoroethylene) Teflon is extrudable in a hydraulic ram-type process and lengths are limited due to amount of material in the ram, thickness of the insulation, and core size. TFE must be extruded over silver-coated or nickel-coated wire. The nickel and silver-coated designs are rated +260°C and +200°C maximum (500°F and 392°F), respectively, which is the highest temperature for common plastics. The cost of Teflon is approximately eight to ten times more per pound than vinyl insulations. The dielectric constant for solid Teflon is 2.1, the lowest of all solid plastics. Foam Teflon (FEP) has a dielectric constant as low as 1.35. Teflon is produced by and a trademark of DuPont Corporation.

18.5.4 Polypropylene

Polypropylene is similar in electrical properties to polyethylene and is primarily used as an insulation material. Typically, it is harder than polyethylene, which makes it suitable for thin wall insulations. UL maximum temperature rating may be 60°C or 80°C (140°F or 176°F). The dielectric constant is 2.25 for solid and 1.55 for cellular designs.

18.6 Thermoset Compounds

As the name implies, *thermoset* compounds are produced by heat (thermo) but are set. That is, the process cannot be reversed as in thermoplastics. They cannot be recycled into new products as thermoplastic materials can.

18.6.1 Silicone

Silicone is a very soft insulation which has a temperature range from −80°C to +200°C (−112°F to +392°F). It has excellent electrical properties plus ozone resistance, low moisture absorption, weather resistance, and radiation resistance. It typically has low mechanical strength and poor scuff resistance. Silicone is seldom used because it is very expensive.

18.6.2 Neoprene

Neoprene has a maximum temperature range from −55°C to +90°C (−67°F to +194°F). The actual range depends on the formulation used. Neoprene is both oil and sunlight resistant making it ideal for many outdoor applications. The most stable colors are black, dark brown, and gray. The electrical properties are not as good as other insulation material; therefore, thicker insulation must be used for the same insulation.

18.6.3 Rubber

The description of *rubber* normally includes natural rubber and styrene-butadiene rubber (SBR) compounds. Both can be used for insulation and jackets. There are many formulations of these basic materials and each formulation is for a specific application. Some formulations are suitable for −55°C (−67°F) minimum while others are suitable for +75°C (+167°F) maximum. Rubber jacketing compounds feature exceptional durability for extended cable life. They withstand high-impact and abrasive conditions better than PVC and are resistant to degradation or penetration by water, alkali, or acid. They have excellent heat resistant properties, and also provide greater cable flexibility in cold temperatures.

18.6.4 EPDM

EPDM stands for *ethylene-propylene-diene monomer*. It was invented by Dr. Waldo Semon in 1927 (see Section 18.5.1). It is extremely rugged, like natural rubber, but can be created from petroleum byproducts ethylene and propylene gas. Like rubber, it is a thermoset compound.

18.7 Single Conductor

Single conductor wire starts with a single wire, either solid or stranded. It can be bare, sometimes called *buss bar*, or can be jacketed. There is no actual limit to how small, or how large, a conductor could be. Choice of size (AWG) will be based on application and the current or wattage delivery required. If jacketed, the choice of jacket can be based on performance, ruggedness, flexibility, melting point, or any other requirement.

There is no single conductor plenum rating because the NEC (National Electrical Code) only applies to cables, more than one conductor. However, Articles 300 and 310 of the NEC are sometimes cited when installing single conductor wire for grounds and similar applications.

18.8 Multiconductor

Bundles of two or more insulated wires are considered multiconductor cable. Besides the requirements for each conductor, there is often an overall jacket, chosen for whatever properties would be appropriate for a particular application.

There are specialized multiconductor cables, such as power cordage used to deliver ac power from a wall outlet (or other source) to a device. There are UL safety ratings on such a cable to assure users will not be harmed. Some of these safety ratings (such as “Class 1”) are now being applied to high-power loudspeaker cables, see Section 18.27.

There are other multiconductor applications such as VFD (variable frequency drive) cables, specially formulated to minimize standing waves and arcing discharge when running variable frequency motors. Since a multi-conductor cable is not divided into pairs, resistance is still the major performance parameter to be determined, although reactions between conductors (as in VFD) can also be considered.

Multiconductor Insulation Color Codes

The wire insulation colors help trace conductors or conductor pairs. There are many color tables; Table 18-10 is one example.

18.9 Pairs and Balanced Lines

Twisting two insulated wires together makes a twisted pair. Since two conductive paths are needed to make a circuit, twisted pairs give users an easy way to connect power or signals from point to point. Sometimes the insulation color is different to identify each wire in each pair. Pairs can have dramatically better performance than multiconductor cables because pairs can be driven as a balanced line.

Table 18-10. Color Code for Nonpaired Cables per ICEA #2 and #2R

Conductor	Color	Conductor	Color	Conductor	Color	Conductor	Color
1st	Black	14th	Green/White	27th	Blue/Blk/Wht	40th	Red/Wht/Grn
2nd	White	15th	Blue/White	28th	Blk/Red/Grn	41st	Grn/Wht/Blue
3rd	Red	16th	Black/Red	29th	Wht/Red/Grn	42nd	Org/Red.Grn
4th	Green	17th	White/Red	30th	Red/Blk/Grn	43rd	Blue/Red/Grn
5th	Orange	18th	Orange/Red	31st	Grn/Blk/Org	44th	Blk/Wht/Blue
6th	Blue	19th	Blue/Red	32nd	Org/Blk/Grn	45th	Wht/Blk/Blue
7th	White/Black	20th	Red/Green	33rd	Blue/Wht/Org	46th	Red/Wht/Blue
8th	Red/Black	21st	Orange/Green	34th	Blk/Wht/Org	47th	Grn/Orn/Red
9th	Green/Black	22nd	Blk/Wht/Red	35th	Wht/Red/Org	48th	Org/Red/Blue
10th	Orange/Black	23rd	Wht/Blk/Red	36th	Org/Wht/Blue	49th	Blue/Red/Org
11th	Blue/Black	24th	Red/Blk/Wht	37th	Wht/Red/Blue	50th	Blk/Org/Red
12th	Black/White	25th	Grn/Blk/Wht	38th	Blk/Wht/Grn		
13th	Red/White	26th	Org/Blk/Wht	39th	Wht/Blk/Grn		

Courtesy Belden.

A balanced line is a configuration where the two wires are electrically identical. The electrical performance is referred to ground, the zero point in circuit design. Balanced lines reject noise, from low frequencies, such as 50/60Hz power line noise, up to radio frequency signals in the Megahertz, or even higher.

When the two conductors are electrically identical, or close to identical, there are many other parameters, besides resistance, that come into play. These include capacitance, inductance, and impedance. And when we get to high-frequency pairs, such as data cables, we even measure the variations in resistance (resistance unbalance), variations in capacitance (capacitance unbalance), or even variations in impedance (return loss). Each of these has a section farther on in this chapter.

Most broadcast products use the “Resistor Color Code” because this color code has been memorized by most designers and installers. Black, which indicates ‘0’ in the original color code, is shifted to indicate color ‘10’, Table 18-11.

Table 18-11. Standard Resistor Color Code

Number	Color	Number	Color
1	Brown	6	Blue
2	Red	7	Violet
3	Orange	8	Gray
4	Yellow	9	White
5	Green	10	Black

After ten colors, there is no common standard, and other color codes are often employed, as shown in Table 18-10.

These are often used as the jacket colors for audio and video cable, especially where there are bundled cables. In single cables, the colors have no standard meaning, with the exception of red, green and blue, which sometimes indicate RGB component analog video.

Balanced lines work because they have a transformer at each end, a device made of two coils of wire wound together. Many modern devices now use circuits that act electrically the same as a transformer, an effect called *active balancing*. The highest-quality transformers can be extremely expensive, so high-performing balanced-line chips have been improving, some getting very close to the coils-of-wire performance.

It should be noted that virtually all professional installations use twisted pairs for audio because of their noise rejection properties. In the consumer world, the cable has one hot connection and a grounded shield around it and is called an unbalanced cable. These cables are effective for only short distances and have no other inherent noise rejection besides the shield itself.

18.9.1 Multipair

As the name implies, multipair cables contain more than one pair. Sometimes referred to as *multicore* cables, these can just be grouped bare pairs, or each pair could be individually jacketed, or each pair could be shielded (shielding is outlined below), or the pairs could even be individually shielded and jacketed. All of these options are easily available. Where there is an overall jacket, or individual jackets for each pair, the jacket material for each pair is chosen with regard to price, flexibility, ruggedness, color, and any other parameter required.

It should be noted that the jackets on pairs, or the overall jacket, has almost no effect on the performance of the pairs. One could make a case that, with individually jacketed pairs, the jacket moves the pairs apart and therefore improves crosstalk between pairs. It is also possible that poorly extruded jackets could leak the chemicals that make up the jacket into the pair they are protecting, an effect called *compound migration*, and therefore affect the performance of the pair.

Table 18-12 shows a common color code for paired cables where they are simply a bundle of pairs. The color coding is only to identify the pair and the coloring of the insulation has no effect on performance. If this cable were individually jacketed pairs, it would be likely that the two wires in the pair would be identical colors such as all black-and-red, and the jackets would use different colors to identify them as shown in Table 18-13.

18.9.2 Analog Multipair Snake Cable

Originally designed for the broadcast industry, hard-wire multipair audio snake cables feature individually shielded pairs for optimum noise rejection, and sometimes with individual jackets on each pair for improved physical protection and are color-coded for identification. These cables are ideal, carrying multiple line-level or microphone-level signals. They will also interconnect audio components such as multichannel mixers and consoles for recording studios, radio and television stations, postproduction facilities, and sound system installations. Snakes offer the following features:

Table 18-12. Color Codes for Paired Cables (Belden Standard)

Pair No.	Color Combination	Pair No.	Color Combination	Pair No.	Color Combination	Pair No.	Color Combination
1	Black/Red	11	Red/Yellow	21	White/Brown	31	Purple/White
2	Black/White	12	Red/Brown	22	White/Orange	32	Purple/Dark Green
3	Black Green	13	Red/Orange	23	Blue/Yellow	33	Purple/Light Blue
4	Black/Blue	14	Green/White	24	Blue/Brown	34	Purple/Yellow
5	Black/Yellow	15	Green/Blue	25	Blue/Orange	35	Purple/Brown
6	Black/Brown	16	Green/Yellow	26	Brown/Yellow	36	Purple/Black
7	Black/Orange	17	Green/Brown	27	Brown/Orange	37	Gray/White
8	Red/White	18	Green/Orange	28	Orange/Yellow		
9	Red/Green	19	White/Blue	29	Purple/Orange		
10	Red/Blue	20	White/Yellow	30	Purple/Red		

Courtesy Belden.

- A variety insulation materials, for low capacitance, ruggedness, or fire ratings.
- Spiral/serve, braid, French Braid™, or foil shields.
- Jacket and insulation material to meet ruggedness or NEC flame requirements.
- High temperature resistance in some compounds.
- Cold temperature pliability in some compounds.
- Low-profile appearance, based mostly on the gage of the wires, but also on the insulation.
- Some feature overall shields to reduce crosstalk and facilitate star grounding.

- Allows easier and cheaper installs than using multiple single channel cables.

Snakes come with various terminations and can be specified to meet the consumer's needs. Common terminations are male or female XLR (microphone) connectors and ¼ inch male stereo connectors on one end, and either a junction box with male or female XLR connectors and ¼ inch stereo connectors or pigtails with female XLR connectors and ¼ inch connectors on the other end. There are hundreds of other connector types that have been used for audio applications. There are also punch-down terminals available set up to accept the three wires (twisted pair and drain wire) commonly used.

For stage applications, multipair individually shielded snake cables feature lightweight and small diameter construction, making them ideal for use as portable audio snakes. Individually shielded and jacketed pairs are easier to install with less wiring errors. In areas that subscribe to the NEC guidelines, the need for conduit in studios is eliminated when CM-rated snake cable is used through walls between rooms. Vertically between floors, snakes rated CMR (riser) do not need conduit. In plenum areas (raised floors, drop ceilings) CMP, plenum rated snake cables can be used without conduit. Color codes for snakes are given in Table 18-13.

18.9.3 High Frequency Pairs

Twisted pairs were original conceived to carry low-frequency signals, such as telephone audio. Beginning in the 1970s research and development was producing cables such as twinax that had reasonable performance to a bandwidth of several megahertz. IBM

Type 1 was the breakthrough product that proved that twisted-pairs could indeed carry data. This led directly to the Category premise/data cable of today.

There are now myriad forms of high-frequency, high-data rate cable including DVI, USB, HDMI, IEEE 1394 FireWire, and many others. All of these are commonly used to transport audio and video signals, Table 18-14.

18.9.3.1 DVI

DVI (Digital Visual Interface) is used extensively in the computer-monitor interface market for flat panel LCD monitors.

The DVI connection between local monitors and computers includes a serial digital interface and a parallel interface format, somewhat like combining the broadcast serial digital and parallel digital interfaces.

Transmission of the TMDS (transition minimized differential signaling) format combines four differential, high-speed serial connections transmitted in a parallel bundle. DVI specifications that are extended to the dual mode operation allow for greater data rates for higher display resolutions. This requires seven parallel, differential, high-speed pairs. Quality cabling and connections become extremely important. The nominal DVI cable length limit is 4.5m (15ft). Electrical performance requirements are signal rise time of 0.330ns (nanoseconds), and a cable impedance of 100Ω. FEXT (far end crosstalk) should be less than 5%, and signal rise time degradation is a maximum of 160ps (picoseconds). Cable for DVI is application specific since the actual bit rate per channel is 1.65Gbps.

Table 18-13 Color Codes for Snake Cables

Pair No.	Color Combination	Pair No.	Color Combination	Pair No.	Color Combination	Pair No.	Color Combination	Pair No.	Color Combination
1	Brown	13	Lt. Gray/Brown stripe	25	Lt. Blue/Brown stripe	37	Lime/Brown stripe	49	Aqua/Brown stripe
2	Red	14	Lt. Gray/Red stripe	26	Lt. Blue/Red stripe	38	Lime/Red stripe	50	Aqua/Red stripe
3	Orange	15	Lt. Gray/Orange stripe	27	Lt. Blue/Orange stripe	39	Lime/Orange stripe	51	Aqua/Orange stripe
4	Yellow	16	Lt. Gray/Yellow stripe	28	Lt. Blue/Yellow stripe	40	Lime/Yellow stripe	52	Aqua/Yellow stripe
5	Green	17	Lt. Gray/Green stripe	29	Lt. Blue/Green stripe	41	Lime/Green stripe	53	Aqua/Green stripe
6	Blue	18	Lt. Gray/Blue stripe	30	Lt. Blue/Blue stripe	42	Lime/Blue stripe	54	Aqua/Blue stripe
7	Violet	19	Lt. Gray/Violet stripe	31	Lt. Blue/Violet stripe	43	Lime/Violet stripe	55	Aqua/Violet stripe
8	Gray	20	Lt. Gray/Gray stripe	32	Lt. Blue/Gray stripe	44	Lime/Gray stripe	56	Aqua/Gray stripe
9	White	21	Lt. Gray/White stripe	33	Lt. Blue/White stripe	45	Lime/White stripe	57	Aqua/White stripe
10	Black	22	Lt. Gray/Black stripe	34	Lt. Blue/Black stripe	46	Lime/Black stripe	58	Aqua/Black stripe
11	Tan	23	Lt. Gray/Tan stripe	35	Lt. Blue/Tan stripe	47	Lime/Tan stripe	59	Aqua/Tan stripe
12	Pink	24	Lt. Gray/Pink stripe	36	Lt. Blue/Pink stripe	48	Lime/Pink stripe	60	Aqua/Pink stripe

Table 18-14. Comparing Twisted-Pair High-Frequency Formats

Standard	Format	Intended Use	Connector Style	Cable Type	Transmission Distance[1]	Sample Rate	Data Rate (Mbps)	Guiding Document
DV	serial	professional/ consumer		(see IEEE 1394)	4.5m/15ft	20.25MHz	25	IEC 61834
IEEE 1394 (FireWire)	serial	professional/ consumer	1394	6 conductors, 2-STPs/2 pwr	4.5m/15ft	n/a	100, 200, 400	IEEE 1394
USB 1.1	serial	consumer	USB A & B	4 conductors, 1 UTP/ 2 pwr	5m/16.5ft	n/a	12	USB 1.1 Promoter Group
USB 2.0	serial	professional/ consumer	USB A & B	4 conductors, 1-UTP/ 2 pwr	5m/16.5ft	n/a	480	USB 2.0 Promoter Group
DVI	serial/ parallel	consumer	DVI (multipin D)	Four STPs	10m/33ft	To 165MHz	1650	DDWG; DVI 1.0
HDMI	parallel	consumer	HDMI (19 pin)	Four STPs + 7 conductors	Unspecified	To 340MHz	To 10.2Gbps	HDMI LLC
DisplayPort	parallel	consumer	20 pin	Four STPs + 8 conductors	15m	To 340MHz	To 10.8Gbps	VESA

[1] Transmission distances may vary widely depending on cabling and the specific equipment involved. STP = shielded twisted pair, UTP = unshielded twisted pair, n/a = not applicable

Picture information or even the entire picture can be lost if any vital data is missing with digital video interfaces. DVI cable and its termination are very important and the physical parameters of the twisted pairs must be highly controlled as the specifications for the cable and the receiver are given in fractions of bit transmission.

Requirements depend on the clock rate or signal resolution being used. Transferring the maximum rate of 1600 × 1200 at 60Hz for a single link system means that the time to write one bit across a screen is 0.606ns.

The DVI receiver specification allows 0.40 × bit time, or 0.242ns intrapair skew within any twisted pair. The pattern at the receiver must be very symmetrical. The interpair skew, which governs how bits will line up in time at the receiving decoder, may only be 0.6 × pixel time, or 0.364ns. These parameters control the transmission distances for DVI.

Also, the cable should be evaluated on its insertion loss for a given length. DVI transmitter output is specified into a cable impedance of 100Ω with a signal swing of ±780mV with a minimum signal swing of ±200mV. When determining DVI cable, assume minimum performance by the transmitter—i.e., 200mV—and best sensitivity by the receiver which must operate on signals ±75mV. Under these conditions the cable attenuation can be no greater than 8.5dB at 1.65GHz (10bits/pixel × 165MHz clock) which is relatively difficult to maintain on twisted-pair cable.

DVI connections combine the digital delivery, described above, with legacy analog component delivery. This allows DVI to be the transition delivery scheme between analog and digital applications.

18.9.3.2 HDMI

HDMI (high definition multimedia interface) is similar to DVI except that it is digital-only delivery. Where DVI has found its way into the commercial space as well as consumer applications, HDMI is almost entirely consumer-based. It is configured into a 19 pin connector which contains four shielded twisted pairs (three pairs data, one pair clock) and seven wire for HDCP (copy protection), devices handshaking, Ethernet and power. The standard versions of HDMI are nonlocking connector, attesting to its consumer-only focus. There are currently in access of three billion devices with an HDM port in the world today.

18.9.3.3 IEEE –1394 or FireWire Serial Digital

FireWire, or IEEE –1394, is used to upload DV, or digital video, format signals to computers etc. DV, sometimes called DV25, is a serial digital format of 25Mbps. IEEE 1394 supports up to 400Mbps. The specification defines three signaling rates, S100 (98.304Mbps), S200 (196.608Mbps), and S400 (393.216Mbps).

IEEE 1394 can interconnect up to sixty three devices in a peer-to-peer configuration so audio and video can be transferred from device to device without a computer, D/A, or A/D conversion. IEEE 1394 is hot plugable from the circuit while the equipment is turned on.

The IEEE 1394 system uses two shielded twisted pairs and two single wires, all enclosed in a shield and jacket, Fig. 18-2. Each pair is shielded with 100% coverage foil and a minimum 60% coverage braid. The outer shield is 100% coverage foil and a minimum 90% coverage braid. Each pair is shielded with aluminum foil and is equal to or greater than 60% braid. The twisted pairs handle the differential data and strobe (assists in clock regeneration) while the two separate wires provide the power and ground for remote devices. Signal level is 265mV differential into 110Ω.

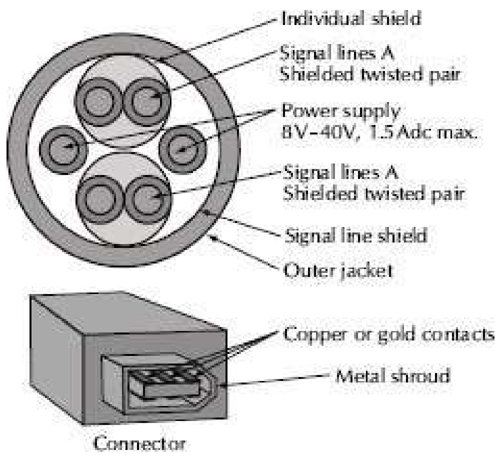


Figure 18-2. IEEE 1394 cable and connector.

The IEEE 1394 specification cable length is a maximum of 4.5m (15ft). Some applications may run longer lengths when the data rate is lowered to the 100Mbps level. The typical cable has 28gage copper twisted pairs and 22gage wires for power and ground. The IEEE 1394 specification provides the following electrical performance requirements:

- Pair-to-pair data skew is 0.40ns.
- Crosstalk must be maintained below –26dB from 1MHz to 500MHz.
- Velocity of propagation must not exceed 5.05ns/m.

Table 18-15 gives details of the physical interface system for IEEE 1394.

Table 18-15. Critical IEEE 1394 Timing Parameters

Parameter	100 Mbps	200 Mbps	400 Mbps
Max *Tr/Tf*	3.20ns	2.20ns	1.20 ns
Bit Cell Time	10.17ns	5.09ns	2.54ns
Transmit Skew	0.40 ns	0.25ns	0.20ns
Transmit Jitter	0.80ns	0.50ns	0.25ns

Receive End Skew	0.80ns	0.65ns	0.60ns
Receive End Jitter	1.08 ns	0.75ns	0.48ns

18.9.3.4 USB

The USB, universal serial bus, simplifies connection of computer peripherals. USB 1.1 is limited to a communications rate of 12Mbps, while USB2.0 supports up to 480Mbps communication. The USB cable consists of one twisted pair for data and two untwisted wires for powering downstream appliances. A full-speed cable includes a 28 gage twisted pair, and an untwisted pair of 28 to 20 gage power conductors, all enclosed in an aluminized polyester shield with a drain wire.

Nominal impedance for the data pair is 90Ω. The maximum cable length is determined by the signal propagation delay which must be less than 26ns from end to end. Table 18-16 lists some common plastics and the theoretical distance each could go based on 26ns. With an additional allowance of 4ns, which is split between the sending device connection and the receiver connection/response function, the entire one-way delay is a maximum of 30ns. The cable velocity of propagation must be less than 5.2ns/m and the length and twist of the data pair must be matched so time skew is no more than 0.10ns between bit polarities. The nominal differential signal level is 800mV.

Table 18-16. Dielectric Constant, Delay, and Transmission Distance of Various Plastics

Material	Dielectric Constant	Delay ns/ft	Maximum USB Distance in ft
Foam, Air-Filled Plastic	1.35	1.16	22.4

Solid Teflon™	2.1	1.45	18
Solid Polyethylene	2.3	1.52	17
Solid Polyvinyl Chloride	3.5-6.5	1.87-2.55	10-14

18.9.3.5 DisplayPort

DisplayPort is an emerging protocol for digital video. Its original intention was the transfer of images from a PC or similar device to a display. It has some significant advantages over DVI and HDMI. DisplayPort is by design backward-compatible to single link DVI and HDMI. Those are both severely distance-limited by the delay skew of the three data pairs when compared to the clock pair. With DisplayPort the clock is embedded with the video, much as the clock is embedded with the audio bit stream in AES digital audio, so the distance limitations on DisplayPort are less likely to involve clock timing problems.

However, display port is also a nonlocking connector, of 20 pins, and is intended for maximum distance 15m (50ft). These cables are, like HDMI and DVI, only available in assemblies. Raw cable and connectorization in the field do not currently look like an option for the professional installer. All these factors make it less likely to be embraced by the professional broadcast video arena.

There are field installed DIY connectors available for HDMI but they are difficult to install and their performance is questionable and the complexity and time to install is not the same as common connectors such as a BNC or XLR.

18.9.3.6 Premise/Data Category Cables

While premise/data category cables were never intended to be audio or video cables, their high performance and low cost, and their ubiquitous availability, have seen them pressed into service carrying all sorts of nondata signals.

It should also be noted that high-speed Ethernet networks are routinely used to transport these audio and video signals in data networks. The emergence of 10GBaseT, 10 gigabit networks, will allow the transport of even multiple uncompressed 1080p/60 video images. For AES digital audio bit streams, 10GBaseT can carry more than 3000 channels at 48kHz sampling rate. The digital nature of most entertainment content, with the ubiquitous video server technology in use today, makes high-bandwidth, high-data-rate networks in audio, video, broadcast, and other entertainment facilities, an obvious conclusion.

However, with the emergence of 4K and 8K video, even 10BaseT bandwidth (10Gbps) may be inadequate. Coming 40GBaseT and 100baseT Ethernet networks may help resolve this problem.

18.9.3.6.1 Cabling Definitions

- *Telcom Closet* (TC). Location where the horizontal cabling and backbone cabling are made.
- *Main Cross-Connect* (MXC). Often called the equipment room and is where the main electronics are located.
- *Intermediate Cross-Connect* (IXC). A room between the TC and the MXC are terminated. Rarely used in LANs.
- *Horizontal Cabling*. The connection from the telcom closet to the work area.
- *Backbone Cabling*. The cabling that connect all of the hubs together.

- *Hub*. The connecting electronic box that all of the horizontal cables connect to which are then connected to the backbone cable.
- *Ethernet*. A 10, 100, or 1000Mb/s LAN. The 10Mbps version is called 10BaseT. The 100Mbps version is called Fast Ethernet and 1000Mbps version is called Gigabit Ethernet.

18.9.3.6.2 Structured Cabling

Structured cabling, also called communications cabling, data/voice, low voltage, or limited energy is the standardized infrastructure for telephone and local area network (LAN) connections in most commercial installations. The architecture for the cable is standardized by Electronic Industries Association and Telecommunications Industry Association (EIA/TIA), an industry trade association. EIA/TIA 568, referred to as 568, is the main document covering structured cabling. IEEE 802.3 also has standards for structured cabling.

The current standard, as of this writing, is EIA/TIA 568-B.2-10 that covers all active standards up to 10GbaseT, 10 gigabit cabling.

18.9.3.6.3 Types of Structured Cables

Following are the types of cabling, Category 1 though Category 7, often referred to as Cat 1 through Cat 7. The standard TIA/EIA 568A no longer recognizes Cat 1, 2, or 4. As of July 2000, the FCC mandated the use of cable no less than Cat 3 for home wiring. The naming convention specified by ISO/IEC 11801 is shown in Fig. 18-3.

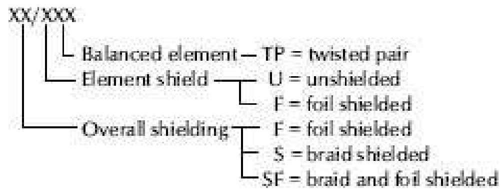


Figure 18-3. ISO/IEC 11801 cable naming convention.

Table 18-17 gives the equivalent TIA and ISO classifications for structured cabling.

Table 18-17. TIA and ISO Equivalent Classifications

Frequency bandwidth	TIA		ISO	
	Components	Cabling	Components	Cabling
1–100MHz	Cat 5e	Cat 5e	Cat 5e	Class D
1–250MHz	Cat 6	Cat 6	Cat 6	Class E
1–500MHz	Cat 6a	Cat 6a	Cat 6a	Class E_A
1–600MHz	n/s	n/s	Cat 7	Class F
1–1000MHz	n/s	n/s	Cat 7A	Class F_A

Category 1. Meets the minimum requirements for analog voice or plain old telephone service (POTS). This category is not part of the EIA/TIA 568 standard.

Category 2. Defined as the IBM Type 3 cabling system. IBM Type 3 components were designed as a higher grade 100Ω UTP system capable of operating 1Mb/s Token Ring, 5250, and 3270 applications over shortened distances. This category is not part of the EIA/TIA 568 standard.

Category 3. Characterized to 16MHz and supports applications up to 10Mbps. Cat 3 conductors are 24AWG. Applications range from voice to 10BaseT.

Category 4. Characterized to 20MHz and supports applications up to 16Mb/s. Cat 4 conductors are 24AW0. Applications range from voice to 16Mbps Token Ring. This category is no longer part of the EIA/TIA 568 standard.

Category 5. Characterized to 100MHz and supports applications up to 100Mbps. Cat 5 conductors are 24AWG. Applications range from voice to 100BaseT. This category is no longer part of the EIA/TIA 568 standard.

Category 5e. Characterized to 100MHz and supports applications up to 1000Mbps/1Gbps. Cat 5e conductors are 24AWG. Applications range from voice to 1000BaseT. Cat 5e is specified under the TIA standard ANSI/TIA/EIA-568-B.2. Class D is specified under ISO standard ISO/IEC 11801, 2nd Ed.

Category 6. Characterized to 250MHz, in some versions bandwidth is extended to 600MHz, and supports 1000Mbps/1Gbps and future applications and is backward compatible with Cat 5 cabling systems. Cat 6 conductors are 23AWG. This gives improvements in power handling, insertion loss, and high-frequency attenuation. Fig. 18-4 shows the improvements of Cat 6 over Cat 5e. Cat 6 is specified under the TIA standard ANSI/TIA/EIA-568-B.2-1. Class E is specified under ISO standard ISO/IEC 11801, 2nd Ed. Cat 6 is available most commonly in the United States as UTP.

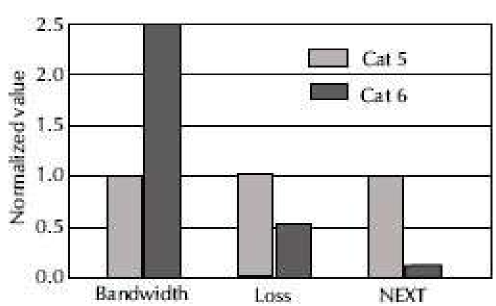


Figure 18-4. Normalized comparison of Cat 5e and Cat 6.

Category 6 F/UTP. Cat 6 F/UTP (foiled unshielded twisted pair) or ScTP (screened twisted pair) consists of four twisted pairs enclosed in a foil shield with a conductive material on one side. A drain wire runs adjacent to the conductive side of the shield, Fig. 18-5. When appropriately connected, the shield reduces ANEXT, RFI, and EMI. Cat 6 FTP can only be designed to 250MHz per TIA/EIA 568B.2-1.

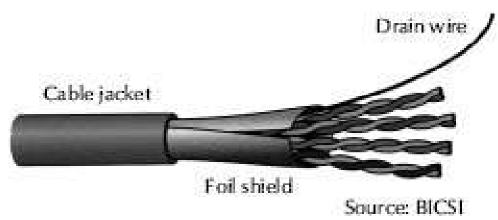


Figure 18-5. Cat 6 F/UTP.

Category 6a. Cat 6a (Augmented Category 6) is characterized to 500MHz, and in special versions to 625MHz, has lower insertion loss, and has more immunity to noise. Cat 6a is often larger than the other cables. 10GBaseT transmission uses digital signal processing (DSP) to cancel out some of the internal noise created by NEXT and FEXT between pairs. Cat 6a is specified under the TIA standard ANSI/TIA/EIA 568-B.2-10. Class E_A is specified under ISO standard ISO/IEC 11801, 2nd Ed. Amendment 1. Cat 6a is available as UTP or FTP.

Category 7. Cat 7 is another four pair data cable. In this case the pairs are individually shielded with a drain wire, much like audio snake cable. However, this cable intended for a bandwidth up to 1GHz per pair. Common in Europe, it is almost unknown in the USA. In fact, almost all the Cat 7 in the USA is to wire up specialized equipment made in Europe and exported to the States.

Cat 7 is most often used for horizontal and building backbone cable and to support current and future Cat 6a and 7 applications, such as 10GBase-T (10Gigabit Ethernet), 1000Base-T (Gigabit Ethernet), 100 Base-T, 10 Base-T, FDDI, ATM

Construction and dimensions are as follows, see Fig. 18-6. The typical performance characteristics are given in Table 18-18.

1.	Conductor	
	Material	Solid bare copper ETP
	Diameter	AWG 23
2.	Insulation	
	Material	Foamed polyethylene
	Nominal diameter 1.45mm over insulation	
3.	Cable core	
	Pair	2 twisted insulated conductors with overall foil
	Foil	Laminated aluminum-polyester Aluminum facing outside
	Number of shielded pairs 4, all twisted together	
	Color code pair 1 White/Blue	
	Color code pair 2 White/Orange	
	Color code pair 3 White/Green	
	Color code pair 4 White/Brown	
4.	Braid	
	Material	Solid tinned copper
	Coverage	≥ 30%
5.	Jacket	
	Material	LSNH

Diameter	8.0 ± 0.3mm
Ripcord	Nylon ripcord under jacket

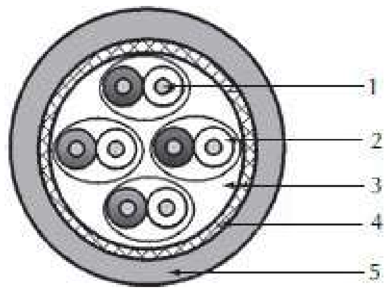


Figure 18-6. Cat 7 construction. Courtesy Belden.

Besides the advantages of bandwidth, there are a number of disadvantages. First is the very fact that the pairs are shielded. This means that the possibilities of "ground loops" (mentioned on Section 18.16.1) is now a serious problem in the data world. In the audio world, this is solved by cutting one of the drain wires at one end. The correct end would be the destination wire, leaving the source drain wire connected. However, that instantly reduces the shield effectiveness by 50% (one wire instead of two). Further, if you cut the wrong end, you get a very nice comb filter. If you cut the correct end, you get exactly the same effect only lower in amplitude. Fig. 18-7 shows the effect of cutting the CORRECT drain wire.

Most of the Cat 7 used is punched down onto terminal blocks. There are no RJ-45's male plugs available for this cable. Most often, it is used as trunk cable and overall shielded cable is used as the drop cable to the user. This, of course, limits the bandwidth.

Category 8. The next step in Ethernet is going to be 40 gigabit networking. The standards group is ANSI/TIA 568C.2-1. This standard already exists in single mode fiber called 40GbaseSR4. The number at the end indicates that there will be four fibers in each direction, a total of eight fibers. Essentially, this is to keep the cost down by using multiple lanes of 10 gigabit fiber, with existing 10gig source and destination devices. At some point in the future, transmitters and receivers will be able to accomplish 40 gig on one singlemode fiber. Most installs even today use 12-fiber bundles. In fact, this is the most popular fiber count. Clearly, this would cover the eight lanes of 40GbaseSR4, with four spare fibers as back-up. There will also be multimode fiber available, called 40GbaseLR4. Still 10gig per fiber times 8 fibers, but more distance limited than singlemode.

Table 18-18. Cat 7 Typical Performance Characteristics

TYPE	1*	4	10	16	31.2	62.5	100	125	200	250	300	600	1000	MHz
Attenuation	1.8	3.4	5.5	6.9	9.7	13.9	17.7	19.9	25.6	28.8	31.8	46.6	62.2	dB/100m
NEXT	103	100	98	97	95	94	93	92	91	90	90	89	88	dB/100m
PS NEXT	100	97	95	94	92	91	90	89	88	87	87	86	85	dB/100m
ACR	101	97	92	90	85	80	75	72	65	61	58	42	26	dB/100m
PS ACR	98	94	89	87	82	77	72	69	64	58	55	39	23	dB/100m
ACR-F	95	94	93	91	90	87	85	83	77	74	74	60	50	dB/100m
PS ACR-F	92	91	92	88	87	84	82	80	74	71	71	57	47	dB/100m
Return Loss	27	30	32	32	35	33	32	31	30	25	25	23	21	dB/100m

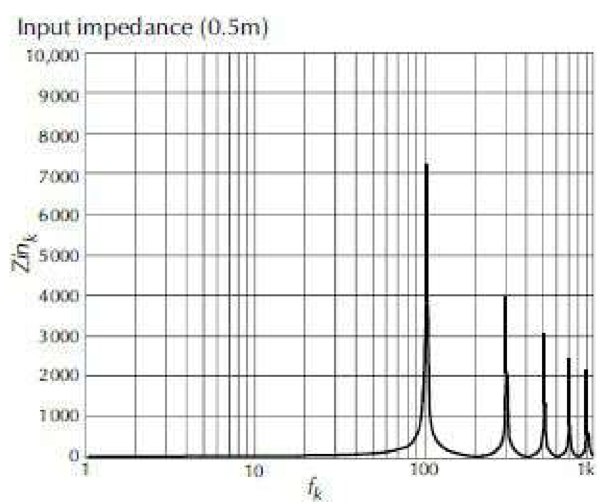


Figure 18-7. The effect of cutting the correct drain wire. Courtesy

Belden.

For twisted pair, the standard will be 40GbaseT (Cat 8) and the study group is not expected to be finished and ratified until 2015. This is expected to be the same individual foil shields as Cat 7, so installers in the USA will begin to get used to this format. The bandwidth is expected to be 2GHz per pair. Because of the immense bandwidth, it is unlikely that any copper cable will go a full 100m (328 ft). It is expected that Cat 8 will be limited to 30m (100ft). While it has been stated that this will cover 90% of the data center wiring, it is a serious question for users such as broadcasters or motion picture production who will want to go much farther. Fiber might be the only option.

18.9.3.6.4 Multimode and Singlemode Fiber

As with most technologies, the prices of fiber are constantly in flux. Multimode used to be the cheap alternative, while single mode was the expensive choice. This has all changed. In fact, for many manufacturers, multimode is now more expensive to produce than single mode, meaning that yet another barrier to implementing single mode has fallen. If you want bandwidth, if you want distance, go singlemode. And, while singlemode transceivers are still more expensive, even there the prices are equalizing. A lot of it is economy of scale. All the long-run telecommunications is on singlemode. While it is a little more finicky (alignment of connectors must be perfect) things are getting more “idiot proof” every day.

This is not to say that you can buy one of these simple do-it-yourself fiber connectors and you are a certified installer. You can have a terrible copper network installation if you don't know what you are doing. (You could have a terribly soldered XLR, too!) The whole point is take a class, learn how fiber works, how to put it in, how to take care of it (how to clean the connectors for instance, a major source of fiber failure).

18.9.3.6.5 Comparisons

Table 18-19 compares network data rates for Cat 3 through Cat 6a and Table 18-20 compares various characteristics of Cat 5e, 6, and 6a. Fig. 18-8 compares the media distance-bandwidth product of Cat 5e and Cat 6a with 802.11 (a, b, g, n) wireless media, often called Wi-Fi.

Table 18-19. Network Data Rates, Supporting Cable Types, and Distance

Minimum Performance	Token Ring	Ethernet	Maximum Distance
Cat 3	4Mb/s	10Mbps	100m/328ft
Cat 4	16 Mb/s	-	100m/328ft
Cat 5	-	100Mbps	100m/328ft
Cat 5e		1000 Mbps	100m/328ft
Cat 6	-	10Gbps	55m/180ft
Cat 6a		10 Gbps	100m/328ft

Table 18-20. Characteristics of Cat 5e, Cat 6, and Cat 6a

Cabling Type	Cat 5e	Cat 6	Cat 6a
Relative Price (%)	100	135-150	165-180
Available Bandwidth	100MHz	250MHz	500MHz
Data rate Capability	1.2 Gbps	2.4Gbps	10 Gbps
Noise Reduction	1.0	0.5	0.3

Broadband Video Channels 6 MHz/channels per pair	17	42	83
Broadband Video Channels rebroadcast existing channels	6	28	60+
No. of Cables in Pathway 24in × 4in	1400	1000	700

New cable designs can affect size and pathway load so consult the manufacturer. Note that cable density is continually changing with newer, smaller cable designs. Numbers in Table 18-20 should be considered worst case. Designers and installers of larger systems should get specific dimensional information from the manufacturer.

Fig. 18-9 shows various problems that can be found in UTP cabling. Fig. 18-10 gives the maximum distances for UTP cabling as specified by ANSI/TIA.

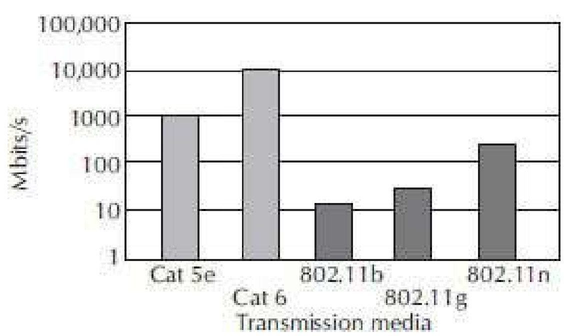


Figure 18-8. Comparison of media distance to bandwidth

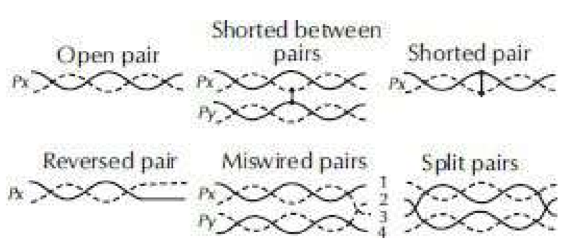


Figure 18-9. Paired wiring faults. Courtesy Belden.

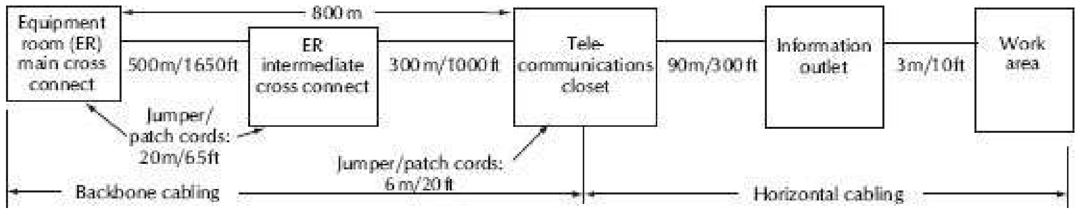


Figure 18-10. Maximum distances between areas for UTP cable.

Four (4) pair 100Ω ±15% UTP Cat 5e cabling is the recommended minimum requirement for residential and light commercial installations because it provides excellent flexibility. Pair counts are four pair for desktop and twenty five pair for backbone cabling. The maximum length of cable is 295ft (90m) with another 33ft (10m) for patch cords.

Unshielded twisted pairs (UTP) and shielded twisted pairs (STP) are used for structured cabling. Unshielded twisted pairs (UTP) are the most common today. These cables look like the POTS cable, however, their construction makes them usable in noisy areas and at high frequencies because of the short, even twisting of the two wires in each pair. All four pairs are twisted differently, have different “lay lengths” to meet the crosstalk requirements of that particular category. When the two wires separate because of bending and flexing during installation or to aid in the use of some RJ-45 connectors, this can increase crosstalk, increase signal egress and ingress, affect impedance and return loss, and increase noise. The twist must be even and tight so complete noise cancellation occurs along the entire length of the cable. To best keep the twist tight and even, better cable has the two wires bonded together so they will not separate when bent or flexed. Patch cable is flexible so twist and impedance are not as well controlled. The color codes for the pairs are given in

Table 18-21. Color Code for UTP Cable

Pair No.	1st Conductor Base/Band	2nd Conductor
1	White/Blue	Blue
2	White/Orange	Orange
3	White/Green	Green
4	White/Brown	Brown

Cable diameter varies for the different types of cable. TIA recommends that two Cat 6 cables but only one Cat 6a cable can be put in a ¾in (21mm) conduit at 40% fill. The diameter and the stiffness of the cables determine their bend radius and therefore the bend radius of conduits and trays, Table 18-22.

Table 18-22 Diameter and Bend Radius for 10GbE Cabling

Cable	Diameter	Bend Radius
Category 6	0.22in (5.72mm)	1.00in (4 × OD)
Category 6a	0.35in (9mm)	1.42in (4 × OD)
Category 6 FTP	0.28in (7.24mm)	2.28in (8 × OD)
Category 7 STP	0.33in (8.38mm)	2.64in (8 × OD)

Fig. 18-11 shows the construction of UTP and screened UTP cable.

18.9.3.6.6 Critical Parameters

Critical parameters for UTP cable are: NEXT, PS-NEXT, FEXT, ELFEXT, PS-ELFEXT, RL, ANEXT.

NEXT. *NEXT*, or near-end crosstalk, is the unwanted signal coupling from the near end of one sending pair to a receiving pair.

PS-NEXT. *PS-NEXT*, or power-sum near-end crosstalk, is the crosstalk between all of the sending pairs to a receiving pair. With four-pair cable, this is more important than NEXT.

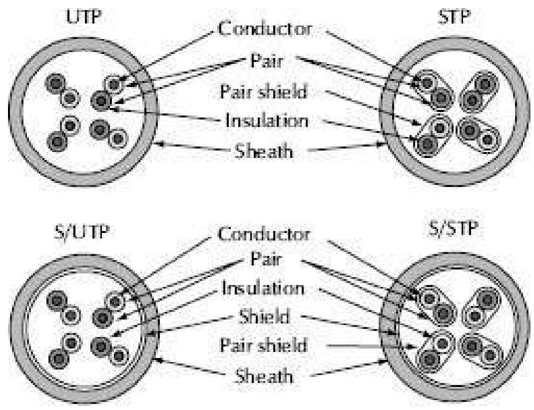


Figure 18-11. UTP and S/UTP cable design.

FEXT. *FEXT*, or far-end crosstalk, is the measure of the unwanted signal from the transmitter at the near end coupling into a pair at the far end.

EL-FEXT. *EL-FEXT*, or equal level far-end crosstalk, is the measure of the unwanted signal from the transmitter end to a neighboring pair at the far end relative to the received signal at the far end. The equation is

$$EL-FEXT = FEXT - Attenuation \qquad (18\text{-}3)$$

Power sum equal-level far-end crosstalk is the computation of the unwanted signal coupling from multiple transmitters at the near end into a pair measured at the far end relative to the received signal level measured on the same pair.

Return Loss (*RL*). *RL* is a measure of the reflected energy from a transmitted signal and is expressed in -dB, the higher the value, the better. The reflections are caused by impedance mismatch caused by connectors, improper installation such as stretching the cable or too sharp a bend radius, improper manufacturing, or improper load.

The desired impedance for various cable types is:

- All UTP data cables (Cat 5, 5e, 6, 6a) is 100Ω.
- All passive video, HD, HD-SDI or 1080p/60 components is 75Ω.
- All digital audio twisted pairs is 110Ω.
- All digital audio on coaxial cable is 75Ω.

Broadcasters are very familiar with return loss, calling it by a different name, SWR (standing wave ratio) or VSWR (voltage standing wave ratio). In fact, return loss measurements can easily be converted into VSWR values, or vice versa. Return loss can be found with the equation

$$RL = 20\log\frac{Difference}{Sum} \qquad (18\text{-}4)$$

where,

Difference is the difference (absolute value) between the desired impedance and the actual measured impedance,

Sum is the desired impedance and the actual measured impedance added together.

With 1000BaseT systems, the pairs simultaneously transmit and receive. As the transmitter sends data, it is also listening for data being sent from the opposite end of the same pair. Any reflected signal from the sending end that reflects back to the sending end mixes with the sending signal from the far end, reducing intelligibility. With 10BaseT or 100BaseT data networks, one pair transmits while another receives, so reflections *(RL*, return loss) are not a major consideration and were not required to be measured. Now, with pairs simultaneously transmitting and receiving, called duplex mode, RL is a critical measurement for data applications.

Delay Skew. Since every pair (and every cable) takes a specific amount of time to deliver a signal from one end to the other, there is a delay. Where four pairs must deliver data to be recombined, the delay in each pair should, ideally, be the same. However, to reduce crosstalk, the individual pairs in any Category cable have different twist rates. This reduces the pair-to-pair crosstalk but affects the delivery time of the separate parts. This is called *delay skew*.

While delay skew affects the recombining of data, in 1000BaseT systems, for instance, the same delay skew creates a problem when these UTP data cables are used to transmit component video or similar signals, since the three colors do not arrive at the receiving end at the same time, creating a thin bright line on the edge of dark images. Some active baluns have skew correction built in, see Section 18.12.5.

ANEXT. *ANEXT*, or alien crosstalk, is coupling of signals between cables. This type of crosstalk cannot be cancelled by DSP at the switch level. Alien crosstalk can be reduced by overall shielding of the pairs, or by inserting a nonconducting element inside that cable to push away the cables around it.

18.9.3.6.7 Terminating Connectors

All structured cabling use the same connector, an RJ-45. In LANs (local area networks) there are two possible pin-outs, 568A and 568B. The difference is pair 2 and pair 3 are reversed. Both work equally well as long as they are not intermixed. The termination is shown in Fig. 18-12.

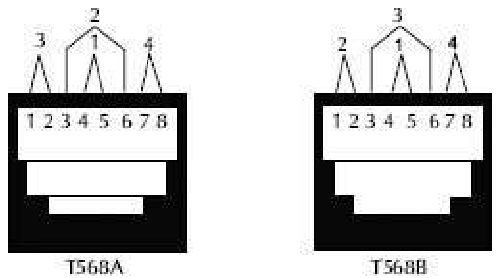


Figure 18-12. Termination layout for EIA/TIA 568-B.2 cable.

In the past decade, the B wiring scheme has become the most common. However, if you are adding to or extending and existing network, you must determine which wiring scheme was used and continue with that scheme. A mixed network is among the most common causes of network failure.

It is very important that the pairs be kept twisted as close to the connector as possible. For 100BaseT (100MHz, 100Mbps) applications, a maximum of ½in should be untwisted to reduce crosstalk and noise pickup. In fact, with Cat 6 (250MHz) or Cat 6a (500MHz) it is safe to say that any untwisting of the pairs will affect performance. Therefore there are many connectors, patch panels, and punch-down blocks that minimize the untwisting of the pairs.

18.9.3.6.8 Baluns

Baluns (*Bal*anced-*Un*balanced) networks are a method of connecting devices of different impedance and different formats. Baluns have been commonly used to convert unbalanced coax, to balanced twinlead for television antennas, or to match coaxial data formats (coaxial Ethernet) to balanced line systems (10BaseT, 100BaseT etc.). Other balun designs can allow unbalanced sources, such as video or consumer audio, for instance, to be carried on balanced lines, such as UTP Cat 5e, 6, etc.

Since there are four pairs in a common data cable, this can carry four channels. Since category cables are rarely tested below 1MHz, the audio performance was originally suspect. Crosstalk at audio frequencies in UTP has been measured and is consistently better than –90dB even on marginal Cat 5. On Cat 6, the crosstalk at audio frequencies is below the noise floor of most network analyzers.

Baluns are commonly available to handle such signals as analog and digital audio, composite video, S-video, RGB or other component video (VGA, Y/R-Y/B-Y, Y/Cr/Cb), broadband RF/CATV, and even DVI and HDMI. The limitations to such applications are the bandwidth specified on the cable and the performance of the cable (attenuation, return loss, crosstalk, etc.) at those higher frequencies.

Passive baluns can also change the source impedance in analog audio devices. This dramatically extends the effective distance of such signals from only a few feet to many hundreds of feet. Consult the balun manufacturer for the actual output impedance and effective distance of their designs.

Some baluns can include active amplification, equalizations, or skew (delivery timing) compensation. While more expensive, these active baluns can dramatically increase the effective distance of even marginal cable. The down side is the added expense and reduced reliability of one more link in the signal chain that could fail.

18.9.3.6.9 Adaptors

Users and installers should be aware there are adaptors, often that fit in wall plates, where keystone data jacks are intended to be snapped in place. These adaptors often connect consumer audio and video (RCA connectors) to 110 blocks or other twisted pair connection points. However, there is no unbalanced-to-balanced device (i.e. transformer) in these, so the noise rejection inherent in twisted pairs when run as a balanced line is compromised. These adaptors simply unbalance the twisted pair and offer dramatically short effective distances. Further, baluns can change the source impedance and extend distance. Adaptors with no transformers or similar components cannot extend distance and often reduce the effective distance. These devices should be avoided unless they contain an actual balun.

One exception to this is where like signals are matched, such as a balanced line to a balanced line. For instance, there are adaptors that convert XLR balanced lines to an RJ-45, where the Cat 5 or 6 data cable is also a balanced line. The same could be said for unbalanced adapters, such as RCA-to-RCA, or BNC-to-BNC, sometimes called bulkhead or feedthrough adaptors. Often these might change the gender of the connection but have no other performance advantage.

18.9.3.6.10 Power Over Ethernet (PoE)

PoE supplies power to various Ethernet services as VoIP (Voice over Internet Protocol) telephones, wireless LAN access points, Blu-tooth access points, and Web cameras. Many audio and video applications will soon use this elegant powering system. IEEE 802.3af-2003 is the IEEE standard for PoE. IEEE 802.3af specifies a maximum power level of 15.4W at the power sourcing equipment (PSE) and a maximum of 12.95W of power over two pairs to a powered device (PD) at the end of a 100m (330ft) cable.

The PSE can provide power by one of three configurations, Fig. 18-13:

1. Alternative A, sometimes called phantom powering, supplies the power over pairs 2 and 3.
2. Alternative B supplies power over pairs 1 and 4.
3. Alternative C supplies power over All pairs.
4. High power (60–90W) uses all four pairs; the two parallel power sources are then applied through two diode bridges as input sources to the powered device.

The voltage supplied is nominally 48Vdc with a minimum of 44Vdc, a maximum of 57Vdc, and the maximum current per pair is 350mAdc, or 175mAdc per conductor. For a single solid 24AWG wire, common to many category cable designs, of 100m length (328ft) this would be a resistance of 8.4Ω. Each conductor would dissipate 0.257W or 1.028W per cable (0.257W × 4 conductors). This causes a temperature rise in the cable and conduit which must be taken into consideration when installing PoE.

18.9.3.6.11 Power Over Ethernet Plus (PoE Plus)

PoE Plus is defined in IEEE 802.3at and is capable of delivering up to 30W. Work is being done to approach 60W or even greater. This requires the voltage supply to be 50 to 57Vdc. Assuming a requirement of 42W of power at the endpoint at 50Vdc, the total current would be 0.84A, or 0.21A per pair, or 0.105A (105mA) per conductor, or a voltage drop of only 0.88V in one 24AWG wire.

Some non-standard systems, such as HDbaseT, claim to be able to handle 100W over standard Ethernet cable. Because this does not meet the PoE standard, it is called PoH (or PoHDbaseT). The ability of small conductors to handle this much power and over long distances (100m) should be considered and calculated before installing. It should be noted that many Cat 6 or 6a cables are 23AWG. The intent is to power up the device and feed signals to and from the device all with one cable.

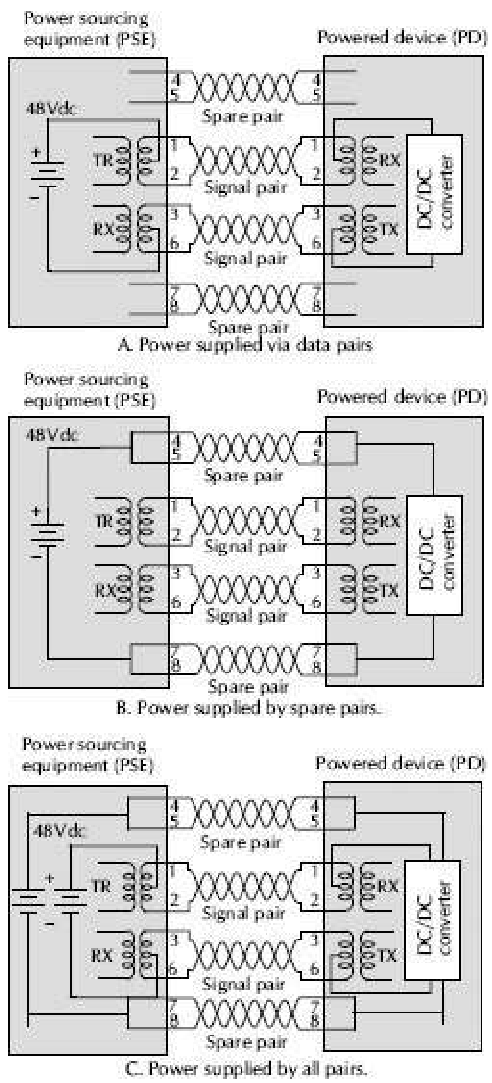


Figure 18-13. Three methods of supplying power via the Ethernet. Courtesy of Panduit Intel Corp.

18.10 Coaxial Cable

Coaxial cable is a design in which one conductor is accurately centered inside another with both conductors carrying the desired signal currents (source to load and return), as shown in Fig. 18-14. Coaxial cable is so called because if you draw a line through the center of a cross-sectional view, you will dissect all parts of the

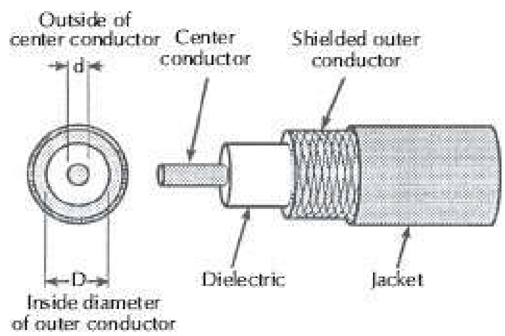


Figure 18-14. Construction of a coaxial cable.

18.10.1 History of Coaxial Cable

It has been argued that the first submarine telegraph cable (1858) was coaxial, Fig. 18-15. While this did have multiple layers, the outer layer was not part of the signal-carrying portion. It was a protective layer.

Figure 18-15. First submarine cable.

Modern coaxial cable was invented on May 23, 1929 by Lloyd Espenscheid and Herman Affel of Bell Laboratories. Often called *coax*, it is often used for the transmission of high-frequency signals. At high frequencies, above 100kHz, coax has a dramatically better performance than twisted pairs. However, coax lacks the ability to reject noise that twisted pairs can do when configured as balanced lines. Coaxial cable was first installed in 1931 to carry multiple telephone signals between cities.

18.10.2 Coaxial Cable Construction

The insulation between the center conductor and the shield of a coaxial cable affects the impedance and the durability of the cable. The best insulation to use between the center conductor and the shield would be a vacuum. The second best insulation would be dry air, the third, nitrogen. The latter two are familiar insulators in hard-line transmission line commonly used to feed high-power antenna in broadcasting.

A vacuum is not used, even though it has the lowest dielectric constant of "1," because there would be no conduction of heat from the center conductor to the outer conductor and such a transmission line would soon fail. Air and nitrogen are commonly used under pressure in such transmission lines. Air is occasionally used in smaller, flexible cables.

Polyethylene (PE) was common as the core material in coaxial cables during WW II. Shortly after the war, polyethylene was declassified and most early cable designs featured this plastic. Today most high-frequency coaxial cables have a chemically formed foam insulation or a nitrogen gas injected foam. The ideal foam is high-density hard cell foam, which approaches the density of solid plastic but has a high percentage of nitrogen gas. Current state-of-the-art polyethylene foam velocity is 86% (dielectric constant: 1.35) although most digital video cables are 82–84% velocity of propagation. High-density foam of this velocity resists conductor migration when the cable is bent, keeping impedance variations to a minimum. This high velocity improves the high-frequency response of the cable.

A problem with soft foam is it easily deforms, which changes the distance between the center conductor and the shield, changing the cable impedance. This can be caused by bending the cable too sharply, or running over it, or pulling it too hard, or any other possibility. To reduce this problem, a hard cell foam is used. Some cable that is rated as having a very high velocity of propagation might use very soft foam. A simple test can be performed where the user squeezes the foam dielectric of various cables. It will be immediately apparent that some cables have a density (crush resistance) double that of other designs. Soft foam can lead to conductor migration over time which will change timing, impedance, return loss, and bit errors over distance.

Coaxial cable is used quite extensively with various types of test equipment. When such cable is replaced, the capacitance per foot, which is determined by the dielectric constant of the insulator, must be taken into consideration, particularly for oscilloscope probes.

18.10.2.1 CCTV Cable

CCTV (closed circuit television) cable has a 75Ω characteristic impedance. CCTV is a baseband signal comprised of low-frequency vertical and horizontal sync pulse information and high-frequency video information. Since the signal is broadband, only cable with a center conductor of solid copper should be used.

If the cable is constantly in motion as in pan and tilt operation, a stranded center conductor should be used as a solid conductor will work-harden and break. There are also robotic coaxes designed to flex millions of times before failure for intense flexing applications.

Shielding for CCTV cable should have a copper or tinned-copper braid of at least 80% coverage, for low-frequency noise rejection. If an aluminum foil shield is used in conjunction with a braid, either tinned copper or aluminum only may be used for the shield. A bare copper braid can result in a galvanic reaction if exposed to air and common pollutants such as water, sulfur, chlorine, salt, etc. that may be contained in the air.

CCTV Distances

For common CCTV 75Ω cables, their rule-of-thumb transmission distances are shown in Table 18-23. These distances can be extended by the use of in-line booster amplifiers.

Table 18-23. Transmission Distances for CCTV Cable

RG-59	1000 ft
RG-6	1500 ft
RG-11	3000ft

18.10.2.2 CATV Broadband Cable

For higher-frequency applications, such as carrying radio frequencies or television channels, only the skin of the conductor is working (see Section 18.2.8, Skin Effect). Television frequencies in the United States, for instance, start with Channel 2 (54MHz), FM radio frequencies in the United States start at 88MHz, either of which are definitely in the area of skin effect. So these cables can use center conductors that have a layer of copper over a steel wire, since only the copper layer will be working.

If one uses a copper-clad steel conductor for applications below 50MHz, the conductor has a dc resistance from four to seven times that of a solid copper conductor. If a copper-clad cable is used on a baseband video signal, for instance, the sync pulses may be attenuated too much. If such a cable is used to carry audio, almost the entire audio signal will be running down the steel wire.

Cables that carry dc as well as RF high-frequency signals also need an all copper construction as the dc (at a frequency of "zero") would travel within the steel wire of a copper-clad construction.

CATV/broadband cable should have a foil shield for good high-frequency noise rejection. CATV cable should also have a braid shield to give the connector something to grab onto, 40% to 60% aluminum braid being the most common. Multiple layer shields are also available such as tri-shielded (foil-braid-foil) and quad shields (foil-braid-foil-braid). Assumptions that quad shields give the best shield effectiveness are erroneous, there being single foil/braid and tri-shield configurations that are measurably superior. Refer to Section 18.8.6 on shield effectiveness.

Modern CATV/broadband cable will use a foamed polyethylene or foamed FEP dielectric, and preferably one with gas injected foam. This will reduce the losses in the cable. The jacket material is determined by the environment that the cable will be working in (see Sections 18.4, 18.5, 18.6).

18.10.3 Coaxial Cable Installation Considerations

18.10.3.1 Indoor Installation

Indoor environments are the most common for coaxial cable installations. A few tips on installing coaxial cable are as follows:

1. First and foremost, follow all NEC requirements when installing coaxial cables.
2. Distribute the pulling tension evenly over the cable and do not exceed the minimum bend radius of ten times the diameter. Exceeding the maximum pulling tension or the minimum bend radius of a cable can cause permanent damage both mechanically and electrically to the cable.
3. When pulling cable through conduit, clean and deburr the conduit completely and use proper cable-friendly lubricants in long runs.

18.10.3.2 Outdoor Installation

Outdoor installations require special installation techniques that will enable the cable to withstand harsh environments. When using cable in an aerial application, lash the cable to a steel messenger, or buy cable with a built-in steel messenger. This will help support the cable and reduce the stress on the cable during wind, snow, and ice storms. When direct burying a cable, lay the cable without tension so it will not be stressed when earth is packed around it. When burying in rocky soil, fill the trench with sand. Lay the cable and then place pressure-treated wood or metal plates over the cable. This will prevent damage to the cable from rocky soil settling; in cold climate areas, bury the cable below the frost line. Buy direct burial cable designed to be buried.

18.10.4 Coaxial Cable Termination Techniques

18.10.4.1 Soldering

Soldering offers several advantages as it can be used on solid or stranded conductors and it creates both a solid mechanical and electrical connection. The disadvantage is that it takes more time to terminate than other methods and cold solder joints can cause problems if the connector is not soldered to the cable properly. The use of lead-based solder might also be a consideration if RoHS (reduction of hazardous substances) requirements are part of the installation. Soldering is not recommended for high-frequency applications, such as HD-SDI or 1080p/60 as the variations in dimensions will show up as variations in impedance and contribute to return loss (see Section 18.10.5).

18.10.4.2 Crimping

Crimping is probably the most popular method of terminating BNC and F connectors on coax cable. Like the solder method, it can be used on solid or stranded conductors and provides a good mechanical and electrical connection. This method is the most popular because there is no need for soldering so installation time is reduced. It is very important to use the proper size connector for a tight fit on the cable. Always use the proper tool. Never use pliers as they are not designed to place the pressure of the crimp evenly around the connector. Pliers will crush the cable and can degrade the electrical properties of the cable.

18.10.4.3 Twist-On Connectors

Twist-on connectors are the quickest way of terminating a coaxial cable; however, they do have some drawbacks. When terminating the cable with this type of connector, the center conductor is scored by the center pin on the connector, thus too much twisting can cause damage to the center conductor. It is not recommended for pan and tilt installations as the constant movement of the cable may work the connector loose. Because there is no mechanical or electrical crimp or solder connection, this connector is not as reliable as the other methods.

18.10.4.4 Compression Connectors

One of the most dramatic improvements in the last few years has been the rise of the truly professional one-piece compression connector. Once regarded as a low-quality alternative, with poor high frequency performance and poor mechanical pull-off resistance, this has all changed. There are now one-piece compression connectors with bandwidth out to 4.5GHz, for 1080p/60 3G-SDI video. These same connectors have a pull strength on RG-6 cable exceeding 110lbs. making them almost impossible to pull off. Since the pull tension on HD/3G RG-6 cable is approximately 69lbs, it is interesting to note that, should you succeed in pulling off the connector, you have long since ruined the cable, Fig. 18-16.

Figure 18-16. RG-6 cable connector. Courtesy Belden.

This is not to say that all compression connectors are of this quality. Manufacturer's specification sheets and test data should be requested to verify any performance claims. There are still low-quality compression connectors available in the marketplace.

18.10.5 Digital Signals

In all digital signals, there is a clock frequency that establishes the sampling of the original signal. It was determined in the 1920s by Harry Nyquist that the maximum sampling rate can only go up to half the clock frequency. This has been called the "Nyquist Limit". If you exceed the Nyquist Limit you begin to create artifacts, digital signals that are not real signals, not related to the original material. Therefore, in a digital signal, the actual usable data fills only half-way to the clock. This is sometimes called the "occupied bandwidth".

Passing Digital Signals

A perfect square wave consists of a clock frequency (sine wave) with an infinite number of harmonics added. If we are trying to measure a component, such as cable, and determine its suitability to carry digital signals, we need to take these harmonics into consideration. While it is impossible to measure “infinite harmonics” a finite number can be chosen. Based on existing test equipment, the common harmonic chosen is the third harmonic. What this means is that, given a specific clock frequency, we would divide this by two (Nyquist Limit) and then multiply that occupied bandwidth by three (third harmonic). We would then test everything (cable, connectors, patch panels, patch cords, bulkheads, adaptors etc.) to that frequency. We could also calculate the quarter-wavelength at that frequency. This will indicate what components are not critical, and do not need to be precisely 75Ω.

18.10.6 Return Loss

At high frequencies, where cable and connectors are a significant percentage of a wavelength, the impedance variation of cable and components can be a significant source of signal loss. When the signal sees something other than 75Ω, a portion of the signal is reflected back to the source. Table 18-24 shows the wavelength and quarter-wavelength at various frequencies. One can see that this was a minor problem with analog video (quarter wave is 59ft) since the distances are so long. However, with HD-SDI and higher signals, the quarter wave can be 1in or less, meaning that everything in the line is critical: cable, connectors, patch panels, patch cords, adaptors, bulkhead/feedthrough connectors, etc.

In fact, Table 18-24 above is not entirely accurate. The distances should be multiplied by the velocity of propagation of the cable or other component, to get the actual length, so they are even shorter still.

Since everything is critical at high frequencies, it is appropriate to ask the manufacturers of the cable, connectors, patch panels, and other passive components, how close to 75Ω their products are, and at the third harmonic of the occupied bandwidth. This can be established by asking for the return loss of each component. Table 18-25 will allow the user to roughly translate the answers given.

Table 18-24. Wavelength and Quarter Wavelength of Various Signals at Various Frequencies

Signal	Clock Frequency	Nyquist Limit	Third Harmonic	Wave-length	Quarter Wave-length
SD-SDI	270MHz	135MHz	405MHz	2.43ft 0.74m	7.3in 18.5cm
HD-SDI	1.5GHz	750MHz	2.25GHz	5.3in 13.46cm	1.3in 33mm
1080p/60	3GHz	1.5GHz	4.5GHz	2.6in 66mm	0.66in 16.8mm

Table 18-25. Return Loss versus % of Signal Received and Reflected

Return Loss	% of Signal Received	% Reflected
−50dB	99.999%	0.001%
−40dB	99.99%	0.01%
−30dB	99.9%	0.1%
−20dB	99.0%	1.0%
−10dB	90.0%	10.0%

Most components intended for HD can pass −20dB return loss. In fact, −20dB return loss at 2GHz is a good starting point for passive components intended for HD-SDI. Better components will pass −30dB at 2GHz. Better still (and rarer still) would be −30dB at 3GHz. There are currently no components that are consistently

−40dB return loss at any reasonable frequency. In Table 18-24, it can be seen that 1080p/60 signals need to be tested to 4.5GHz. This requires expensive custom-built matching networks. Belden was the first company to make such an investment.

Note that the number of nines in the Signal Received column is the same as the first digit of the return loss (i.e., −30dB = 3nines = 99.9%). There are similar tests, such as SRL (structural return loss). This test only partially shows total reflection. Do not accept values measured in any way except return loss. The SMPTE maximum amount of reflection on a passive line (with all components measured and added together) is −15dB or 96.84% received, 3.16% reflected. A line with an RL of −10dB (10% reflected) will probably fail.

18.10.7 Video Triaxial Cable

Video triaxial cable is used to interconnect analog video cameras to their related equipment. Triaxial cable contains a center conductor and two isolated shields, allowing it to support many functions on one cable. The center conductor and inner shield carry the video signals. The center conductor and outer shield carry intercoms, monitoring devices, tally lights, teleprompters, remote camera controls and many other functions, The two braids commonly carry camera power. The center shield carries the video signal ground or common. Triax cable is usually of the RG-59 or RG-11 type. There are also European version measured in millimeters.

There are camera manufacturers who offer digital cameras that run over triax. However, these convert the digital signal to analog such as broadband RGB or Y-C, which are analog signals, then convert them back at the other end. The users loses many of the advantages of an all-digital signal, not the least of which is sending a signal multiple times will show profound noise and poor video quality, just like the old days of analog video.

18.10.8 S-Video

S-video requires a duplex (dual) coaxial cable to allow separate transmission of the luminance (Y) and the chrominance (C). The luminance signal is black or white or any gray value while the chrominance signal contains color information. This transmission is sometimes referred to as Y-C. Separating signals provides greater picture detail and resolution and less noise interference.

S-video is sometimes referred to as S-VHS™ (Super-Video Home System). While its intention was for improved consumer video quality, these cameras were also used for the lower end of the professional area, where they were used for news, documentaries, and other less-critical applications.

Since this is an analog system, S-video or Y-C has all but disappeared from the consumer, commercial and professional marketplaces.

18.10.9 RGB

RGB stands for red-green-blue, the primary colors in color television. It is often called component video since the signal is split up to its component colors. When these analog signals are carried separately much better image resolution can be achieved. RGB can be carried on multiple single video cables, or in bundles of cables made for this application. With separate cables, all the cables used must be precisely the same electrical length. This may or may not be the same as the physical length. Using a vectorscope, it is possibly to determine the electrical length and compare the RGB components. If the cables are made with poor quality control, the electrical length of the coaxes may be significantly different (i.e., one cable may have to be physically longer than the others to align the component signals). Cables made with very good quality control can simply be cut at the same physical length.

Bundles of RGB cables should be specified by the amount of timing error, the difference in the delivery time on the component parts. For instance, all Belden bundled coax cables are guaranteed to be 5ns (nanosecond) difference per 100ft of cable. Other manufacturers should have a similar specification and/or guarantee. The de facto timing requirement for broadcast RGB is a maximum of 40ns. Timing cables by hand with a vectorscope allows the installer to achieve timing errors of >1ns. Bundled cables made for digital video can also be used for RGB analog, and similar signals (Y, R–Y, B–Y or Y, Pb, Pr or YUV or VGA, SVGA, XGA, etc.) although the timing requirements for VGA and that family of signals has not been established.

These bundled coaxes come in other version besides just three coax RGB. Often the horizontal and vertical synchronizing signals (H and V) are carried with the green video signal on the green coax. For even greater control, these signals can be carried by a single coax (often called RGBS) or five coaxes, one for each signal (called RGBHV). These cables are becoming more common in the home, where they are often referred to as five-wire video. There are also four-pair UTP data cables made especially to run RGB and VGA signals. Some of these have timing tolerance (called delay skew in the UTP world) that is seriously superior to bundled coaxes. However, the video signals would have to be converted from 75Ω to

100Ω, and the baluns to do this, one for each end of the cable, would be added to the cost of the installation. Further, the impedance tolerance of coax, even poorly made coax, is dramatically superior to twisted pairs. Even bonded twisted pairs are, at best, ±7Ω, where most coaxial cables are ±3Ω, with precision cables being twice as good as that, or even better.

18.10.10 VGA and Family

VGA stands for video graphics array. It is an analog format to connect progressive video source to displays, such as projectors and screens. VGA comes in a number of formats, based on resolution. These are shown in Table 18-26.

There are many more variations in resolution and bandwidth than the ones shown in Table 18-26. While many of these are high resolution, these are still analog signals. VGA and this family might be the last holdout of analog in our increasingly digital world.

Table 18-26. Resolution of Various VGA and Family Formats

Signal Type	Resolution
VGA	640 × 480
SVGA	800 × 600
XGA	1024 × 168
WXGA	1280 × 120
SXGA	1280 × 1024
SXGA-HD	1600 × 1200
WSXGA	1680 × 1050
QXGA	2048 × 1536
QUSXG	3840 × 2400

18.11 Digital Video

There are many formats for digital video, for both consumer, commercial and professional applications. This section concentrates on the professional applications, mainly SD-SDI (standard definition–serial digital interface) and HD-SDI (high-definition–serial digital interface.) There are sections on related consumer standards such as DVI (Section 18.9.4.1) and HDMI (Section 18.9.4.2).

18.11.1 Digital Signals and Digital Cable

Control communications, or data communications, uses digital signals. Digital video signals require wide bandwidth cabling. Control communications and slow-speed data communications use lower-performance cabling because they carry less information, requiring less bandwidth. High-speed data communications systems have significant overhead added to handle error correction so if data is lost, it can be re-sent. Digital video has some error correction capabilities, however, if all of the data bits required to make the system work are not received, picture quality is reduced or lost completely. Table 18-27 compares various digital formats.

18.11.2 Coax and SDI

Most professional broadcast formats (SDI and HD-SDI) are in a serial format and use a single coaxial cable with BNC connectors. Emerging higher resolution formats, such as 1080p/60, are also BNC based. Some work with smaller connectors for dense applications, such as patch panels and routers, which use subminiature connectors such as LCC, DIN 1.0/2.3 or DIN 1.0/2.5. Proprietary miniature BNC connectors are also available.

Table 18-27. Comparing Coaxial Digital Formats

Standard	Format	Intended Use	Connector Style	Cable Type	Transmission Distance[2]	Sample Rate	Data Rate (Mbps)	Guiding Document
SDI	serial	broadcast	one BNC	coax[1]	300m/1000ft	27MHz	270	SMPTE 259
SDTI	serial	data transport	one BNC	coax[1]	300m/1000ft	variable	270 or 360	SMPTE 305
SDTV	serial	broadcast	one BNC	coax[1]	300m/1000ft	27MHz	3 to 8	ATSC; N53
HDTV	serial	broadcast	one BNC	coax[1]	122m/400ft	74.25MHz	19.4	ATSC; A/53
HD-SDI	serial	broadcast	one BNC	coax[1]	122m/400ft	74.25MHz	1500	SMPTE 292M
1080p/60	serial	Master format	one BNC	coax[1]	80m/250ft	148.5MHz	3000	SMPTE 424M
1080p/60	serial	Master format	two BNC	dual coax	122m/400ft	74.25MHz	1500 × 2	SMPTE 424M

[1] Also implemented over fiber systems

[2] Transmission distances may vary widely depending on cabling and the specific equipment involved.

18.11.3 Cables and SDI

The most common form of SDI, component SDI, operates at data rates of 270Mbps with a clock frequency of 270MHz (occupied bandwidth–135MHz). Cable loss specifications for standard SDI are specified in SMPTE 259M and ITUR BT.601. The maximum cable length is specified as 30dB signal loss at one-half the clock frequency and is acceptable because serial digital receivers have signal recovery processing.

HD-SDI, whose cable loss is governed by SMPTE 292M, operates at a data rate of 1.5Gbps (clock 1.5GHz, occupied bandwidth 750MHz). The maximum cable length is specified at 20dB signal loss at one-half the clock frequency. 1080p/60 applications (sometimes called 3G or 3G-SDI) are covered under SMPTE 424M. The data rate is 3Gbps (clock 3GHz, occupied bandwidth 1.5GHz).

18.11.4 Receiver Quality

The quality of the receiver is important in the final performance of a serial digital system. The receiver has a greater ability to equalize, reclock, and recover the signal with SDI signals. SMPTE 292M describes the minimum capabilities of a type A receiver and a type B receiver. SDI receivers are considered adaptive because of their ability to amplify, equalize, and filter the information. Rise time is significantly affected by distance, and all quality receivers can recover the signal from a run of HD-SDI RG-6 (such as Belden 1694A) for a minimum distance of 122m (400ft). The most important losses that affect serial digital are rise time/fall time degradation and signal jitter. Serial digital signals normally undergo reshaping and reclocking as they pass through major network hubs or matrix routers.

Table 18-28 gives the specifications mandated in SMPTE 259M and SMPTE 292M in terms of rise/fall time performance and jitter. If the system provides this level of performance at the end of the cable run, the SDI receiver should be able to decode the signal.

18.11.5 Serial Digital Video

Serial digital video (SDI) falls under standards by the Society of Motion Picture and Television Engineers (SMPTE) and ITU and falls under the following categories:

SMPTE 259M	Digital video transmissions of composite NTSC 143Mb/s (Level A) and PAL 177Mb/s (Level B). It also covers 525/625 component transmissions of 270Mb/s (Level C) and 360Mb/s (Level D).
SMPTE 292M	HDTV transmissions at 1.485Gb/s
SMPTE 344M	Component widescreen transmission of 540Mb/s
ITU-R BT.601	International standard for PAL transmissions of 177Mb/s

These standards can work with standard analog video coax cables, however, the newer digital cables provide the more precise electrical characteristics required for high-frequency transmission.

SDI cable utilizes a solid bare-copper center-conductor which improves impedance stability and reduced return loss (*RL*). Digital transmissions contain both low-frequency and high-frequency signals so it is imperative that a solid-copper center-conductor is used rather than a copper-clad steel center conductor. This allows the low frequencies to travel down the center of the conductor and the high frequencies to travel on the outside of the conductor due to the skin effect. Since digital video consists of both low and high frequencies, foil shields work best. All SDI cable should be sweep tested for return loss to the third harmonic of the occupied bandwidth (Nyquist limit). For HD-SDI which is 1.485Gb/s or has a 750MHz occupied bandwidth, the cable is swept at 2.25GHz. *RL* can be no greater than 15dB at this frequency.

Table 18-28. SMPTE Serial Digital Performance Specifications

Parameter	SMPTE 259				SMPTE 292M	
	Level A NTSC 4fsc Composite	Level B PAL 4fsc Composite	Level C 525/625 Component	Level D 525/625 Component	Level D 1920 × 1080 Interlaced	Level L 1280 × 720 Progressive
Data Rate in Mbps (clock)	143	177	270	360	1485	1485
½ Clock Rate in MHz	71.5	88.5	135	180	742.5	742.5
Signal Amplitude (p-p)	800mV	800mV	800mV	800mV	800mV	800mV
dc Offset (volts)	0 ±0.5	0 ±0.5	0 ±0.5	0 ±0.5	0 ±0.5	0 ±0.5
Rise/Fall Time Max. (ns)	1.50	1.50	1.50	1.5	0.27	0.27
Rise/Fall Time Min. (ns)	0.40	0.40	0.40	0.40	–	–
Rise/Fall Time Differential (ns)	0.5	0.5	0.5	0.5	0.10	0.10
% Overshoot Max.	10	10	10	10	10	10
Timing Jitter (ns)	1.40	1.13	0.74	0.56	0.67	0.67
Alignment Jitter (ns)	1.40	1.13	0.74	0.56	0.13	0.13

BNC 50Ω connectors are often used to terminate digital video lines. This is probably acceptable for SD-SDI if only one or two connectors are used. However, for HD-SDI and above, 75Ω connectors are required to reduce RL. Cable, connectors and all passive components should exhibit a stable 75Ω impedance out to 2.25GHz, the third harmonic of the 750MHz occupied bandwidth. For 3G-SDI, with an occupied bandwidth of 1.5GHz, the third harmonic is 4.5GHz, all passive components should be tested and verified to that frequency.

18.12 Radio Guide Designations

From the late 1930s the U.S. Army and Navy began to classify different cables by their constructions. Since the intent of these high-frequency cables, both coaxes and twisted pairs, was to guide radio frequency signals, they carried the designation RG for radio guide.

There is no correlation between the number assigned and any construction factor of the cable. Thus an RG-8 came after an RG-7 and before an RG-9, but could be completely different and unrelated designs. For all intents and purposes, the number simply represents the page number in a book of designs. The point was to get a specific cable design, with predictable performance, when ordered for military applications.

As cable designs changed, with new materials and manufacturing techniques, variations on the original RG designs began to be manufactured. Some of these were specific targeted improvement, such as a special jacket on an existing design. These variations are noted by an additional letter on the designation. Thus RG-58C would be the third variant on the design of RG-58.

The test procedure for many of these military cables is often long, complicated, and expensive. For the commercial user of these cables, this is a needless expense. So many manufacturers began to make cables that were identical to the original RG specification except for testing. These were then designated utility grade and a slash plus the letter U is placed at the end. RG-58C/U is the utility version of RG-58C, identical in construction but not in testing.

Often the word *type* is included in the RG designation. This indicates that the cable under consideration is based on one of the earlier military standards but differs from the original design in some significant way. At this point, all the designation is telling the installer is that the cable falls into a family of cables. It might indicate the size of the center conductor, the impedance, and some aspects of construction, with the key word being *might*.

By the time the RG system approached RG-500, with blocks of numbers abandoned in earlier designs, the system became so unwieldy and unworkable that the military abandoned it in the 1970s and instituted MIL-C-17 (Army) and JAN C-17 (Navy) designations that continue to this day. RG-6, for instance, is found under MIL-C-17G.

18.13 Velocity of Propagation

Velocity of propagation, abbreviated V_p, is the ratio of the speed of transmission through the cable versus the speed of light in free space, about 186,282 miles per second (mi/s) or 299,792,458 meters per second (m/s). For simplicity, this is usually rounded up to 300,000,000 meters per second (m/s). Velocity of propagation is a good indication of the quality of the cable. Solid polyethylene has a V_p of 66%. Chemically formed foam has a V_p of 78%, and nitrogen gas injected foam has a V_p up to 86%, with current manufacturing techniques. Some hardline, which is mostly dry air or nitrogen dielectric, can exceed 95% velocity.

Velocity of propagation is the velocity of the signal as it travels from one end of a transmission line to the other end. A transmission line, like all electrical circuits, possesses three inherent properties: resistance, inductance, and capacitance. All three of these properties will exist regardless of how the line is constructed. Cables and hardline can be constructed to reduce these characteristics, but not eliminate them.

Under the foregoing conditions, the velocity of the electrical pulses applied to the line is slowed down in its transmission. The elements of the line are distributed evenly and are not localized or present in a lumped quantity.

The velocity of propagation (V_p) in flexible cables will vary from 50% to a V_p of 86%, depending on the insulating composition used. V_p is directly related to the dielectric constant (DC) of the insulation chosen. The equation for determining the velocity of propagation is

$$V_p = \frac{100}{\sqrt{DC}} \tag{18-5}$$

where,

V_p is the velocity of propagation,

DC is the dielectric constant.

Velocity can apply to any cable, coax or twisted pairs, although it is much more common to be expressed for cables intended for high-frequency applications. The velocity of propagation of coaxial cables is the ratio of the dielectric constant of a vacuum to the square root of the dielectric constant of the insulator, and is expressed in percent.

$$\frac{V_L}{V_S} = \frac{1}{\sqrt{\varepsilon}} \quad (18\text{-}6)$$

or

$$V_L = \frac{V_S}{\sqrt{\varepsilon}} \quad (18\text{-}7)$$

where,

V_L is the velocity of propagation in the transmission line,

V_s is the velocity of propagation in free space,

ε is the dielectric constant of the transmission line insulation.

Various dielectric constants (e) are given in Table 18-29.

Table 18-29. Dielectric Constants (e)

Material	Dielectric Constant
Vacuum	1.00
Air	1.0167
Gas injected foam PE as low as	1.35
Teflon	2.1
Polyethylene	2.25
Polypropylene	2.3
PVC	3.0 to 6.5

18.14 Shielding

From outdoor news gathering to studios and control rooms to sound reinforcement systems, the audio industry faces critical challenges from EM/RF interference (EMI and RFI). Shielding cable and twisting pairs insures signal integrity and provides confidence in audio and video transmissions, preventing downtime and maintaining sound and picture clarity.

Cables can be shielded or unshielded, except for coaxial cable which is, by definition, a precise constructions of a shielded single conductor. There are a number of shield constructions available. Here are the most common.

18.14.1 Serve or Spiral Shields

Serve or spiral shield are the simplest of all wire-based shields. The wire is simply wound around the inner portions of the cable. Spiral shields can be either single or double spirals. They are more flexible than braided shields and are easier to terminate. Since spiral shields are, in essence, coils of wire, they can exhibit inductive effects which make them ineffective at higher frequencies. Therefore, spiral/serve shields are relegated to low frequencies and are rarely used for frequencies above analog audio. Serve or spiral shields tend to open up when the cable is bent or flexed. So shield effectiveness is less than ideal, especially at high frequencies.

18.14.2 Double Serve Shields

Serve or spiral shields can be improved by adding a second layer. Most often, this is run at a 90° angle to the original spiral. This does improve coverage although the tendency to open up is not significantly improved and so this is still relegated to low-frequency or analog audio applications. This double serve or spiral construction is also called a *Reussen* shield (pronounced roy-sen).

18.14.3 French Braid™

The French Braid shield by Belden is an ultraflexible double spiral shield consisting of two spirals of bare or tinned copper conductors tied together with one weave. The shield provides the long flex life of spiral shields and greater flexibility than braided shields. It also has about 50% less microphonic and triboelectric noise. Because the two layers are woven along one axis, they cannot open up as dual spiral/serve constructions can. So French Braid shields are effective up to high frequencies, and are used up to the Gigahertz range of frequencies.

18.14.4 Braid

Braid shields provide superior structural integrity while maintaining good flexibility and flex life. These shields are ideal for minimizing low-frequency interference and have lower dc resistance than foil. Braid shields are effective at low frequencies, as well as RF ranges. Generally, the higher the braid coverage, the more effective the shield. The maximum coverage of a single braid shield is approximately 95%. The coverage of a dual braid shield can be as much as 98%. One hundred percent coverage with a braid is not physically possible.

18.14.5 Foil

Foil shields can be made of bare metal, such as a bare copper shield layer, but more common is an aluminum-polyester foil. Foil shields can offer 100% coverage. Some cables feature a loose polyester-foil layer. Other designs can bond the foil to either the core of the cable or to the inside of the jacket of the cable. Each of these presents challenges and opportunities.

The foil layer can either face out, or it can be reversed and face in. Since foil shields are too thin to be used as a connection point, a bare wire runs on the foil side of the shield. If the foil faces out, the drain wire must also be on the outside of the foil. If the foil layer faces in, then the drain wire must also be inside the foil, adjacent to the pair.

Unbonded foil can be easily removed after cutting or stripping. Many broadcasters prefer unbonded foil layers in coaxial cable to help prevent thin slices of foil that can short out BNC connectors. If the foil is bonded to the core, the stripping process must be much more accurate to prevent creating a thin slice of core-and-foil.

However, with F connectors, which are pushed onto the end of the coax, unbonded foil can bunch up and prevent correct seating of these connectors. This explains why virtually all coaxes for broadband/CATV applications have the foil bonded to the core—so F connectors easily slip on.

In shielded paired cables, such as analog or digital audio paired cables, the foil shield wraps around the pair. Once the jacket has been stripped off, the next step is to remove the foil shield. These cables are also available where the foil is bonded (glued) to the inside of the jacket. When the jacket is removed, the foil is also removed, dramatically speeding up the process. Bonded foil pairs then require that the foil be facing inside the cable and the drain wire be inside with the pair. There is some evidence that having the drain wire inside with the pair affects the symmetry of the pair, especially at high frequencies. In that case, a non-bonded foil with outside drain wire, or a symmetrical braid around the foil, would be preferred. See also Section 18.15.

A shorting fold technique is often used to maintain metal-to-metal contact for improved high-frequency performance. Without the shorting fold, a slot is created through which signals can leak. A isolation fold also helps prevent the shield of one pair contacting the shield of an adjacent pair in a multipair construction. Such contact significantly increases crosstalk between these pairs.

An improvement on the traditional shorting fold used by Belden employs the Z-Fold™, designed for use in multipair applications to reduce crosstalk, Fig. 18-17. The Z-Fold combines an isolation fold and a shorting fold. The shorting fold provides metal-to-metal contact while the isolation fold keeps shields from shorting to one another in multipair, individually shielded cables.

Since the wavelength of high frequencies can eventually work through the holes in a braid, foil shields are most effective at those high frequencies. Essentially, foil shields represent a skin shield at high frequencies, where skin effect predominates.

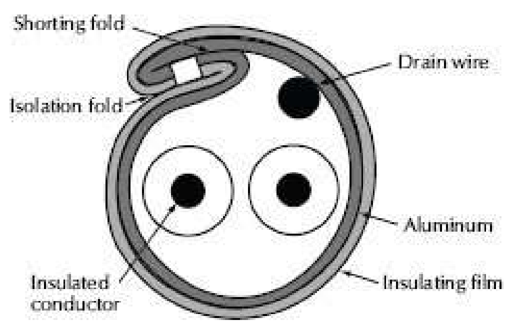


Figure 18-17. Z-Fold foil type shielded wire improves high frequency performance. Courtesy Belden.

18.14.6 Combination Shields

Combination shields consist of more than one layer of shielding. They provide maximum shield efficiency across the frequency spectrum. The combination foil-braid shield combines the advantages of 100% foil coverage and the strength and low dc resistance of a braid. Other combination shields available include various foil-braid-foil, braid-braid, and foil-spiral designs.

18.14.6.1 Foil + Serve

Because of the inductive effects of serve/spiral shields, which relegate them to low-frequency applications, this combination is rarely seen.

18.14.6.2 Foil + Braid

This is the most common combination shield. With a high-coverage braid (95%) this can be extremely effective over a wide range of frequencies, from 1kHz to many GHz. This style is commonly seen on many cables, including precision video cable.

18.14.6.3 Foil + Braid + Foil

Foil-braid-foil is often called a tri-shield. It is most commonly seen in cable television (CATV) broadband coaxial applications. The dual layers of foil are especially effective at high frequencies. However, the coverage of the braid shield in between is the key to shield effectiveness. If it is a reasonably high coverage (>80%) this style of braid will have excellent shield effectiveness.

One other advantage of tri-shield coax cable is the ability to use standard dimension F connectors since the shield is essentially the same thickness as the common foil + braid shield of less expensive cables. This style of shield is beginning to show up in precision video cables, especially those used for 3G-SD (1080p/60) which are tested to 4.5Hz. One example is Belden 1794A.

18.14.6.4 Foil + Braid + Foil + Braid

Foil-braid-foil-braid is often called *quad-shield* or just *quad* (not to be confused with starquad microphone cable or old POTS quad telephone hookup cable). Like tri-shield above, this is most common in cable television (CATV) broadband coaxial applications. Many believe this to be the ultimate in shield effectiveness. However, this is often untrue.

If the two braids in this construction are high coverage braids (>80%) then, yes, this would be an exceptional cable. But most quad-shield cable uses two braids that are 40% and 60% coverage, respectively. With that construction, the tri-shield with an 80%+ braid is measurably superior. Further, quad-shield coaxial cables are considerably bigger in diameter and therefore require special connectors.

Table 18-30 shows the shield effectiveness of different shield constructions at various frequencies. This table shows "Transfer Impedance" in ohms, the lower the number, the better the shield effectiveness. Note that all the braids measured are aluminum braids except for the last cable mentioned. That last cable is a digital precision video (such as Belden 1694A) and is many times the cost of any of the other cables listed.

Table 18-30. Shield Effectiveness of Different Shield Constructions

Shield Type (Aluminum Braid)	5 MHz	10 MHz	50 MHz	100 MHz	500 MHz
60% braid, bonded foil	20	15	11	20	50
60% braid, tri-shield	3	2	0.8	2	12
60%/40% quad shield	2	0.8	0.2	0.3	10
77% braid, tri-shield	1	0.6	0.1	0.2	2
95% copper braid, foil	1	0.5	0.08	0.09	1

18.15 Shield Current Induced Noise

There is significant evidence that constructions that feature bonded foil with an internal drain wire may affect the performance of the pairs, especially at high frequencies. Since an ideal balanced line is one where the two wires are electrically identical, having a drain wire in proximity would certainly seem to affect the symmetry of the pair. This would be especially critical where strong RF fields are around audio cables.

Despite this evidence, there are very few cables made with appropriate symmetry. This may be based on lack of end-user demand, as manufacturers would be glad to redesign their cables should the demand arise. The drain wire could be easily substituted with a symmetrical low-coverage braid, for instance.

18.16 Grounds of Shields

With any combination shield, the braid portion is the part that is making the connection. Even if we are shielding against high-frequency noise, in which case the foil is doing the actual work, the noise gets to ground by way of the braid which is much lower in resistance than the foil.

Where the foil uses a drain wire, it is that drain wire that is the shield connection. Therefore, that drain wire must be bare so it can make contact with the foil. If the foil is floating, not glued or bonded to the core of the cable, then another plastic layer is used to carry the foil. The foil itself is much too thin and weak to even be applied in the factory by itself. The second plastic layer adds enough strength and flex-life (flexes until failure) to allow the foil to be used.

The drain wire, therefore, must be in contact with the foil. In some cables, the foil faces out, so the drain wire must be on the outside of the foil, between the foil and the jacket. If the foil faces in, then the drain wire must be on the inside of the foil, adjacent to the pair (or other components) inside the cable.

With an internal drain wire, there are a number of additional considerations. One is SCIN, shield current induced noise, mentioned earlier in Section 18.15. Another is the ability to make a multipair shielded cable where the shields are facing in and the plastic facing out. This allows the manufacturer to color code the pairs by coloring the plastic holding the foil.

If you have a multipair cable, with individual foil shields, it is important that these foil shields do not touch. If the shields touch, then any signal or noise that is on one foil will be instantly shared by the other. You might as well put a foil shield in common around both pairs. Therefore, it is common to use foil shields facing in which will help prevent them from touching. These can then be color coded by using various colors of plastic with each foil to help identify each pair.

However, simply coiling the foil around the pair still leaves the very edge of the foil exposed. In a multipair cable with many individual foils, where the cable is bent and flexed to be installed, it would be quite easy for the edge of one foil to touch the edge of another foil, thus compromising shield effectiveness. The solution for this is a Z-fold invented by Belden in 1960, shown in Fig. 18-17. This does not allow any foil edge to be exposed no matter how the cable is flexed.

18.16.1 Ground Loops

In many installations, the ground potential between one rack and another, or between one point in a building and another, may be different. If the building can be installed with a star ground, the ground potential will be identical throughout the building. Then the connection of any two points will have no potential difference.

When two points are connected that do have a potential difference, this causes a ground loop. A ground loop is the flow of electricity down a ground wire from one point to another. Any RF or other interference on a rack or on an equipment chassis connected to ground will now flow down this ground wire, turning that foil or braid shield into an antenna and feeding that noise into the twisted pair. Instead of a small area of interference, such as where wires cross each other, a ground loop can use the entire length of the run to introduce noise.

If one cannot afford the time or cost of a star ground system, there are still two options. The first option is to cut the ground at one end of the cable. This is called a telescopic ground.

18.16.2 Telescopic Grounds

Where a cable has a ground point at each end, disconnecting one end produces a telescopic ground. Installers should be cautioned to disconnect only the destination (load) end of the cable, leaving the source end connected.

For audio applications, the effect of telescopic grounds will eliminate a ground loop, but at a 50% reduction in shield effectiveness (one wire now connected instead of two). If one disconnects the source end, which in analog audio is the low-impedance end, and maintains the destination (load) connection, this will produce a very effective R-L-C filter at audio frequencies.

At higher frequencies, such as data cables, even a source-only telescopic shield can have some serious problems. Fig. 18-18 shows the effect of a telescopic ground on a Cat 6 data cable. The left column shows the input impedance, the impedance presented to any RF traveling on the shield, at frequency F_k (bottom scale).

You will note that at every half-wavelength, the shield acts like an open circuit. Since most audio cables are foil shielded, and the foil is effective only at high frequencies, this means that even a correctly terminated telescopic shield is less effective at RF frequencies.

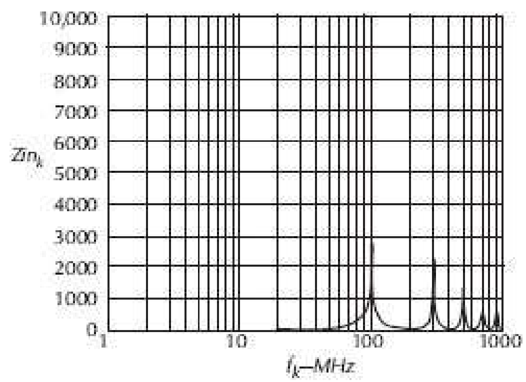


Figure 18-18. Effect of a telescopic ground on a Cat 6 cable.

18.17 UTP and Audio

One other solution for ground loops is to have no ground connection. For the seasoned audio and video professional, this solution may require a leap of faith. In can clearly be seen that, with a cable that has no shield, no drain wire, and no ground wire, so no ground loop can develop. This is a common form of data cable called UTP, unshielded twisted pairs.

With such a cable, having no shield means that you are totally dependent on the balanced line to reject noise. This is especially true, where you wish to use the four pairs in a Cat 5e, 6, or 6a cable to run four unrelated audio channels. Tests were performed on low-performance (stranded) Cat 5e patch cable (Belden 752A) looking at crosstalk between the pairs. This test shows the average of all possible pair combinations, the worst possible case, and covered a bandwidth of 1kHz to 50kHz. The results are shown in Fig. 18-19.

You will note that the worst case is around 40kHz where the crosstalk is slightly better than −95dB. In the range of common audible frequencies (20kHz) the pair-to-pair crosstalk approaches −100dB. Since a noise floor of −90dB is today considered wholly acceptable, a measurement of −95dB or −100dB is even better still.

A number of data engineers questioned these numbers based on the fact that these measurements were FEXT, far-end crosstalk, where the signals are weakest in such a cable. So measurements were also taken of NEXT, near-end crosstalk, where the signals are strongest. Those measurements are shown in Fig. 18-20.

The NEXT measurements are even better than the previous FEXT measurements. In this case, the worst case is exactly −95dB at just under 50kHz. At 20kHz and below, the numbers are even better than the previous graph, around −100dB or better.

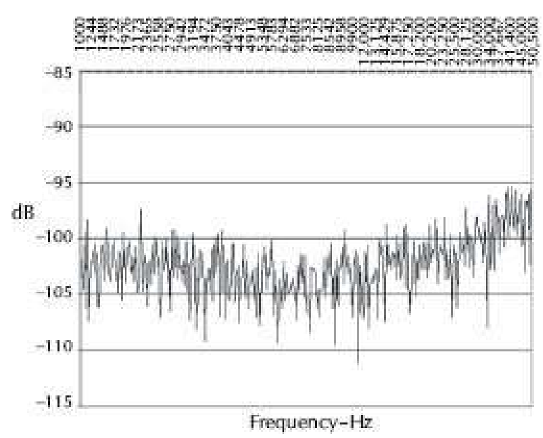


Figure 18-19. Crosstalk between Cat 5e patch cable.

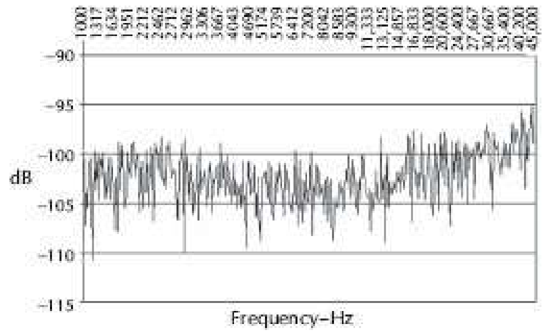


Figure 18-20. NEXT crosstalk.

There were attempts made to test a much better cable (Belden 1872A MediaTwist). This unshielded twisted-pair cable is now a Cat 6 bonded-pair design. After weeks of effort, it was determined that the pair-to-pair crosstalk could not be read on an Agilent 8714ES network analyzer. The crosstalk was somewhere below the noise floor of the test gear. The noise floor of that instrument is −110dB. With a good cable, the crosstalk is somewhere below −110dB.

So Why Shields?

These experiments with unshielded cable beg the question, why have a shield? In fact, the answer is somewhat reversed. The pairs in data cables are dramatically improved over the historic audio pairs. The bandwidth alone, 500MHz for Cat 6a, for instance, indicates that these are not the same old pairs but something different. In fact, what has happened is that the wire and cable (and data) industries have fixed the pairs.

Before, with a poorly manufactured pair, a shield would help prevent signals from getting into, or leaking out of, a pair. The fact that either effect, ingress or egress, occurred indicated the poor balance, the poor performance of the pair.

This does not mean shields are dead. There are data cables with overall shields (FTP), even individually shielded pairs (Cat 7) common in Europe. However, these are subject to the same problems as all shielded, grounded cables in terms of ground loops and wavelength effects as shown in sections 18.8.6.5 and 18.8.6.6.

The truth to the efficacy of unshielded twisted pairs running audio, video, data and many other signals is commonplace today. Many audio devices routinely use UTP for analog and digital connections. Where the source is not a balanced line, a device must change from balanced (UTP) to unbalanced (coax, for instance). Such a device matches *Bal*anced-to-*Un*balanced and is therefore called a balun. There is more on baluns in Section 18.9.3.6.7, Baluns.

18.18 AES/EBU Digital Audio Cable

Digital technology has been around for decades, but until recently it has not been used much for audio. This has now changed and digital audio is overtaking analog audio. For this reason it is important that the cable used for digital signals meet the digital requirements. The Audio Engineering Society (AES) and the European Broadcast Union (EBU) have set standards for digital audio cable. The most common sampling rates and equivalent bandwidth are shows in Table 18-31.

It is important that the line impedance be maintained to eliminate reflections that degrade the signal beyond recovery. Standard analog cable can be used for runs under 50ft (15m) but beyond that, reliability decreases. At high sampling rates, failure due to use of analog cable can be as short at 5m (16ft). The impedance and capacitance of analog cable is 40 to 70Ω and 20 to 50pF/ft. The impedance and capacitance for digital cable is 110Ω and 13pF/ft with a velocity of propagation of 78%. Proper impedance match and low capacitance are required so the square wave signal is not distorted, reflected, or attenuated.

Digital audio cable is most often 24AWG (7 × 32) tinned copper wire with short overall twist lengths, low-loss foam insulation, and 100% aluminum polyester foil shield for permanent installations. Braided shields are also available for portable use. If required, 22 to 26 wire can be obtained. Digital audio cable also comes in multiple pairs with each pair individually shielded, and often jacketed, allowing each pair and its shield to be completely isolated from the others. One pair is capable of carrying two channels of digital audio. Cables are terminated with either XLR connectors or are punched down or soldered in patch panels.

AES/EBU Digital Coaxial Cable

Digital audio requires a much wider bandwidth than analog. As the sampling rate doubles, the bandwidth also doubles, as shown in Table 18-31.

Table 18-31 Sampling Rate versus Bandwidth

Sampling Rate kHz	Bandwidth MHz	Sampling Rate kHz	Bandwidth MHz
32.0	4.096	48.0	6.144
38.0	4.864	96.0	12.228
44.1	5.6448	192.0	24.576

While still part of the standard, 32 and 38kHz sampling are rarely used. The vast majority of all applications use one of the following three; 44.1, 48, or 96kHz sampling. Digital audio can be transmitted farther distances over coax than over twisted pairs. The coax should have a 75Ω impedance, a solid copper center conductor, and have at least 90% shield coverage. When transmitting audio over an unbalanced coax line, the use of baluns may be required to change from balanced to unbalanced and back unless the device contains AES/EBU unbalanced coax inputs and outputs. The baluns change the impedance from 110Ω balanced to 75Ω unbalanced and back.

18.19 Triboelectric Noise

Noise comes in a variety of types such as EMI (electromagnetic interference) and RFI (radio frequency interference). There are also other kinds of noise problems that concern cables. These are mechanically generated or mechanically induced noise, commonly called *triboelectric noise*.

Triboelectric noise is generated by mechanical motion of a cable causing the wires inside the shield to rub against each other. Triboelectric noise is actually small electrical discharges created when conductors position changes relative to each other. This movement sets up tiny capacitive changes that eventually pop. Highly amplified audio can pick this up.

Fillers, nonconductive elements placed around the conductors, help keep the conductor spacing constant while semiconductive materials, such as carbon-impregnated cloth or carbon-plastic layers, help dissipate charge buildup. Triboelectric noise is measured through low noise test equipment using three low noise standards: NBS, ISA-S, and MIL-C-17.

Mechanically induced noise is a critical and frequent concern in the use of high-impedance cables such as guitar cords and unbalanced microphone cables that are constantly moving. The properties of special conductive tapes and insulations are often employed to help prevent mechanically induced noise. Cable without fillers can often produce triboelectric noise. This is why premise/data category cables are not suitable for flexing, moving audio applications. There are emerging flexible tactical data cables, especially those using bonded pairs, that might be considered for these applications. The use of bonded pairs for audio, such as Belden 1353A, prevents the movement of the two wires in the pair and thus prevents triboelectric and other impact or motion-induced noise.

18.20 Conduit Fill

To find the conduit size required for any cable, or group of cables, do the following:

1. Square the OD (outside diameter) of each cable and total the results.
2. To install only one cable: multiply that number by 0.5927.
3. To install two cables: multiply by 1.0134.
4. To install three or more cables: multiply the total by 0.7854.
5. From step #2 or #3 or #4, select the conduit size with an area equal to or greater than the total area. Use the ID (inside diameter) of the conduit for this determination.

This is based on the NEC ratings of

• Single cable	53% fill
• Two cables	31% fill
• Three or more cables	40% fill

If the conduit run is 50ft to 100ft, reduce the number of cables by 15%. For each 90° bend, reduce the conduit length by 30ft. Any run over 100ft requires a pull box at some midpoint.

18.21 Long Line Audio Twisted Pairs

As can be seen in Table 18-33, low frequency signals, such as audio, rarely go a quarter-wavelength and, therefore, the attributes of a transmission line, such as the determination of the impedance and the loading/matching of that line, are not considered.

However, long twisted pairs are common for telephone and similar applications, and now apply for moderate data rate, such as DSL. A twisted-pair transmission line is loaded at stated intervals by connecting an inductance in series with the line. Two types of loading are in general usage—lumped and continuous. Loading a line increases the impedance of the line, thereby decreasing the series loss because of the conductor resistance.

Although loading decreases the attenuation and distortion and permits a more uniform frequency characteristic, it increases the shunt losses caused by leakage. Loading also causes the line to have a cutoff frequency above which the loss becomes excessive. In a continuously loaded line, loading is obtained by wrapping the complete cable with a high-permeability magnetic tape or wire. The inductance is distributed evenly along the line, causing it to behave as a line with distributed constants.

In the lumped loading method, toroidal wound coils are placed at equally spaced intervals along the line, as shown in Fig. 18-21. Each coil has an inductance on the order of 88mH. The insulation between the line conductors and ground must be extremely good if the coils are to function properly.

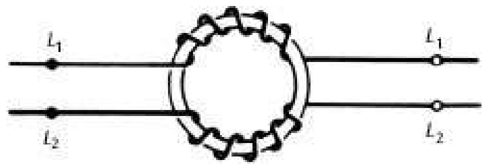


Figure 18-21. Loading coil connected in a balanced transmission line.

Loading coils will increase the talking distance by 35 to 90 miles for the average telephone line.

If a high-frequency cable is not properly terminated, some of the transmitted signal will be reflected back toward the transmitter, reducing the output.

18.22 Delay and Delay Skew

The fact that every cable has a velocity of propagation, obviously means that it takes time for a signal to go down a cable. That time is called *delay*, normally measured in nanoseconds (*DN*). V_p can easily be converted into delay. Since V_p is directly related to dielectric constant (*DC*), they are all directly related as shown in Eq. 18-8 and determine the delay in nanoseconds-per-foot (ns/ft).

$$Dn = \frac{100}{V_p} = \sqrt{DC} \qquad (18\text{-}8)$$

While these equations will give you a reasonable approximate value, the actual equations should be

$$Delay = \frac{101.67164}{V_p} = 1.0167164\sqrt{DC} \qquad (18\text{-}9)$$

Delay becomes a factor in broadcasting when multiple cables carry a single signal. This commonly occurs in RGB, VGA, or other analog component video delivery systems. It is also a factor for HDMI, Display-Port, and similar systems that divide the data between multiple pairs. Delay also appears in high-data rate UTP, such a 1000BaseT (1GBaseT) and beyond where data is split between the four pairs and combined at the destination device.

Where signals are split up and recombined, the different cables supplying the components will each have a measurable delay. The trick is for all the component cables to have the *same delay* to deliver their portions at the same time. The de facto maximum timing variation in delay for RGB analog is delivery of all components within 40ns. Measuring and adjusting cable delivery is often called timing. By coincidence, the maximum delay difference in the data world is 45ns, amazingly close. In the data world, this is called *skew* or *delay skew*, where delivery does not line up.

In the RGB world, where separate coax cables are used, they have to be cut to the same electrical length. This is not necessarily the same physical length. Most often, the individual cables are compared by a Vector-scope, which can show the relationship between components, or a TDR (time domain reflectometer) that can establish the electrical length (delay) of any cable.

Any difference in physical versus electrical length can be accounted for by the velocity of propagation of the individual coaxes, and therefore, the consistency of manufacture. If the manufacturing consistency is excellent, then the velocity of all coaxes would be the same, and the physical length would be the same as the electrical length. Where cables are purchased with different color jackets, to easily identify the components, they are obviously made at different times in the factory. It is then a real test of quality and consistency to see how close the electrical length matches the physical length.

Where cables are bundled together, the installer then has a much more difficult time in reducing any timing errors. Certainly in UTP data cables, there is no way to adjust the length of any particular pair. In all these bundled cables, the installer must cut and connectorize.

This becomes a consideration when four-pair UTP data cables (category cables) are used to deliver RGB, VGA, and other nondata component delivery systems. The distance possible on these cables is therefore based on the attenuation of the cables at the frequency of operation, and on the delay skew of the pairs. Therefore, the manufacturers measurement and guarantee (if any) of delay skew should be sought if nondata component delivery is the intended application.

This same guarantee is often available for multiple coax constructions. Consult your manufacturer for details.

18.23 Attenuation

All cable has attenuation and the attenuation varies with frequency. Attenuation can be found with the equation

$$A = 4.35\frac{R_t}{Z_o} + 2.78pf\sqrt{\varepsilon} \tag{18-10}$$

where,

A is the attenuation in dB/100ft,

R_t is the total dc line resistance in Ω/100ft, ε is the dielectric constant of the transmission line insulation,

p is the power factor of the dielectric medium,

f is the frequency,

Z_o is the impedance of the cable.

Table 18-32 gives the attenuation for various 50Ω, 52Ω, and 75Ω cables. The difference in attenuation is due to either the dielectric of the cable or center-conductor diameter.

18.24 Characteristic Impedance of a Cable

The *characteristic impedance* of a cable is the measured impedance of a cable of infinite length. This impedance is an ac measurement, and cannot be measured with an ohmmeter. It is frequency-dependent, as can be seen in Fig. 18-22. This shows the impedance of a coaxial cable from 10Hz to 100MHz.

At low frequencies, where resistance is a major factor, the impedance is changing from a high value (approximately 4000Ω at 10Hz) down to a lower impedance. This is due to skin effect (see Section 18.2.8), where the signal is moving from the whole conductor at low frequencies to just the skin at high frequencies. Therefore, when only the skin is carrying the signal, the resistance of the conductor is of no importance. This can be clearly seen in the equations for impedance, Eq. 18-13, for low frequencies, shows R, the resistance, as a major component. For high frequencies, Eq. 18-14, there is no *R*, no resistance, even in the equation.

Table 18-32 Coaxial Cable Signal Loss (Attenuation) in dB/100ft

Frequency	RG-174/ 8216	RG-58/ 8240	RG-8X/ 9258	RG-8/ 8237	RG-8/ RF-9913F	RG-59/ 8241	RG-6/ 9248	RG-11/ 9292
1MHz	1.9	0.3	0.3	0.2	0.1	0.6	0.3	0.2
10MHz	3.3	1.1	1.0	0.6	0.4	1.1	0.7	0.5
50MHz	5.8	2.5	2.3	1.3	0.9	2.4	1.5	1.0
100MHz	8.4	3.8	3.3	1.9	1.3	3.4	2.0	1.4
200MHz	12.5	5.6	4.9	2.8	1.8	4.9	2.8	2.1
400MHz	19.0	8.4	7.6	4.2	2.7	7.0	4.0	2.9
700MHz	27.0	11.7	11.1	5.9	3.6	9.7	5.3	3.9
900MHz	31.0	13.7	13.2	6.9	4.2	11.1	6.1	4.4
1000MHz	34.0	14.5	14.3	7.4	4.5	12.0	6.5	4.7
Characteristic impedance–Ω	50.0	50.0	50.0	52.0	50.0	75.0	75.0	75.0
Velocity of propagation–%	66	66%	80%	66%	84%	66%	82%	78%
Capacitance pF/ft, pF/m	30.8/101.0	29.9/98.1	25.3/83.0	29.2/96.8	24.6/80.7	20.5/67.3	16.2/53.1	17.3/56.7

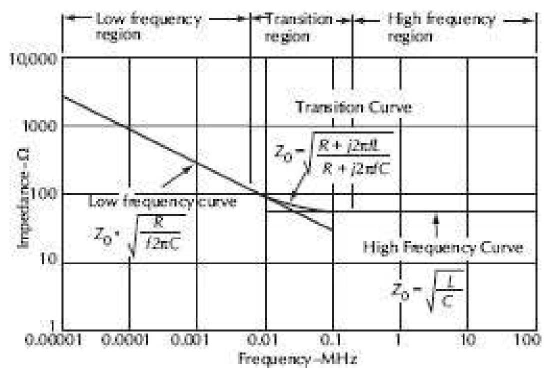


Figure 18-22. Impedance of coaxial cable from 10Hz to 100 MHz.

Once we enter that high-frequency area where resistance has no effect, around 100kHz as shown in Fig. 18-22, we enter the area where the impedance will not change. This area is called the *characteristic impedance* of the cable.

The characteristic impedance of an infinitely long cable does not change if the far end is open or shorted. Of course, it would be impossible to test this as it is impossible to short something at infinity. It is important to terminate coaxial cable with its rated impedance or a portion of the signal can reflect back to the input, reducing the efficiency of the transmission. Reflections can be caused by an improper load, using a wrong connector—i.e., using a

50Ω video BNC connector at high frequencies rather than a 75Ω connector—a flattened cable, or too tight a bend radius, which changes the spacing between the conductors. Anything that affects the dimensions of the cable, will affect the impedance and create reflective losses. It would just be a question of how much reflection is caused. Reflections thus caused are termed *return loss*.

The characteristic impedance of common coaxial cable can be between 30Ω and 200Ω. The most common values are 50Ω and 75Ω. The characteristic Z_o is the average impedance of the cable equal to

$$Z_o = \frac{138}{\sqrt{\varepsilon}} \log \frac{D}{d} \qquad (18\text{-}11)$$

where

ε is the dielectric constant,

D is the diameter of the inner surface of the outer coaxial conductor (shield) in in,

d is the diameter of the center conductor in in.

The true characteristic impedance, at any frequency, of a coaxial cable is found with the equation

$$Z_o = \sqrt{\frac{R + j2\pi fL}{G + j2\rho C}} \qquad (18\text{-}12)$$

where,

R is the series resistance of the conductor in Ω/circular mil foot per unit length,

f is the frequency in Hz,

L is the inductance in H,

G is the shunt conductance in mhos per unit length,
C is the capacitance in F.

At low frequencies, generally below 100kHz, the equation for coaxial cable simplifies to

$$Z_o = \sqrt{\frac{R}{j2\rho\pi C}} \quad (18\text{-}13)$$

At high frequencies, generally above 100kHz, the equation for coaxial cable simplifies to

$$Z_o = \sqrt{\frac{L}{C}} \quad (18\text{-}14)$$

18.25 Characteristic Impedance of a Transmission Line

The *characteristic impedance* of a transmission line is equal to the impedance that must be used to terminate the line in order to make the input impedance equal to the terminating impedance. For a line that is longer than a quarter-wavelength at the frequency of operation, the input impedance will equal the characteristic impedance of the line, irrespective of the terminating impedance.

This means that low-frequency applications often have quarter-wavelength distance way beyond common practical applications. Table 18-33 shows common signals, with the wavelength of that signal and the quarter-wavelength. To be accurate, given a specific cable type, these numbers would be multiplied by the velocity of propagation.

The question is very simple: will you be going as far as the quarter-wavelength, or farther? If so, then the characteristic impedance becomes important. As that distance gets shorter and shorter, this distance becomes critical. With smaller distances, patch cords, patch panels, and eventually the connectors themselves become just as critical as the cable. The impedance of these parts, especially when measured over the desired bandwidth, becomes a serious question. To be truly accurate, the quarter-wavelength numbers in Table 18-33 need to be multiplied by the velocity of propagation of each cable. So, in fact, the distances would be even shorter than what is shown.

It is quite possible that a cable can work fine with lower-bandwidth applications and fail when used for higher-frequency applications. The characteristic impedance will also depend on the parameters of the pair or coax cable at the applied frequency. The resistive component of the characteristic impedance is generally high at the low frequencies as compared to the reactive component, falling off with an increase of frequency, as shown in Fig. 18-22. The reactive component is high at the low frequencies and falls off as the frequency is increased.

Table 18-33. Characteristics of Various Signals

Signal Type	Bandwidth	Wave-length	Quarter-Wave-length
Analog audio	20kHz	15km	3.75km
		49,212ft	12,303ft
AES 3—44.1kHz	5.6448MHz	53.15m	13.29m
		174.4ft	43.6ft
AES 3—48kHz	6.144MHz	48.83m	12.21m
		160.2ft	39.76ft
AES 3—96kHz	12.288MHz	24.41m	6.1m
		80.1ft	20ft
AES 3—192kHz	24.576MHz	12.21m	3.05m
		40.06ft	10ft
Analog video (U.S.)	4.2MHz	71.43m	17.86m
		234.35ft	58.6ft
Analog video (PAL)	5MHz	60m	15m
		196.84ft	49.21ft
SD-SDI	270MHz clock	1.11m	27.75cm
		3.64ft	10.14in
SD-SDI	405MHz third harmonic	74cm	18.5cm
		29.13in	7.28in
HD-SDI	1.5GHz clock	20cm	5cm
		7.87in	1.97in
HD-SDI	2.25GHz third harmonic	13cm	3.25cm
		5.12in	1.28in
1080P/50-60	3GHz clock	10cm	2.5cm
		3.94in	0.984in
1080P/50-60	4.5GHz third harmonic	66mm	16.5mm
		2.6in	0.65in

The impedance of a uniform line is the impedance obtained for a long line (of infinite length). It is apparent, for a long line, the current in the line is little affected by the value of the terminating impedance at the far end of the line. If the line has an attenuation of 20dB and the far end is short circuited, the characteristic impedance as measured at the sending end will not be affected by more than 2%.

18.26 Twisted-Pair Impedance

For shielded and unshielded twisted pairs, the characteristic impedance is

$$Z_0 = \frac{101670}{C(V_p)} \qquad (18\text{-}15)$$

where,

Z_0 is the average impedance of the line,

C is found with Eqs. 18-16 and 18-17,

V_P is the velocity of propagation.

For unshielded pairs

$$C = \frac{3.68\varepsilon}{\log\left[\frac{2(ODi)}{DC(Fs)}\right]} \qquad (18\text{-}16)$$

For shielded pairs

$$C = \frac{3.68\varepsilon}{\log\left[\frac{1.06(ODi)}{DC(Fs)}\right]} \qquad (18\text{-}17)$$

where,

ε is the dielectric constant,

ODi is the outside diameter of the insulation,

DC is the conductor diameter,

Fs is the conductor stranding factor (solid = 1, 7 strand = 0.939, 19 strand = 0.97.

The impedance for higher-frequency twisted-pair data cables is

$$Z_0 = 276\left(\frac{VP}{100}\right) \times \log\left[2\left(\frac{h}{DC \times Fs}\right) \times \left(\frac{1 - \frac{h}{DC + Fb}^2}{1 + \frac{h}{DC + Fb}^2}\right)\right] \tag{18-18}$$

where,

h is the center to center conductor spacing,

Fb is very near 0. Neglecting Fb will not introduce appreciable error.

Transmission Line Termination

All lines do not need to be terminated. Knowing when to terminate a transmission line is a function of the frequency/wavelength of the signal and the length of the transmission line. Table 18-32 can be guideline, especially where the signal is long compared to the length of the line. If the wavelength of the signal is small compared to the transmission-line length, for instance a 4.5GHz signal, a terminator is required to prevent the signal from reflecting back toward the source and interfering with forward traveling signals. In this case the line must be terminated for any line longer than a quarter of a wavelength.

Transmission-line termination is accomplished using parallel or series termination. Parallel termination connects a resistor between the transmission line and ground at the receiving end of the transmission line while series termination connects a resistor in series with the signal path near the beginning of the transmission line, Fig. 18-23.

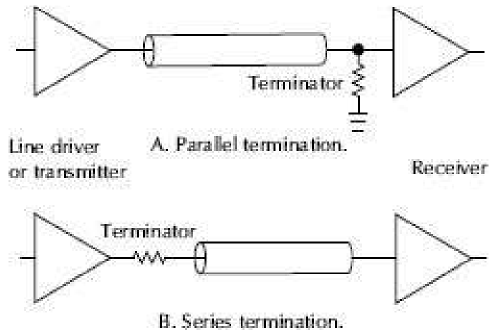


Figure 18-23. Basic termination of transmission lines.

Resistive termination requires a resistor value that matches the characteristic impedance of the transmission line, most commonly a 50Ω or 75Ω characteristic impedance. The termination resistance is matched to the transmission line characteristic impedance so the electrical energy in the signal does not reflect back from the receiving end of the line to the source. If the resistor is perfectly matched to the characteristic impedance, at all frequencies within the desired bandwidth, all of the energy in the signal dissipates as heat in the termination resistor so no signal reflects backwards down the line to the source causing cancellations. As video frequencies increase, proper termination becomes more expensive as the resistive value must be constant at these high frequencies. Older terminators must be discarded unless they can be tested to 2.25GHz (for HD) or 4.5GHz (for 3G) and be shown to be reasonably accurate. A resistor value that is ±3Ω or less at the frequency of interest should be desired for the best accuracy.

18.27 Loudspeaker Cable

Much has been said about wire for connecting loudspeakers to amplifiers. Impedance, inductance, capacitance, resistance, loading, matching, surface effects, etc. are constantly discussed.

Most home and studio loudspeaker runs are short (less than 50ft, or 15m) and therefore do not constitute a transmission line. When runs are longer, it is common to connect the loudspeakers as a 70.7V (normally called a 70V system), or 100V distributed system to reduce line loss caused by the wire resistance so the power lost in the line does not appreciably effect the power delivered to the loudspeaker. For instance, if a 4Ω loudspeaker is connected to an amplifier with a cable which measures 4Ω resistance, 50% of the power will be dissipated in the cable. If the loudspeaker was connected to a 70V system, and the loudspeaker was taking 50W from the amplifier, the loudspeaker/transformer impedance would be 100Ω; therefore, the 4Ω line resistance would dissipate 4% of the power.

When using a 70V loudspeaker system, the choice of wire size for loudspeaker lines is determined by an economic balance of the cost of copper against the cost of power lost in the line. Power taken from the amplifier is calculated from the equation

$$P = \frac{V^2}{Z} \qquad (18\text{-}19)$$

where,

P is the power delivered by the amplifier,

V is the voltage delivered by the amplifier,

Z is the impedance of the load.

For a 70V system

$$P = \frac{5000}{Z} \qquad (18\text{-}20)$$

If the voltage is 70.7V and the load is 50Ω, the power would be 100W. However, if the amplifier was connected to a 50Ω load with 1000 ft of #16 wire (2000ft round trip) or 8Ω of wire resistance the power from the amplifier would be

$$\begin{aligned} P &= \frac{5000}{50\,\Omega + 8\,\Omega} \\ &= 86.2\text{ W} \end{aligned}$$

The current through the system is found with

$$I = \sqrt{\frac{P}{R}} \tag{18-21}$$

or in this case

$$\begin{aligned} I &= \sqrt{\frac{P}{R}} \\ &= \sqrt{\frac{86.2}{58}} \\ &= 1.21\text{ A} \end{aligned}$$

The power to the 50Ω load would be found with

$$P = I^2 R \tag{18-22}$$

or in this case

$$\begin{aligned} P &= I^2 R \\ &= 1.21^2 R \\ &= 74.3\text{ W} \end{aligned}$$

Only 11.7W are lost to the line, the other 14W cannot be taken from the amplifier because of the impedance mismatch. While high-power amplifiers are relatively inexpensive, it is still practical to use heavy enough wire so the amplifier can output almost its full power to the loudspeakers. Table 18-34 shows the characteristics of various cables which could be used for loudspeaker wire and Table 18-35 is a cable selection guide for loudspeaker cable. Fig. 18-24 and Eq. 18-23 were used to calculate Table 18-35.

Table 18-34. Frequency Limitations for 33ft (10m) Lengths of Cable with Various Loads

Cable Type	Upper Corner Frequency, kHz		Resonant Measurement Phase (°)	
	2Ω Load	4Ω Load	4μF Load	4Ω Load
No. 18 zip cord	75	136	35	3
No. 16 zip cord	61	114	32	2
No. 14 speaker cable	82	156	38	2
No. 12 speaker cable	88	169	40	2
No. 12 zip cord	55	106	32	4
Welding cable	100	200	44	2
Braided cable	360	680	80	1
Coaxial dual cylindrical	670	1300	112	
Coaxial RG-8	450	880	92	

Table 18-35. Loudspeaker Cable Transmission Distance as a Function of Conductor Size vs. Loss. Courtesy Belden

AWG		Power Loss in Cable (% Loss & dB Loss)								
		4Ω Loudspeaker			8Ω Loudspeaker			70V Loudspeaker*		
% loss		11%	21%	50%	11%	21%	50%	11%	21%	50%
dB loss		0.5	1.0	3.0	0.5	1.0	3.0	0.5	1.0	3.0
6	ft	277	571	1930	554	1141	3859	13,580	27,965	94,548
	m	84	174	588	187	347	1176	4139	8524	28818
8	ft	174	359	1214	349	718	2428	8546	17,598	59,498
	m	53	109	370	106	219	740	2604	5364	18135
10	ft	110	226	764	219	452	1528	5377	11,072	37,434
	m	34	69	233	67	137	466	1639	3375	11410
12	ft	69	142	480	138	284	959	3376	6952	23,505
	m	21	43	146	42	87	292	1030	2119	7164
14	ft	43	89	302	87	179	604	2127	4380	14,809
	m	13	27	92	22	55	184	648	1335	4514
16	ft	27	55	185	53	110	371	1305	2687	9085
	m	8.2	17	56	16	34	113	398	819	2770
18	ft	17	35	117	34	69	234	823	1694	5726
	m	5.2	11	36	10.4	21	71	251	516	1745
20	ft	11	22	74	21	44	147	518	1068	3610
	m	3.4	6.7	23	6.4	13.4	45	158	325	1100
22	ft	7	13	46	13	27	91	321	661	2234
	m	2.1	4	14	4	8.2	27.7	98	201	681
24	ft	4	9	29	8	17	57	202	417	1409
	m	1.2	2.7	8.8	2.5	5.2	17.4	62	127	430

The number of feet or meters you can run for a given loss and performance budget

*70V line drives systems, while considered a potential for Hi-Fi performance, follow the same cable loss physics as higher current (lower impedance) system. For the sake of this calculation a 25W, 70V system (196Ω) was used.

To use the table:

1. Select the appropriate loudspeaker impedance column.
2. Select the appropriate power loss column deemed to be acceptable.
3. Select the applicable wire gage size and follow the row over to the columns determined in steps one and two. The number listed is the maximum cable run length.

Example. The maximum run for 12AWG in a 4W loudspeaker system with 11% or 0.5dB loss is 140ft.

Fig. 18-24 and the derivation equations, Eq. 18-23 were used to produce Table 18-33

Courtesy Belden.

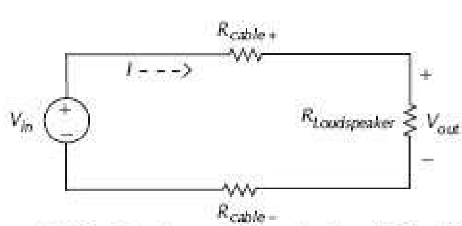


Figure 18-24. Circuit used to calculate Table 18-32. Courtesy

Belden.

$$
\begin{aligned}
R_{cable} &= R_{cable+} + R_{cable-} \\
V_{in} &= I \times (R_{Loudspeaker} + R_{Cable}) \\
V_{out} &= I \times R_{Loudspeaker} \\
V_{in} &= \frac{V_{out}}{R_{Loudspeaker} \times (R_{Loudseaker} + R_{Cable})} \\
\frac{V_{out}}{V_{in}} &= \frac{R_{loudspeaker}}{(R_{Loudspeaker} + R_{Cable})} \\
\alpha &= 20\log\left(\frac{V_{out}}{V_{in}}\right) \\
10^{\frac{\alpha}{20}} &= \frac{R_{loudspeaker}}{(R_{Loudspeaker} + R_{Cable})} \\
R_{Cable} &= R_{Loudspeaker} \times \left(10^{\frac{-\alpha}{20}} - 1\right) \\
R_{Loudspeaker} &= 10^{\frac{\alpha}{20} \times (R_{Loudspeaker} + R_{Cable})}
\end{aligned}
\tag{18-23}
$$

For simplicity, assumptions include use of tin coated copper conductors at 20°C, Minimally compliant dc resistance (ASTM), and a flat response with ideal source, cable, & load. Larger, solid, and/or uncoated conductors will transmit farther than the values presented. Use of an electrical model other than the purely resistive model shown will yield differing results. Performance in any system may vary with frequency. Damping factor, among other considerations, are outside the scope of this table and should be considered if required by the intended application. 70 volt line drive systems, while considered a potential for Hi-Fi performance, follow the same cable loss physics as the higher current (lower impedance) system. For the sake of this calculation, a 25W, 70V system (196Ω) was used.

18.27.1 Damping Factor

The damping factor of an amplifier is the ratio of the load impedance (loudspeaker plus wire resistance) to the amplifier internal output impedance. The damping factor of the amplifier acts as a short circuit to the loudspeaker, controlling the overshoot of the loudspeaker. Present day amplifiers have an output impedance of less than 0.05Ω which translates to a damping factor over 150 at 10kHz, for instance, so they effectively dampen the loudspeaker as long as the loudspeaker is connected directly to the amplifier. Damping factor is an important consideration when installing home systems, studios, or any system where high-quality sound, especially at the low frequencies, is desired. As soon as wire resistance is added to the circuit, the damping factor reduces dramatically, reducing its effect on the loudspeaker. For instance, if a 16AWG 50ft loudspeaker cable (100ft round trip) is used, the wire resistance would be 0.4Ω making the damping factor only 18, considerably less than anticipated.

It is not too important to worry about the effect the damping factor of the amplifier has on the loudspeakers in a 70V system as the 70V loudspeaker transformers wipe out the effects of the wire resistance.

Consider the line as a lump sum, Fig. 18-25. The impedance of the line varies with wire size and type. Table 18-36 gives typical values of R, *C*, and *L* for 33ft (10m) long cables. Note, the impedance at 20kHz is low for all but the smallest wire and the −3dB upper frequency is well above the audio range. The worst condition is with a capacitive load. For instance, with a 4μF load, resonance occurs around 35kHz.

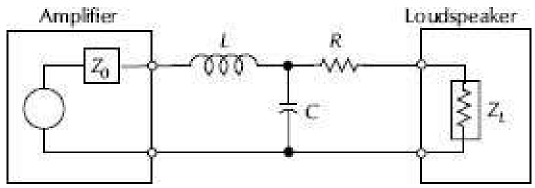


Figure 18-25. Amplifier, cable, loudspeaker circuit using lumped circuit elements to represent the properties of the cable Courtesy Belden

Table 18-36. Lumped Element Values for 33ft (10m) Lengths of Cable

Cable Type	L-μH	C-pF	R_{dc}-Ω	Z-Ω@20kHz
No. 18 zip cord	5.2	580	0.42	0.44
No. 16 zip cord	6.0	510	0.26	0.30
No. 14 speaker cable	4.3	570	0.16	0.21
No. 12 speaker cable	3.9	760	0.10	0.15
No. 12 zip cord	6.2	490	0.10	0.15
Welding cable	3.2	880	0.01	0.04
Braided cable	1.0	16,300	0.26	0.26
Coaxial dual cylindrical	0.5	58,000	0.10	0.10
Coaxial RG-8	0.8	300	0.13	0.13

The results of the above are as follows:

1. Make the amplifier to loudspeaker runs as short as possible.
2. Use a wire gage that represents less than 5% of the loudspeaker impedance at any frequency.
3. Use twisted pairs on balanced 70 or 100V distributed systems to reduce crosstalk (amplifier output is often fed back into the amplifier as negative feedback).
4. Use good connectors to reduce resistance.

18.27.2 Crosstalk

When a plurality of lines, carrying different programs or signals, are run together in the same conduit, or where multiple pairs or multiple coax cables are bundled, they tend to induce crosstalk currents into each other. Crosstalk is induced by two methods:

1. Electromagnetically through unbalanced coupling between one circuit and others.
2. Electrostatically through unbalanced capacitance to other circuits, or to the conduit if it carries current. This develops a voltage difference between one circuit and the others, or to its own or other shields carrying current.

If the line is less than a quarter-wavelength at the frequency of operation, then the cable does not have to have a specific impedance, or does not need to be terminated in a specific impedance. The terminating impedance could then be small compared to the open line characteristic impedance. The net coupling with unshielded pairs would then be predominantly magnetic. If the terminating impedance is much larger than the characteristic impedance of the wires, the net coupling will be predominantly electric.

Two wires of a pair must be twisted; this insures close spacing and aids in canceling pickup by transposition. In the measurements in Fig. 18-26, all pickup was capacitive because the twisting of the leads effectively eliminated inductive coupling.

One application that is often ignored regarding crosstalk is speaker wiring, especially 70V distributed loudspeaker wiring. You will note in the first drawing that the two wires are not a balanced line. One is hot the other is ground. Therefore, that pair would radiate some of the audio into the adjoining pair, also unbalanced.

Twisting the pairs in this application would do little to reduce crosstalk.

The test was made on a 250ft twisted pair run in the same conduit with a similar twisted pair, the latter carrying signals at 70.7V. Measurements made for half this length produced half the voltages, therefore the results at 500ft and 1000ft were interpolated.

The disturbing line was driven from the 70V terminals of a 40W amplifier and the line was loaded at the far end with 125Ω, thus transmitting 40W. The crosstalk figures are for 1kHz. The voltages at 100Hz and 10kHz are one-tenth and ten times these figures, respectively.

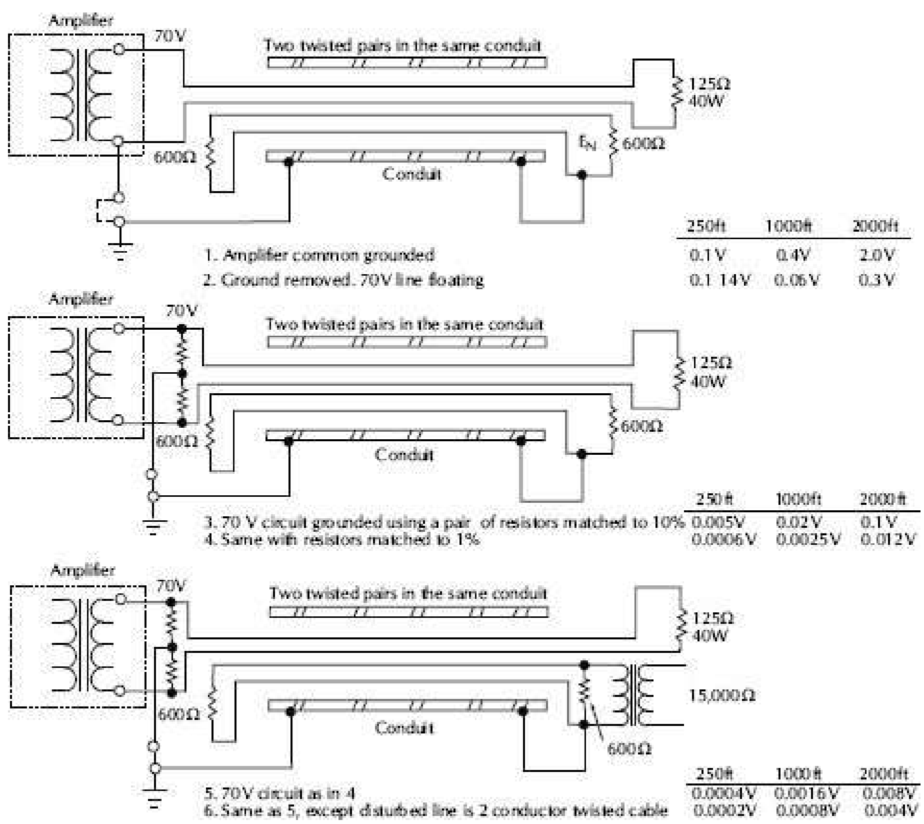


Figure 18-26. Effects of grounding on crosstalk. Courtesy Altec Lansing Corp.

There are two ways to effectively reduce crosstalk. One is to run signals only on balanced-line twisted pairs. Even shielding has a small added advantage compared to the noise and crosstalk rejection of a balanced line. The second way to reduce crosstalk is to move the two cables apart. The inverse-square law tells us that doubling the distance will produce four times less interference. Further, if cables cross at right angles, this is the point where the magnetic fields have minimum interaction. Of course, the latter solution is not an option in a prebundled cable, or in cable trays or installations with multiple cables run from point to point.

18.28 National Electrical Code

The *National Electrical Code* (NEC) is a set of guidelines written to govern the installation of wiring and equipment in commercial buildings and residential areas. These guidelines were developed to insure the safety of humans as well as property against fires and electrical hazards. Anyone involved in specifying cable for installation should be aware of the basics of the code.

The NEC code book is made up of nine chapters, with each chapter divided into separate articles pertaining to specific subjects. Five articles pertain to communication and power-limited cable. The NEC book is written by and available from the NFPA (National Fire Protection Association), 11 Tracy Drive, Avon, MA 02322. They can be reached at 1-800-344-3555 or www.nfpa.org.

Article 725—Class 1, Class 2, Class 3, Remote-Control, Signaling, and Power-Limited Circuits. Article 725 covers Class 1, Class 2, and Class 3 remote control and signaling cables as well as power-limited tray cable. Power-limited tray cable can be used as a Class 2 or Class 3 cable. Cable listed multipurpose, communications, or power-limited fire protective can be used for Class 2 and Class 3 applications. A Class 3 listed cable can be used as a Class 2 cable.

Article 760—Fire Protective Signaling Systems. Article 760 covers power-limited fire-protective cable. Cable listed as power-limited fire-protective cable can also be used as Class 2 and Class 3 cable. Cable listed as communications and Class 3 can be used as power-limited fire protective cable with restrictions to conductor material and type gage size and number of conductors.

Article 770—Fiber Optic Systems. Article 770 covers three general types of fiber optic cable: nonconductive, conductive, and composite. Nonconductive type refers to cable containing no metallic members and no other electrically conductive materials. Conductive type refers to cable containing noncurrent carrying conductive members such as metallic strength members, etc. Composite type refers to cable containing optical fibers and current carrying electrical conductors. Composite types are classified according to the type of electrical circuit that the metallic conductor is designed for.

Article 800—Communication Circuits. Article 800 covers multipurpose and communication cable. Multipurpose cable is the highest listing for a cable and can be used for communication, Class 2, Class 3, and power-limited fire-protective cable. Communication cable can be used for Class 2 and Class 3 cable and also as a power-limited fire protective cable with restrictions.

Article 820—Community Antenna Television. Article 820 covers community antenna television and RF cable. CATV cable may be substituted with multipurpose or communication listed coaxial cable.

18.28.1 Designation and Environmental Areas

The NEC has designated four categories of cable for various environments and they are listed from the highest to the lowest listing. A higher listing can be used as a substitute for a lower listing.

Plenum—Suitable for use in air ducts, plenums, and other spaces used for environmental air without conduit and has adequate fire-resistant and low-smoke producing characteristics. It can also be substituted for all applications below.

Riser—Suitable for use in a vertical run, in a shaft, or from floor to floor, and has fire-resistant characteristics capable of preventing the spread of fire from floor to floor. It can also be substituted for all applications below.

General Purpose—Suitable for general-purpose use, with the exception of risers, ducts, plenums, and other space used for environmental air, and is resistant to the spread of fire. It can be substituted for the applications below.

Restricted Applications—Limited use and suitable for use in dwellings and in raceways and is flame retardant. Restricted use is limited to nonconcealed spaces of 10ft or less, fully enclosed in conduit or raceway, or cable with diameters less than 0.25in for a residential dwelling.

18.28.2 Cable Types

Signal cable used for audio, telephone, video, control applications, and computer networks of less than 50V is considered low-voltage cabling and is grouped into five basic categories by the NEC, Table 18-37.

Table 18-37. The Five Basic NEC Cable Croups

Cable Type	Use
CM	Communications
CL2, CL3	Class 2, Class 3 remote-control, signaling, and power-limited cables
FPL	Power-limited fire-protective signaling cables
MP	Multipurpose cable
PLTC	Power-limited tray cable

All computer network and telecommunications cabling falls into the CM class. The A/V industry primarily uses CM and CL2 cabling.

Table 18-38 defines the cable markings for various applications. Note plenum rated cable is the highest level because it has the lowest fire load which means it does not readily support fire.

Table 18-38. Cable Applications Designations Hierarchy

	Cable Family					
Application	MP	CM	CL2	CL3	FPL	PLTC
Plenum	MPP	CMP	CL2P	CL3P	FPLP	–
Riser	MPR	CPR	CL2R	CL3R	FPLR	–
General Purpose	MP, MPG	CM, CMG	CL2	CL3	FPL	PLTC
Dwelling	–	CMX	CL2X	CL3X	–	–

18.28.3 NEC Substitution Chart

Fig. 18-27 defines the Canadian Electrical Code (CEC) substitution chart. NEC cable hierarchy, Fig. 18-28 defines which cables can replace other cables. The chart starts with the highest listed cable on the top and descends to the lowest listed cable on the bottom. Following the arrows defines which cable can be substituted for others.

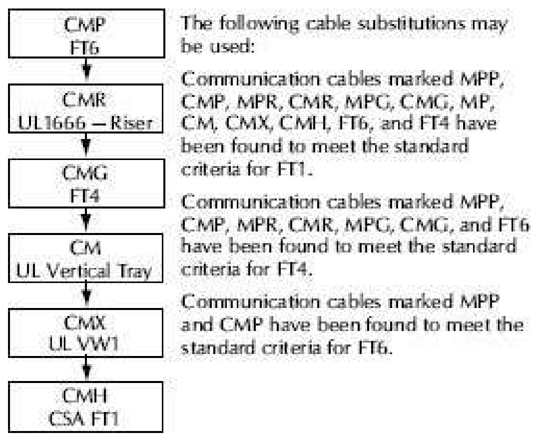


Figure 18-27. Canadian Electrical Code cable substitution hierarchy per C22.2 #214—Communication Cables.

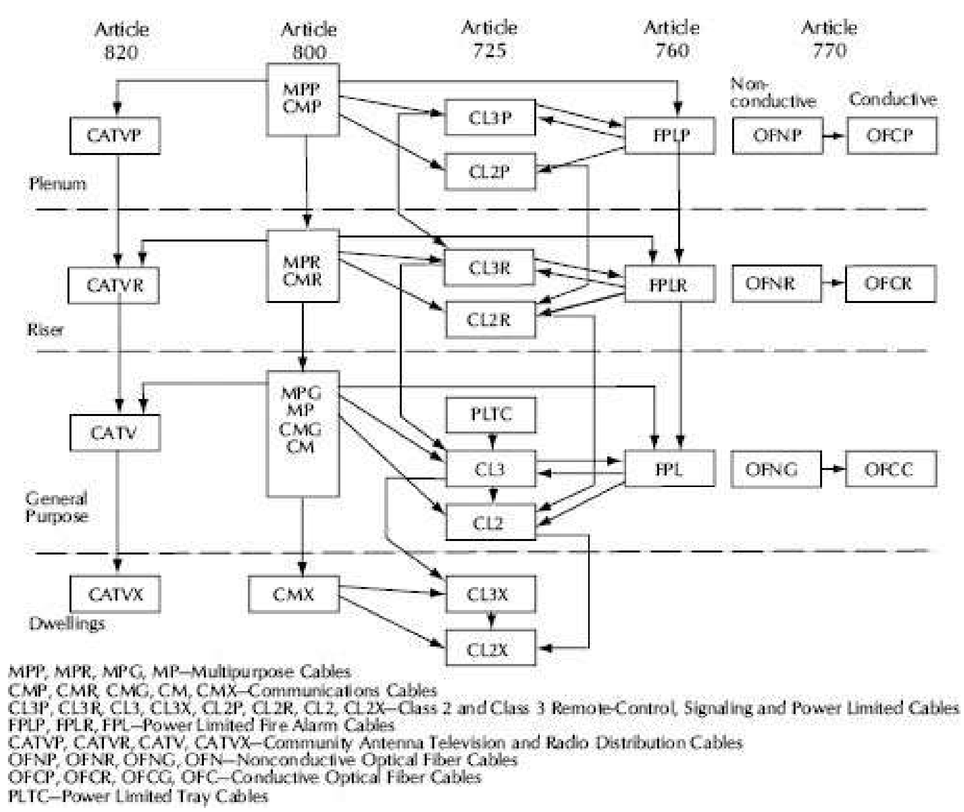


Figure 18-28. National Electrical Code substitution and hierarchy. Courtesy Belden.

18.28.4 Final Considerations

The National Electrical Code is widely accepted as the suggested regulations governing the proper installation of wire and cable in the United States. The code is revised every three years to keep safety in the forefront in wire and cable manufacturing and installation. Even though the code is generally accepted, each state, county, city, and municipality has the option to adopt all of the code, part of the code, or develop one of its own. The local inspectors have final authority of the installation. Therefore, the NEC is a good reference when questions arise about the proper techniques for a particular installation, but local authorities should be contacted for verification.

When choosing cable for an installation, follow these three guidelines to keep problems to a minimum:

1. The application and environment determine which type of cable to use and what rating it should have. Make sure the cable meets the proper ratings for the application.
2. If substituting a cable with another, the cable must be one that is rated the same or higher than what the code calls for. Check with the local inspector as to what is allowed in the local area.
3. The NEC code is a general guideline that can be adopted in whole or in part. Local state, county, city, or municipal approved code is what must be followed. Contact local authorities for verification of the code in the area.

The local inspector or fire marshal has the final authority to approve or disapprove any installation of cable based on the National Electric Code or on the local code.

18.28.5 Plenum Cable

Plenum cable is used in ceilings where the air handling system uses the plenum as the delivery or the return air duct. Because of its flame-resistant and low smoke-emission properties, the special compound used in plenum cable jackets and insulations has been accepted under the provisions of the NEC and classified by Underwriters Laboratories Inc. (UL) for use without conduit in air plenums.

In a typical modern commercial building, cables are installed in the enclosed space between drop ceilings and the floors from which they are suspended. This area is also frequently used as a return air plenum for a building's heating and cooling system. Because these air ducts often run across an entire story unobstructed, they can be an invitation to disaster if fire breaks out. Heat, flames, and smoke can spread rapidly throughout the air duct system and building if the fire is able to feed on combustible materials (such as cable insulations) in the plenum. To eliminate this problem and to keep fumes from entering the air handling system, the NEC requires that conventional cables always be installed in metal conduit when used in plenums.

Plenums, with their draft and openness between different areas, cause fire and smoke to spread, so the 1975 NEC prohibited the use of electrical cables in plenums and ducts unless cables were installed in metal conduit. In 1978, Sections 725-2(b) (signaling cables), 760-4(d) (fire-protection cable), and 800-3(d) (communication/telephone cables) of the NEC allowed that cables “listed as having adequate fire-resistance and low-smoke producing characteristics shall be permitted for ducts, hollow spaces used as ducts, and plenums other than those described in Section 300-22(a).”

While plenum cable costs more than conventional cable, the overall installed cost is dramatically lower because it eliminates the added cost of conduit along with the increased time and labor required to install it.

In 1981 the jacket and insulation compound used in plenum cables was tested and found acceptable under the terms of the NEC and was classified by UL for use without conduit in air return ducts and plenums. Fig. 18-29 shows the UL standard 910 plenum flame test using a modified Steiner tunnel equipped with a special rack to hold test cables.

Virtually any cable can be made in a plenum version. The practical limit is the amount of flammable material in the cable and its ability to pass the Steiner Tunnel Test, shown in Fig. 18-29. Originally plenum cable was all Teflon inside and out. Today most plenum cables have a Teflon core with a special PVC jacket which meets the fire rating. But there are a number of compounds such as Halar® and Solef® that can also be used.

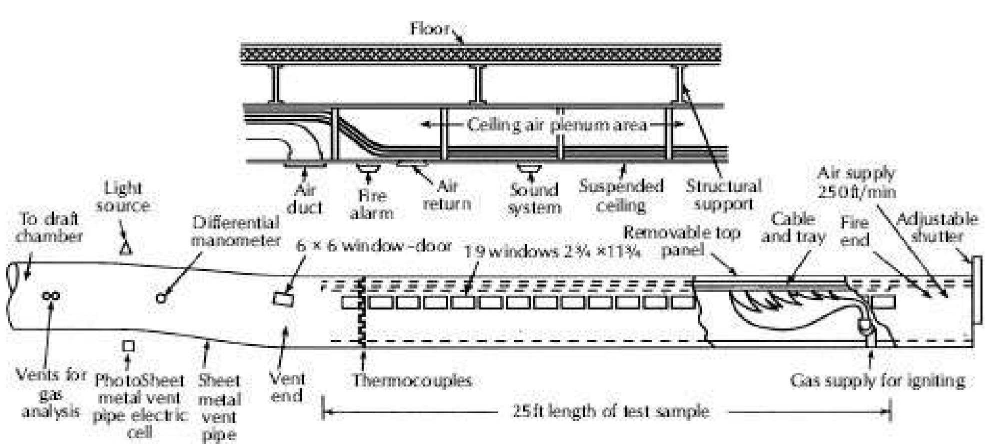


Figure 18-29. Plenum cable flame test, UL standard 910.

18.28.6 AC Power Cords and Receptacles

Ac power cords, like other cables come with a variety of jacket materials for use in various environments. All equipment should be connected with three-wire cords. Never use ground-lift adapters to remove the ground from any equipment. This can be dangerous, even fatal.

The color codes used in North America and Europe for three conductors are shown in Table 18-39. Cables should be approved to a standards shown in Table 18-40.

The UL listing signifies that all elements of the cords and assembly methods have been approved by the Underwriters Laboratories, Inc. as meeting their applicable construction and performance standards. UL listed has become a symbol of safety to millions of Americans and their confidence in it results in easier sales of electrical products.

The U.S. NEMA configurations for various voltage and current general purpose plugs and receptacles are shown in Fig. 18-30.

Table 18-39. Color Codes for Power Supply Cords

Function	North America	CEE and SAA Standard
N—Neutral	White	Light Blue
L—Live	Black	Brown
E—Earth orGround	Green or Green/Yellow	Green/Yellow

Table 18-40. Approved Electrical Standards

Country		Standard
United States	UL	Underwriters Laboratory
Canada	cUL	Canadian Underwriters Laboratory
Germany	GS/TUV	German Product Certification Organization
International	IEC	International Electrotechnical

18.28.7 Shielded Power Cable

Shielding the power cord is an effective means of minimizing the error-generating effects of ambient electrical interference. However, the shield effectiveness of most constructions is mostly medium to high-frequency. For instance, braid shields, even high-density high-coverage braids, have little effectiveness below 1000Hz. So, if the intent is to shield the 50/60Hz from adjacent cables or equipment, buying a shielded power cord will not be effective. In that case, steel conduit is recommended. But even solid steel conduit, perfectly installed, only has an effectiveness of approximately 30dB at 50/60Hz.

The standard power cable shielding consists of aluminum polyester film providing 100% shield coverage and radiation reduction. A spiral-wound drain wire provides termination to ground. These shields are highly effective at high frequencies, generally above 10MHz. Power cords used in applications involving extremely high EMI and RFI environments require shield constructions such as Belden Z-Fold™ foil providing 100% coverage, plus another layer of tinned copper braid of 85% coverage, or greater. This provides the maximum shielding available in a flexible power cord.

Shield effectiveness is an important benefit where interference-sensitive electronic devices, such as computer and laboratory test equipment are concerned. However, any designer or installer should realize that the ultimate protection between power cable and other cables or equipment is distance. The inverse-square-law clearly states that double the distance results in four times less interference. Double that distance is sixteen times less, etc.

18.28.8 International Power

Fig. 18-31 shows the plugs used in various countries and referenced in Table 18-41. Table 18-41 gives the current characteristics and plugs and receptacles for various countries and their cities. The superscripts in the Country or City column are:

1. The neutral wire of the secondary distribution system is grounded.
2. A grounding conductor is required in the electrical cord attached to appliances.
3. Voltage tolerance is plus or minus 4 to 9%.
4. Voltage tolerance is plus or minus 10%.
5. Voltage tolerance is plus or minus 20 to 30%.
6. Voltage tolerance is plus or minus 4.5 to 20.5%.

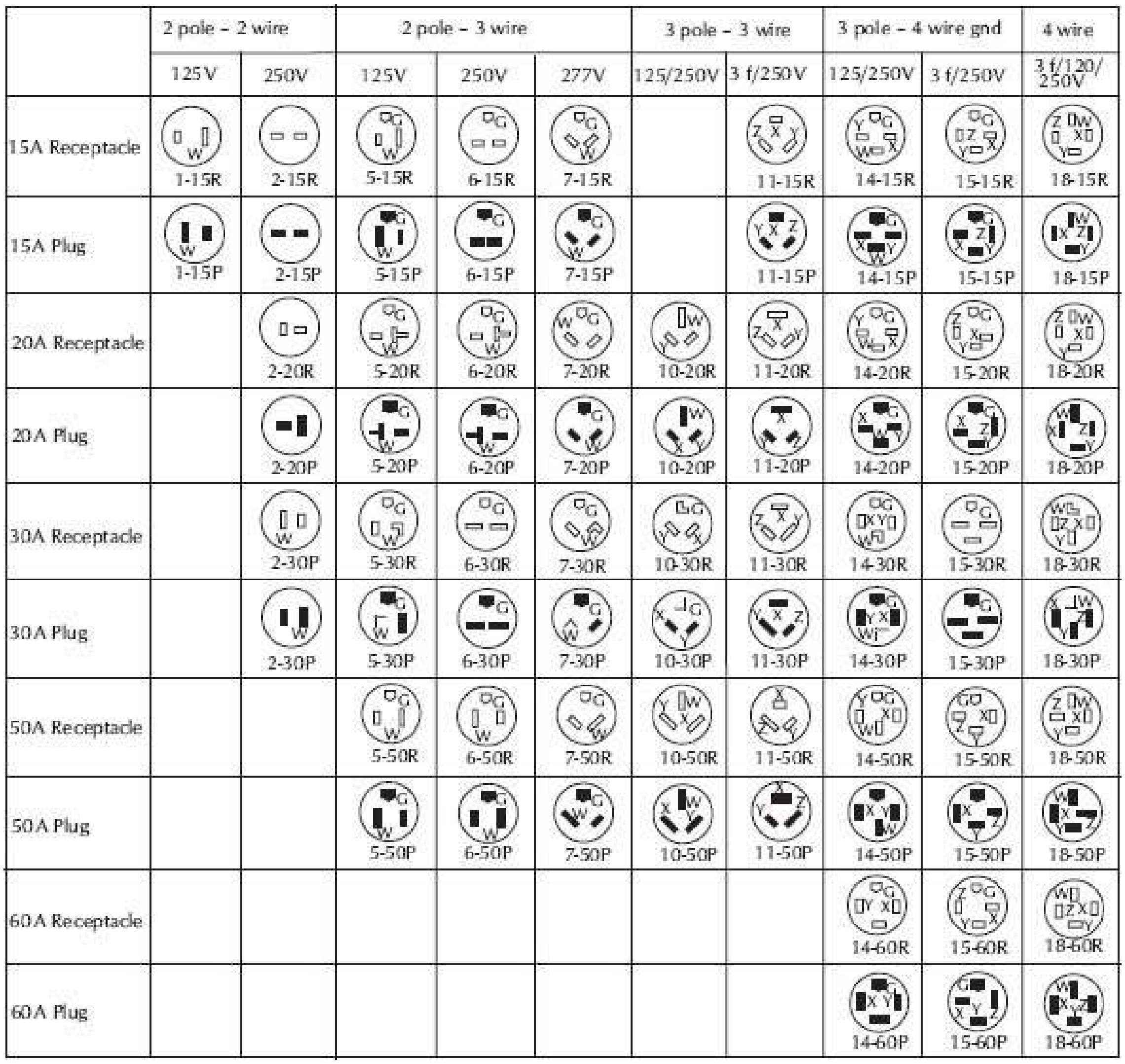


Figure 18-30. NEMA configurations for general-purpose plugs and receptacles.

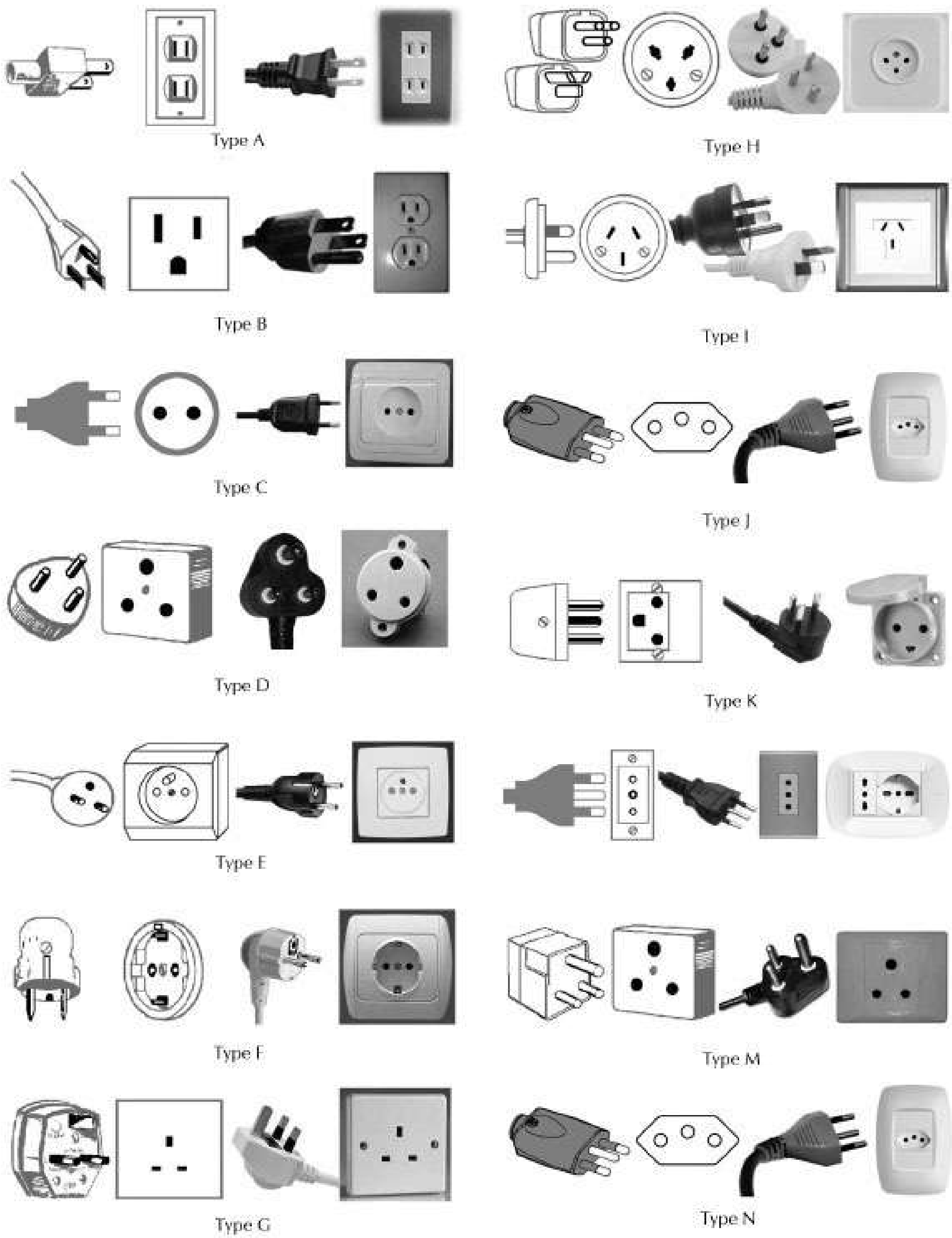


Figure 18-31. International plugs and receptacles.

Table 18-41. Voltage and Current in Various Countries

Country or City	Current	Phase	Voltage	# Wires	Plugs
Afghanistan	a.c. 50	1, 3	220/380	2, 4	D
Albania	a.c. 50	1, 3	220/380	2, 4	C
Algeria	a.c. 50	1, 3	127/220	2, 4	C, F
	a.c. 50		220/380		
Angola1[1,2,3]	a.c. 50	1, 3	220/380	2, 4	C
Argentina	a.c. 50	1, 3	220/380	2, 4	C, I
			220/440	2, 3	
Australia[1,2]	a.c. 50	1, 3	240/415	2, 3, 4	I
Austria[1,2]	a.c. 50	1, 3	220/380	3, 5	C
Bahamas	a.c. 60	1, 3	120/240	2, 3, 4	A, B
			120/208		
Bahrain[1,2]	a.c. 50	1, 3	230/400	2, 3, 4	G
	a.c. 60	1	110/115	3	
Bangladesh[1,2,3]	a.c. 50	1, 3	220/440	3, 4	A, C, D
Barbados[1,2]	a.c. 50	1, 3	115/230	2, 3, 4	A, B, F, H
			115/200		
Belarus	a.c. 50	1, 3	220/380	2, 4	C
Belgium[1,2]	a.c. 50	1, 3	220/400	2, 3, 4	A, C, E
Belize[1,3]	a.c. 60	1, 3	110/220	2, 3, 4	A, B, H
			220/440		
Benin	a.c. 50	1, 3	220/380	2, 4	D
Bermuda[1,2,3]	a.c. 60	1, 3	120/240	2, 3, 4	A, B
			120/208		
Bolivia	a.c. 50	1, 3	110/220	2, 4	A, C
Botswana	a.c. 50	1, 3	231/400	2, 4	C, D, H
Brazil[1]	a.c. 60				A, B, C
Alagoinhas	a.c. 60	1, 3	127/220	2, 3, 4	
Americana	a.c. 60	1, 3	127/220	2, 3, 4	
Anapolis	a.c. 60	1, 3	220/380	2, 3, 4	
Aracaju	a.c. 60	1, 3	127/220	2, 3, 4	
Aracatuba	a.c. 60	1, 3	127/220	2, 3, 4	
Araraquara	a.c. 60	1,	127/220	2, 3, 4	
Bage	a.c. 60	1, 3	220/380	2, 3, 4	
Baixo Guandu	a.c. 60	1, 3	127/220	2, 3, 4	
Barbacena	a.c. 60	1, 3	127/220	2, 3, 4	
Barra Man sa (18-24)	a.c. 60	1, 3	127/220	2, 3, 4	
Barretos	a.c. 60	1, 3	127/220	2, 3, 4	
Bauru	a.c. 60	1, 3	127/220	2, 3, 4	
Belem	a.c. 60	1, 3	127/220	2, 3, 4	
Belo Horizonte	a.c. 60	1, 3	127/220	2, 3, 4	
Blumenau	a.c. 60	1, 3	220/380	2, 3	

Boa Vista (Rio Branco)	a.c. 60	1, 3	127/220	2, 3, 4
Botucatu	a.c. 60	1, 3	127/220	2, 3, 4
Braganca	a.c. 60	1, 3	127/220	2, 3, 4
Brazilia, D.F	a.c. 60	1, 3	220/380	2, 3, 4
Cachoeira	a.c. 60	1, 3	127/220	2, 3, 4
Cachoeira do Itapemirim	a.c. 60	1, 3	127/220	2, 3, 4
Campinas	a.c. 60	1, 3	127/220	2, 3, 4
Campos	a.c. 60	1, 3	127/220	2, 3, 4
Caruaru	a.c. 60	1, 3	220/380	2, 3, 4
Caxias do Sul	a.c. 60	1, 3	220/380	2, 3, 4
Cel Fabriciano	a.c. 60	1, 3	110/220	2, 3
Cidade Industrial (Betim)	a.c. 60	1, 3	127/220	2, 3, 4
Colatina	a.c. 60	1, 3	127/220	2, 3, 4
Corumba	a.c. 60	1, 3	127/220	2, 3, 4
Curitiba	a.c. 60	1, 3	127/220	2, 3, 4
Feira de Santana	a.c. 60	1, 3	127/220	2, 3, 4
Florianopolis	a.c. 60	1, 3	220/380	2, 3, 4
Fortaleza	a.c. 60	1, 3	220/380	2, 3, 4
Franca (Sao Paulo)	a.c. 60	1, 3	127/220	2, 3, 4
Goiania	a.c. 60	1, 3	220/380	2, 3, 4
Goias	a.c. 60	1, 3	220/380	2, 3, 4
Governador Valadares	a.c. 60	1, 3	127/220	2, 3, 4
Ilheus	a.c. 60	1, 3	127/220	2, 3, 4
Itabuana	a.c. 60	1, 3	127/220	2, 3, 4
Itajai	a.c. 60	1, 3	220/380	2, 3
Jequie	a.c. 60	1, 3	220/380	2, 3, 4
Joao Pessoa	a.c. 60	1, 3	220/380	2, 3, 4
Joinville	a.c. 60	1, 3	220/380	2, 3, 4
Juiz de Fora	a.c. 60	1, 3	120/240	2, 3, 4
Jundiai	a.c. 60	1, 3	220	2, 3
Livramento	a.c. 60	1, 3	220/380	2, 3, 4
Londrina	a.c. 60	1, 3	127/220	2, 3, 4
Macapa	a.c. 60	1, 3	127/220	2, 3
Maceio (Alagoas)	a.c. 60	1, 3	220/380	2, 3, 4
Manaus	a.c. 60	1, 3	110/220	2,
Marilia	a.c. 60	1, 3	127/220	2, 3, 4
Mossoro	a.c. 60	1, 3	220/380	2, 3, 4
Natal (Rio Grando Norte)	a.c. 60	1, 3	220/380	2, 3, 4
Niteroi	a.c. 60	1, 3	127/220	2, 3, 4
Nova Friburgo	a.c. 60	1, 3	220/380	2, 3

Olinda	a.c. 60	1, 3	220/380	2, 3, 4	
Ouro Preto	a.c. 60	1, 3	127/220	2, 3, 4	
Paranagua	a.c. 60	1, 3	127/220	2, 3	
Parnaiba	a.c. 60	1, 3	220/380	2, 3	
Paulista	a.c. 60	1, 3	127/220	2, 3, 4	
Pelotas	a.c. 60	1, 3	220/380	2, 3, 4	
Petropolis	a.c. 60	1, 3	127/220	2, 3, 4	
Piracicaba	a.c. 60	1, 3	127/220	2, 3, 4	
Ponta Grossa	a.c. 60	1, 3	127/220	2, 3	
Porto Alegre	a.c. 60	1, 3	127/220	2, 3, 4	
Porto Velho	a.c. 60	1, 3	127/220	2, 3	
Recife	a.c. 60	1, 3	220/380	2, 3, 4	
Ribeirao Preto	a.c. 60	1, 3	127/220	2, 3, 4	
Rio Branco	a.c. 60	1, 3	127/220	2, 3, 4	
Rio de Janeiro	a.c. 60	1, 3	127/220	2, 3, 4	
Salvador	a.c. 60	1, 3	127/220	2, 3, 4	
Santo André	a.c. 60	1, 3	127/220 220/380	2, 3	
Santos	a.c. 60	1, 3	127/220	2, 3, 4	
Sao Bernardo do Campo	a.c. 60	1, 3	220/380	2, 3	
Sao Caetano do Sul	a.c. 60	1, 3	115/230	2, 3	
Sao Luis	a.c. 60	1, 3	110/220	2, 3	
Sao Paulo	a.c. 60	1, 3	115/230	2, 3	
Sorocaba	a.c. 60	1, 3	127/220	2, 3, 4	
Teresina	a.c. 6 0	1, 3	110/220	2, 3	
Uberaba	a.c. 60	1, 3	127/220	2, 3, 4	
Vitoria	a.c. 60	1, 3	127/220	2, 3, 4	
Volta Redonda	a.c. 60	1, 3	125/216	2, 3, 4	
Brunei[1,2]	a.c. 50	1, 3	240/415	2, 4	G
Bulgaria	a.c. 50	1, 3	220/380	2, 4	F
Burkina Faso	a.c. 50	1, 3	220/380	2, 4	B, E
Burma[1,2,3]	a.c. 50	1, 3	230/400	2, 4	C, D, F
Burundi[3]	a.c. 50	1, 3	220/380	2, 4	C, E
Cambodia	a.c. 50	1, 3	220/380	2, 3, 4	
Cameroon	a.c. 50	1, 3	220/380	2, 4	C, E
Canada[1]	a.c. 60	1, 3	120/240	3, 4	B
Cape Verde[2]	a.c. 50	1, 3	220/380	2, 3, 4	C, F
Cayman Islands[1,3]	a.c. 60	1, 3	120/240	2, 3	A, B
Central African Republic[2,3]	a.c. 50	1, 3	220/380	2, 4	C, E
Chad	a.c. 50	1, 3	220/380	2, 4	E
Chile	a.c. 50	1, 3	220/380	2, 3, 4	C, F, L

China, Peoples Republic of	a.c. 50	1, 3	220/380	2, 3, 4	C, D, G, H
Colombia	a.c. 60	1, 3	110/220 150/260	2, 3, 4	A, B
Congo, Democratic Republic of the [1,2]	a.c. 50	1, 3	220/380	2, 3, 4	E
Congo, Peoples Republic of[1,2,3]	a.c. 50	1, 3	220/380	2, 4	C, E
Costa Rica	a.c. 60	1, 3	120/240	2, 3, 4	A, B
Cyprus[1,2]	a.c. 50	1, 3	240/415	2, 4	G
Czech Republic	a.c. 50	1, 3	220/380	2, 3, 4	E
Denmark	a.c. 50	1, 3	220/380	2, 3, 4	C, K
Djibouti, Republic of	a.c. 50	1, 3	220/380	2, 4	C, E
Dominican Republic	a.c. 60	1, 3	110/220	2, 3	A
Ecuador[1]	a.c. 60	1, 3	110/220 127/220	2, 3, 4	A, B, C, D
Egypt	a.c. 50	1, 3	220/380	2, 3, 4	C
El Salvador[1]	a.c. 60	1, 3	115/230	2, 3	A, B, C, D, E, F, G, I, J, L
England (See United Kingdom)					
Eritrea	a.c. 50	1, 3	220/380	2, 4	C
Ethiopia	a.c. 50	1, 3	220/380	2, 4	C
Fiji[3]	a.c. 50	1, 3	240/415	2, 3, 4	I
Finland	a.c. 50	1, 3	230/400	2, 4, 5	C, F
France	a.c. 50	1, 3	220/380	2, 4	E
Gabon[1,2]	a.c. 50	1, 3	220/380	2, 4	D, E
Gambia,The[1, 2]	a.c. 50	1, 3	220/380	2, 4	G
Germany, Federal Republic of [1,2,3]	a.c. 50	1, 3	230/400	2, 4	F
Ghana	a.c. 50	1, 3	240/415	2, 4	D, G
Gibraltar	a.c. 50	1, 3	240/415	2, 4	C, G
Great Britain (See United Kingdom)					
Greece	a.c. 50	1, 3	220/380	2, 4	C, F
Greenland	a.c. 50	1, 3	220/380	2, 3, 4	C, K
Grenada[1,2,3]	a.c. 50	1, 3	230/400	2, 4	G
Guatemala	a.c. 60	1, 3	120/240	2, 3, 4	A, B, G, H, I
Guinea	a.c. 50	1, 3	220/380	2, 3, 4	C, F, K
Guinea-Bissau	a.c. 50	1, 3	220/380	2, 3, 4	C
Guyana[1,2]	a.c. 50	1, 3	110/220	2, 3, 4	A, H
Haiti	a.c. 60	1, 3	110/220	2, 3, 4	A, B, H
Honduras	a.c. 60	1, 3	110/220	2, 3	A
Hong Kong	a.c. 50	1, 3	202/415	3, 4	H
Hungary[2,3]	a.c. 50	1, 3	220/380	2, 3, 4	C, F

Iceland	a.c. 50	1, 3	220/380	2, 3, 4	B
India[3]	a.c. 50	1, 3	230/400	2, 4	C, D, G
Indonesia[1]	a.c. 50				C, E, F
Bandjarmasin	a.c. 50	1, 3	127/220	2, 4	
Bandung	a.c. 50	1, 3	220/380	2, 4	
Bogor	a.c. 50	1, 3	220/380	2, 4	
Cilacap	a.c. 50	1, 3	220/380	2, 4	
Cirebon	a.c. 50	1, 3	220/380	2, 4	
Jakarta	a.c. 50	1, 3	220/380	2, 4	
Malang	a.c. 50	1, 3	220/380	2, 4	
Medan	a.c. 50	1, 3	127/220	2, 4	
Padang	a.c. 50	1, 3	127/220	2, 4	
Palembang	a.c. 50	1, 3	127/220	2, 4	
Semarang	a.c. 50	1, 3	220/380	2, 4	
Sukabumi	a.c. 50	1, 3	220/380	2, 4	
Surabaya	a.c. 50	1, 3	220/380	2, 4	
Surakarta	a.c. 50	1, 3	220/380	2, 4	
Ujungpandang	a.c. 50	1, 3	127/220	2, 4	
Yogyakarta	a.c. 50	1, 3	220/380	2, 4	
Ireland[1,2,3]	a.c. 50	1, 3	220/380	2, 4	G
Israel[1,2,4]	a.c. 50	1, 3	220/380	2, 4	C, H
Italy[1,2,4]	a.c. 50				L
Ancona	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Bari	a.c. 50	1, 3	220/380	2, 4	
Bologna	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Brindisi	a.c. 50	1, 3	220/380	2, 4	
Cagliari	a.c. 50	1, 3	220/380	2, 4	
Catania	a.c. 50	1, 3	220/380	2, 4	
Como	a.c. 50	1, 3	127/220	2, 4	
Cremona	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Florence	a.c. 50	1, 3	220/380	2, 4	
Genoa	a.c. 50	1, 3	127/220	2, 4	
			220/380		
La Spezia	a.c. 50	1, 3	220/380	2, 4	
Latina	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Leghorn	a.c. 50	1, 3	220/380	2, 4	
Milan	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Naples	a.c. 50	1, 3	220/380	2, 4	
Palermo	a.c. 50	1, 3	220/380	2, 4	

Midelt	a.c. 50	1, 3	127/220	2, 4	
Nador	a.c. 50	1, 3	127/220	2, 4	
Ouarzazete	a.c. 50	1, 3	127/220	2, 4	
Oued-Zem	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Ouezzane	a.c. 50	1, 3	127/220	2, 4	
Rabat	a.c. 50	1, 3	127/220	2, 4	
Safi	a.c. 50	1, 3	127/220	2, 4	
Sefrou	a.c. 50	1, 3	127/220	2, 4	
Sidi Kacem	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Sidi Slimane	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Souk-El-Arba Gharb	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Settat	a.c. 50	1, 3	127/220	2, 4	
Taza	a.c. 50	1, 3	127/220	2, 4	
Taroudant	a.c. 50	1, 3	127/220	2, 4	
Tiznit	a.c. 50	1, 3	127/220	2, 4	
Tangier	a.c. 50	1, 3	127/220	2, 4	
Tetouan	a.c. 50	1, 3	127/220	2, 4	
Youssoufia	a.c. 50	1, 3	127/220	2, 4	
Mozambique[2]	a.c. 50	1, 3	220/380	2, 4	C, D, F
Namibia[1,2]	a.c. 50	1, 3	220/380	2, 4	C
Nepal[1]	a.c. 50	1, 3	220/380	2, 4	C, D
Netherlands[1]	a.c. 50	1, 3	220/380	2, 4	F
New Zealand[1,2]	a.c. 50	1, 3	230/400	2, 3, 4	H
Nicaragua	a.c. 60	1, 3	120/240	2, 3, 4	A
Niger	a.c. 50	1, 3	220/380	2, 4	A, C, E
Nigeria[1]	a.c. 50	1, 3	220/380	2, 4	C, D, H
Northern Ireland (See United Kingdom)					
Norway	a.c. 50	1, 3	220/380	2, 4	C, F
Oman[2]	a.c. 50	1, 3	240/415	2, 4	H
Pakistan[1]	a.c. 50	1, 3	230/400	3	B, C, D
Palau	a.c. 60	1, 3	120/240	4	A, B
Panama	a.c. 60	1, 3	120/240	2, 4	A, B, I
Paraguay	a.c. 50	1, 3	220/380	2, 4	C
Peru	a.c. 60	1, 3	220/380	2, 4	A, C
Philippines[1,2]	a.c. 60	1, 3	125/216	2, 4	A, B, C
Poland	a.c. 50	1, 3	220/380	3, 4	C, E
Portugal[1]	a.c. 50	1, 3	220/380	2, 3, 4	C, F
Qatar	a.c. 50	1, 3	240/415	2, 3, 4	D, G
Romania[3]	a.c. 50	1, 3	220/380	2, 4	C, F

Russia	a.c. 50	1, 3	220/380	2, 4	C
Rwanda	a.c. 50	1, 3	220/380	2, 4	C, J
Saudi Arabia[3]	a.c. 60	1, 3	127/220	2, 4	A, B, G
Scotland (See United Kingdom)					
Senegal[1,3]	a.c. 50	1, 3	127/220	2, 3, 4	C, D, E, K
Serbia-Montenegro	a.c. 50	1, 3	220/380	3, 4, 5	F
Seychelles	a.c. 50	1, 3	240/450	2, 4	D
Sierra Leone	a.c. 50	1, 3	230/400	2, 4	D, G
Singapore[1]	a.c. 50	1, 3	230/400	2, 3	B, H
Slovak Republic[1]	a.c. 50	1, 3	220/380	2, 4	E
Somalia	a.c. 50				C
Berbera	a.c. 50	1, 3	230	2, 3	
Brava	a.c. 50	1, 3	220/440	2, 4	
Chisimaio	a.c. 50	1, 3	220	2, 3	
Hargeisa	a.c. 50	1, 3	220	2, 3	
Merca	a.c. 50	1, 3	110/220	2, 4	
Mogadishu	a.c. 50	1, 3	220/380	2, 4	
South Africa[1,2,3]	a.c. 50				D
Alberton	a.c. 50	1, 3	220/380	2, 3, 4	
Beaufort West	a.c. 50	1, 3	230/400	2, 4	
Benoni	a.c. 50	1, 3	230/400	2, 3, 4	
Bethlehem	a.c. 50	1, 3	220/380	2, 4	
Bloemfontein	a.c. 50	1, 3	220/380	2, 4	
Boksburg	a.c. 50	1, 3	230/400	2, 4	
Brakpan	a.c. 50	1, 3	220/380	2, 3, 4	
Caledon	a.c. 50	1, 3	220/380	2, 4	
Cape Town	a.c. 50	1, 3	220/380	2, 4	
Carltonville	a.c. 50	1, 3	220/380	2, 4	
Cradock	a.c. 50	1, 3	230/400	2, 4	
De Aar	a.c. 50	1, 3	220/380	2, 4	
Durban	a.c. 50	1, 3	220/380	2, 4	
East London	a.c. 50	1, 3	220/380	2, 4	
Germiston	a.c. 50	1, 3	230/400	2, 3, 4	
Grahamstad	a.c. 50	1, 3	250/430	2, 4	
Johannesburg	a.c. 50	1, 3	220/380	2, 3, 4	
			230/460	2, 3	
Kimberley	a.c. 50	1, 3	220/380	2, 3, 4	
King Williams	a.c. 50	1, 3	220/380	2, 3, 4	
			250/433		
Klerksdorp	a.c. 50	1, 3	230/400	2, 3, 4	
Kroonstad	a.c. 50	1, 3	230/400	2, 3, 4	
Krugersdorp	a.c. 50	1, 3	220/380	2, 4	
Malmesbury	a.c. 50	1, 3	220/380	2, 4	

Ladysmith, N.	a.c. 50	1, 3	220/380	2, 4	
Oudtshoorn	a.c. 50	1, 3	220/380	2, 4	
Paarl	a.c. 50	1, 3	230/400	2, 4	
Parys	a.c. 50	1, 3	220/380	2, 3, 4	
Pietermaritzburg	a.c. 50	1, 3	220/380	2,	
Port Elizabeth	a.c. 50	1, 3	250/433	2,	
Pretoria	a.c. 50	1, 3	240/415	2, 3, 4	
Queenstown	a.c. 50	1, 3	220/380	2, 4	
Robertson	a.c. 50	1, 3	220/380	2, 4	
Roodepoort	a.c. 50	1, 3	230/400	2, 4	
Rustenburg	a.c. 50	1, 3	220/380	2, 4	
Senekal	a.c. 50	1, 3	220/380	2, 3, 4	
Somerset West	a.c. 50	1, 3	230/400	2, 4	
Springs	a.c. 50	1, 3	220/380 230/400	2,	
Stellenbosch	a.c. 50	3	220/380	4	
Tulbagh	a.c. 50	1, 3	220/380	2, 4	
Uitenhage	a.c. 50	1, 3	220/380	2, 4	
Umtata	a.c. 50	1, 3	230/400	2, 3, 4	
Umkomaas	a.c. 50	1, 3	220/380	2, 4	
Upington	a.c. 50	1, 3	230/400	2, 4	
Vereeniging	a.c. 50	1, 3	220/380	2, 4	
Virginia	a.c. 50	1, 3	230/400	2, 4	
Vryheid	a.c. 50	1, 3	230/400	2, 3, 4	
Walvis Bay	a.c. 50	1, 3	230/400	2, 3, 4	
Welkom	a.c. 50	1, 3	220/380	2, 4	
Wellington	a.c. 50	1, 3	230/400	2, 4	
Worcester	a.c. 50	1, 3	230/400	2, 4	
Spain[1]	a.c. 50	1, 3	220/380	2, 3, 4	C, F
Sri Lanka[1,3]	a.c. 50	1, 3	230/400	2, 4	D
Sudan[1]	a.c. 50	1, 3	240/415	2, 4	C, D
Suriname	a.c. 60	1, 3	127/220	2,	C, F
Swaziland	a.c. 50	1, 3	230/400	2, 4	D
Sweden[1,2]	a.c. 50	1, 3	230/400	2, 3, 4, 5	C, F
Switzerland[1,2]	a.c. 50	1, 3	220/380	2, 3, 4	C, E, J
Syria	a.c. 50	1, 3	220/380	2, 3	C
Tahiti	a.c. 60	1, 3	127/220	2, 3, 4	A
Taiwan[1]	a.c. 60	1, 3	110/220	2, 3, 4	A, B
Tajikistan	a.c. 50	1, 3	220/380	2, 3	C, I
Tanzania[1,2,3]	a.c. 50	1, 3	220/380	2, 4	D, G
Thailand[3]	a.c. 50	1, 3	220/380	2, 4	A, B, C, D, E, G, J, K
Togo	a.c. 50	1, 3	127/220	2, 4	C

			220/380		
Trinidad and Tobago[3]	a.c. 60	1, 3	115/230	2, 3, 4	A, B
			230/400		
Tunisia[1,2,3]	a.c. 50	1, 3	220/380	2, 4	C, E
	a.c. 50		220/380		
Turkey[1]	a.c. 50	1, 3	220/380	2, 3, 4	C, F
Turkmenistan	a.c. 50	1, 3	220/380	2, 3	B, F
Uganda[1,2]	a.c. 50	1, 3	240/415	2, 4	G
Ukraine[1]	a.c. 50	1, 3	220/380	2, 4	C
United Arab Emirates	a.c. 50	1, 3	220/380	2, 4	C, D, G
United Kingdom[1,2,3]					
England	a.c. 50	1 3	230/415	2, 4	A, C, H
Northern Ireland	a.c. 50	1, 3	230/415	2, 4	A, C, H
Scotland	a.c. 50	1, 3	230/415	2, 4	A, C, H
Wales	a.c. 50	1, 3	230/415	2, 4	A, C, H
Uruguay[2,3,6]	a.c. 50	1, 3	220/380	2, 4	C, F, I, L
Uzbekistan	a.c. 50	1, 3	220/380	2, 4	C, I
Venezuela	a.c. 60	1, 3	120/240	2, 3, 4	A, B, H
Vietnam	a.c. 50				A, B, C, E, F
Ban Me Thuot (Sic)	a.c. 50	1, 3	220/380	2, 4	
Can Tho	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Dalat	a.c. 50	1, 3	120/208	2, 4	
			220/380		
Da Nang	a.c. 50	1, 3	127/220	2, 4	
Hanoi	a.c. 50	1, 3	127/220	2, 4	
			220/380		
Hue	a.c. 50	1, 3	127/220	2, 4	
Khanh Hung (Soc Trang)	a.c. 50	1, 3	220/380	2, 4	
Nha Trang	a.c. 50	1, 3	127/220	2, 4	
Saigon	a.c. 50	1, 3	120/208	2, 4	
			220/380		
Western Samoa	a.c. 50.	1, 3	230/400	2, 3, 4	H
Wales (See United Kingdom)					
Yemen, Republic of	a.c. 50	1, 3	220/380	2, 4	A, D, G
Zambia[1,2,4]	a.c. 50	1, 3	220/380	2, 4	C, D, G
Zimbabwe[4]	a.c. 50	1, 3	220/380	2, 3, 4	D, G

Acknowledgments

I wish to acknowledge Martin Van Der Burgt, Product Development Manager; David DeSmidt, Senior Product Engineer; Carl Dole, Applications Engineers; and Steve Lampen, Product Line Manager —Entertainment Products—all with Belden.

Bibliography

"Conduit Capacity Chart," Belden Technical Papers.

"Characteristic Impedance of Cables at High and Low Frequencies," Belden Technical Papers. Marty Van Der Burgt, "Solving Signal Problems," Belden Technical Papers.

Paul Vanderlaan, "Installation Effects Upon Alien Crosstalk and Equal Level Far End Crosstalk," Belden Technical Paper, October 1999.

Paul Vanderlaan, "Impedance/Return Loss," Belden Technical Papers.

R. A. Greiner, "Cables and the Amp," *Audio*, August 1989. Parts copied with permission.

Steve Lampen, *The Audio-Video Cable Installer's Pocket Guide*, New York:McGraw-Hill, 2002.

Steve Lampen, "Cable and the Cutting Edge," Belden Technical Papers, July 1997.

Steve Lampen, "How Wire Fails," Belden Technical Papers, July 1997.

Steve Lampen, "Video and UTP," Belden Technical Papers, May 2000.

Steve Lampen, "What Is RG-6 Anyway," CE Pro, July 2000.

Steve Lampen, *Wire, Cable, and Fiber Optics for Video and Audio Engineers*, New York: McGraw Hill.

Steve Lampen, "Wire, Cable, and Fiber Optics," Belden Technical Papers, June 2000.

"Characteristic Impedance of Cables at High and Low Frequencies," Engineering Department, Belden.

"Choosing the Correct Coaxial Cable for CCTV Applications," Technical Bulletin #*TB3-001*, West Penn Wire.

"Data Grade Fire Alarm Cables," West Penn Technical Wire.

"Differences in Coaxial Cable Construction," West Penn Technical Wire.

"Electric Current Abroad," 1998 Edition, Reprinted 2002, U.S. Department of Commerce.

Chapter 19

Transmission Techniques: Fiber Optics

by Ron Ajemian

19.1 History

Fiber optics is the branch of optical technology concerned with the transmission of light through fibers made of transparent materials, such as glass, fused silica, or plastic, to carry information.

Fiber optics has been used by the telephone industry for over thirty years, and has proved itself as being the transmission medium for communications. Past history shows audio follows the telephone industry, therefore fiber optics will soon be a force in audio.

The founder of fiber optics was probably the British physicist, John Tyndall. In 1870 Tyndall performed an experiment before the Royal Society that showed light could be bent around a corner as it traveled in a rush of pouring water. Tyndall aimed a beam of light through the spout along with the water and his audience saw the light follow a zigzag path inside the curved path of the water. His experiment utilized the principle of total internal reflection, which is also applied in today's optical fibers.

About ten years later, William Wheeler, an engineer from Concord, Massachusetts invented a scheme for piping light through buildings. He used a set of pipes with a reflective lining and diffusing optics to transmit light (bright electric arc) through a building, then diffuse it into other rooms. Although Wheeler's light pipes probably didn't reflect enough light to illuminate the rooms, his idea kept coming up again and again until it finally coalesced into the optical fiber.

At about the same time Alexander Graham Bell invented the photophone, Fig. 19-1. Bell demonstrated that a light ray could carry speech through the air. This was accomplished by a series of mirrors and lenses directing light onto a flat mirror attached to a mouthpiece. Speech vibrating the mirror caused the light to modulate. The receiver included a selenium diode detector whose resistance varied with the intensity of light striking it. Thus the modulated light (sunlight, etc.) striking the selenium detector varied the amount of current through the receiver and reproduced speech that could be transmitted over distances of approximately 200m (650ft).

In 1934, an American, Norman R. French, while working with AT&T, received a patent for his optical telephone system. French's patent described how speech signals could be transmitted via an optical cable network. Cables were to be made out of solid glass rods or a similar material with a low attenuation coefficient at the operating wavelength.

Interest in glass waveguides increased in the 1950s, when research turned to glass rods for unmodulated transmission of images. One result was the invention of the fiber scope, widely used in the medical field for viewing the internal parts of the body. In 1956 Brian O'Brien, Sr., in the United States, and Harry Hopkins and Narinder Kapany, in England, found the way to guide light. The key concept was making a two-layer fiber. One layer was called the core the other layer was called the cladding (see section on light). Kapany then coined the term *fiber optics*.

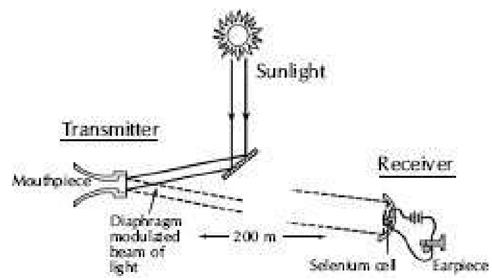


Figure 19-1. Alexander Graham Bell's Photophone.

An efficient light source was needed but it wasn't until 1960 when the first laser light was invented that it became available. A Nobel Prize was awarded to Arthur Schawlow and Charles H. Townes of Bell Laboratories for developing the laser, which was first successfully operated by Theodor H. Maiman of Hughes Research Laboratory. The manufacturing process of lasers from semiconductor material was recognized in 1962. At the same time semiconductor photodiodes were developed for receiver elements. Now the only thing left was to find a suitable transmission medium.

Then in 1966 Charles H. Kao and George A. Hockham, of Standard Telecommunication Labs, England, published a paper proposing that optical fibers could be used as a transmission medium if their losses could be reduced to 20 dB/km. They knew that high losses of over 1000dB/km were the result of impurities in the glass, not of the glass itself. By reducing these impurities a low-loss fiber could be produced for telecommunications.

Finally in 1970, Robert Maurer and associates at Corning Glass Works, New York, developed the first fiber with losses under 20dB/km, and by 1972 lab samples were revealed as low as

4dB/km. Since then the Corning Glass Works and Bell Telephone Labs of the United States; Nippon Sheet Glass Company and Nippon Electric Company of Japan; and at AEG-Tele-funken, Siemens and Halske in Germany, have developed glass fibers with losses at about 0.2dB/km. There is also some plastic materials as well as glass being used for shorter distances.

The practical use of fiber optics for communications began in the mid- and late 1970s with test trials. However, the popularization of fiber optics wasn't until the 1980 Winter Olympics at Lake Placid, New York, when the joint effort of New York Telephone, AT&T, Western Electric, and Bell Labs installed a fiber optic system. Its purpose was to transform the Lake Placid telephone facility into a professional communications center capable of handling a wide range of telecommunications services necessary to support the Olympic events. Today fiber optics is an established technology.

19.2 Advantages of Using Fiber Optics for Audio

There are at least four advantages in using fiber over hardwired systems. One is the superb performance in transmission, allowing extremely large bandwidths and low loss which minimizes the need for preamplifying a signal for long haul applications. Digital data can be easily transmitted with rates of 100Mb/s or higher showing more information handling capability and greater efficiency. Since the optical fiber is nonmetallic (made of glass, plastic, etc.), it is immune to problems caused by electromagnetic interference (EMI) and radio frequency interference (RFI). Also the problem of crosstalk is eliminated—a quality advantage.

With optical fiber one no longer needs to worry about impedance matching, electrical grounding or shorting problems, or no ground loops. Safety is an important feature of fiber optics because a broken cable will not spark, possibly causing shock or an explosion in a dangerous environment.

Another plus is fiber optic cable weighs about 9lbs/1000ft and takes up less space than wire, useful especially when running in conduits. Cost is now less than or comparable to copper. And finally an optical fiber system cannot be easily tapped, which allows for better security.

Applications for Audio

Telephone companies have many fiber links which can connect Asia and Europe to the United States. Think of the many possibilities of doing a multitrack recording from many different places all over the world over a fiber optic cable without worrying about *SNR*, interference, distortion, etc. Top-of-the-line compact disc and DAT players already provide an optical fiber link output. Also, there are companies like Optocore, Klotz Digital of Germany and Wadia Digital Corporation and Emcore of the United States who are manufacturing fiber optic digital audio links, which employ an AES/EBU input and output at each end.

Many recording studios are located in high rise apartment buildings. A perfect application of a digital audio fiber optic link is to connect, for instance, studio A which is located on the 21st floor, to studio B which is located on the 24th floor. This is ideal because the user doesn't have to worry about noise and interference caused by fluorescent lighting and elevator motors, to name a few. Another perfect use is to connect MADI and MIDI stations together.

Another recent advance is a recording studio can record in real time by using DWDM (dense wavelength division multiplexing) lasers and erbium doped optical fibers to send the AES3 audio channels over the Atlantic or Pacific Ocean and then to the appropriate recording studio. The Internet is also being used to establish a fiber optic end-to-end recording session.

19.3 Physics of Light

Before discussing optical fiber, we must understand the physics on light.

Light. Light is electromagnetic energy, as are radio waves, x-rays, television, radar, and electronic digital pulses. The frequencies of light used in fiber optic data transmission are around 200THz–400THz (400×10^{12}), several orders of magnitude higher on the electromagnetic energy spectrum than the highest radio waves, see Fig. 19-2. Wavelength, a more common way of describing light waves, are correspondingly shorter than radio wavelengths. Visible light, with wavelengths from about 400nm for deep violet to 750nm for deep red, is only a small portion of the light spectrum. While fiber optic data transmission sometimes uses visible light in the 600nm to 700nm range, the near infrared region extending from 750nm to 1550nm is of greater interest because fibers propagate the light of these wavelengths more efficiently.

The main distinction between different waves lies in their frequency or wavelength. Frequency, of course, defines the number of sine-wave cycles per second and is expressed in hertz (Hz). Wavelength is the distance between the same points on two consecutive waves (or it is the distance a wave travels in a single cycle). Wavelength and frequency are related. The wavelength (λ) equals

$$\lambda = \frac{v}{f} \quad (19\text{-}1)$$

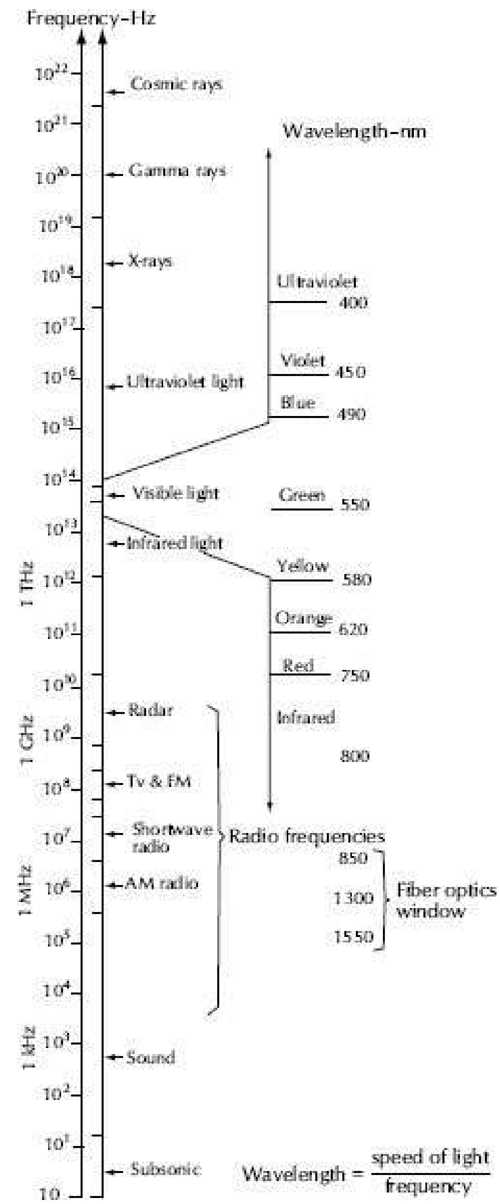


Figure 19-2. The electromagnetic spectrum.

The velocity of electromagnetic energy in free space is generally called *the speed of light*, (186,000mi/s [300,000km/s]). The equation clearly shows that the higher the frequency, the shorter the wavelength.

Light travels slower in other media than a vacuum, and different wavelengths travel at different speeds in the same medium. When light passes from one medium to another, it changes speed, causing a deflection of light called refraction. A prism demonstrates this principle. White light entering a prism is composed of all colors which the prism refracts. Because each wavelength changes speed differently, each is refracted differently, therefore the light emerges from the prism divided into the colors of the visible spectrum, as shown in Fig. 19-3.

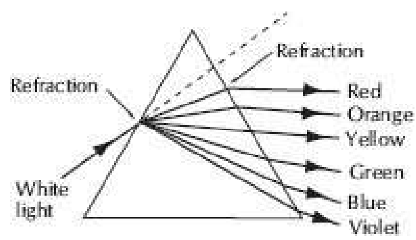


Figure 19-3. Light prism.

The Particle of Light. Light and electrons both exhibit wave- and particle-like traits. Albert Einstein theorized that light could interact with electrons so that the light itself might be considered as bundles of energy or quanta (singular, quantum). This helped explain the photoelectric effect.

In this concept, light rays are considered to be particles that have a zero rest mass called photons.

The energy contained in a photon depends on the frequency of the light and is expressed in Planck's Law, as

$$E = hf \qquad (19\text{-}2)$$

where,
E is the energy in W,
h is Planck's constant, equal to 6.624×10^{-34} J•s,
f is its frequency.

As can be seen from this equation, light energy is directly related to frequency (or wavelength). As the frequency increases, so does the energy, and vice versa. Photon energy is proportional to frequency. Because most of the interest in photon energy is in the part of the spectrum measured in wavelength, a more useful equation which gives energy in electron volts when wavelength is measured in micrometers (μm) is

$$E \text{ in ev} = \frac{1.2406}{\lambda \text{ in } \mu\text{m}} \qquad (19\text{-}3)$$

Treating light as both a wave and a particle aids investigation of fiber optics. We switch back and forth between the two descriptions, depending on our needs.

For example the characteristics of many optical fibers vary with wavelength, so the wave description is used. On the other hand, the emission of a light by a source, a light emitting diode (LED), or its absorption by a positive-intrinsic-negative detector (PIN), is best treated by particle theory.

Light Rays. The easiest way to view light in fiber optics is by using light ray theory, where the light is treated as a simple ray drawn by a line. The direction of propagation is shown on the line by an arrow. The movement of light through the fiber optic system can be analyzed with simple geometry. This approach simplifies the analysis and makes the operation of an optical fiber simple to understand.

Refraction and Reflection. The index of refraction (n) is a dimensionless number expressing the ratio of the velocity of light in free space (c) to its velocity in a specific medium (v)

$$n = \frac{c}{v} \qquad (19\text{-}4)$$

The following are typical indices of refraction:

Vacuum	1.0
Air	1.0003 (generalized to 1)
Water	1.33
Fused Quartz	1.46
Glass	1.5
Diamond	2.0
Gallium Arsenide	3.35
Silicon	3.5
Aluminum Gallium Arsenide	3.6
Germanium	4.0

Although the index of refraction is affected by light wavelength, the influence of wavelength is small enough to be ignored in determining the refractive indices of optical fibers.

Refraction of a ray of light as it passes from one material to another depends on the refractive index of each material. In discussing refraction, three terms are important. The *normal* is an imaginary line perpendicular to the interface of the two materials. The *angle of incidence* is the angle between the incident ray and the normal. The *angle of refraction* is the angle between the normal and the refracted ray.

When light passes from one medium to another that has a higher refractive index, the light is refracted toward the normal as shown in Fig. 19-4A. When the index of the first material is higher than that of the second, most of the light is refracted away from the normal, Fig. 19-4B. A small portion is reflected back into the first material by Fresnel reflection. The greater the difference in the indices of two materials the greater the reflection. The magnitude of the Fresnel reflection at the boundary between any two materials is approximately

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \qquad (19\text{-}5)$$

where,

R is the Fresnel reflection,

n_1 is the index of refraction of material 1,

n_2 is the index of refraction of material 2.

In decibels, this loss of transmitted light is

$$L_F = -10\log(1 - R)\,\text{dB}. \qquad (19\text{-}6)$$

As the angle of incidence increases, the angle of refraction approaches 90° with the normal. The angle of incidence that yields a 90° angle of refraction is called the critical angle, Fig. 19-4C. If the angle of incidence is increased past the critical, the light is totally reflected back into the first material and does not enter the second material and the angle of reflection equals the angle of incidence, Fig. 19-4D.

A single optical fiber is comprised of two concentric layers. The inner layer, the *core*, contains a very pure glass (very clear glass); it has a refractive index higher than the outer layer, or *cladding*, which is made of less pure glass (not so clear glass). Fig. 19-5 shows the arrangement. As a result, light injected into the core and striking the core-to-cladding interface at an angle greater than the critical is reflected back into the core. Since the angles of incidence and reflection are equal, the ray continues zigzagging down the length of the core by total internal reflection, as shown in Fig. 19-6.

The light is trapped in the core, however, the light striking the interface at less than the critical angle passes into the cladding and is lost. The cladding is usually surrounded by a third layer, the buffer, whose purpose is to protect the optical properties of the cladding and core.

Total internal reflection forms the basis for light propagation in optical fiber. Most analyses of light propagation in a fiber evaluate meridional rays—those which pass through the fiber axis each time they are reflected. To help you to understand how an optical fiber works, let us look at Snell's Law which describes the relationship between incident and reflected light as shown in Fig. 19-6.

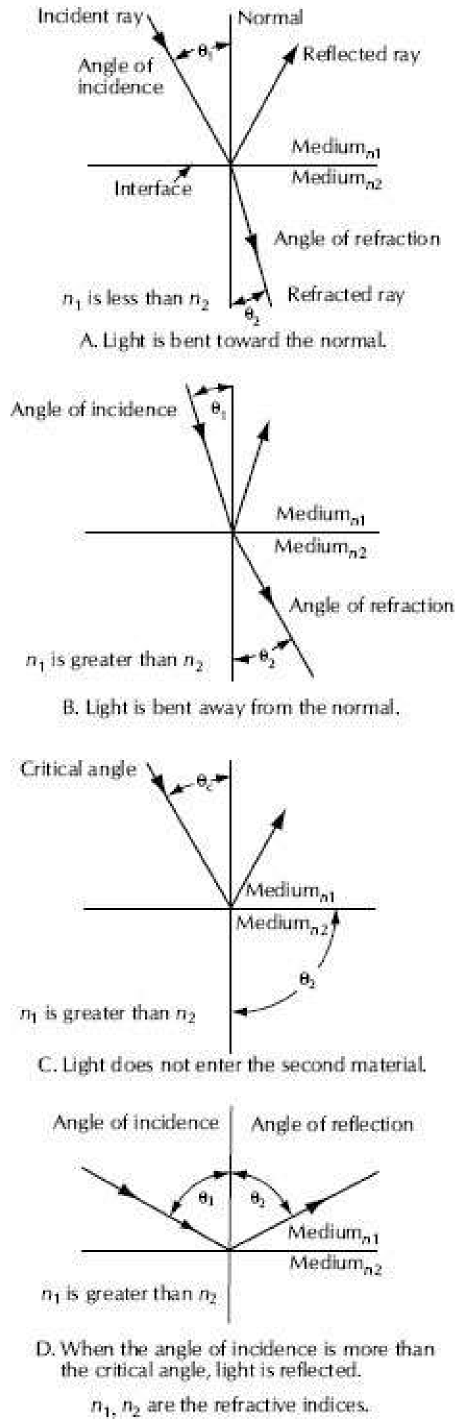


Figure 19-4. Refraction and reflection.

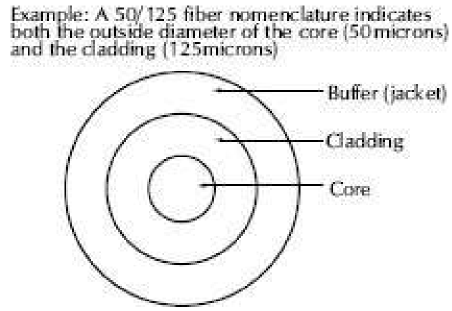


Figure 19-5. Optical fiber cross section.

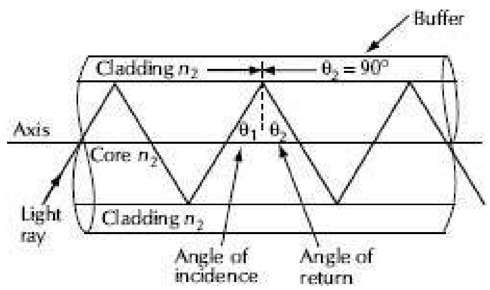


Figure 19-6. Light guided through an optical fiber.

Snell's Law equation is

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{19-7}$$

where,

n_1 is the refractive index of the core,

n_2 is the refractive index of the cladding,

θ_1 is the angle of incidence,

θ_2 is the angle of reflection.

The critical angle of incidence, θ_c, (where $\theta_2 = 90°$) is

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{19-8}$$

At angles greater than θ_c, the light is reflected. Because reflected light means that n_1 and n_2 are equal (since they are in the same material), θ_1 and θ_2, the angles of incidence and reflection are equal. These simple principles of refraction and reflection form the basis of light propagation through an optical fiber.

Fibers also support skew rays, which travel down the core without passing through the fiber axis. In a straight fiber, the patch of a skew ray is typically helical. Because skew rays are very complex to analyze, they are usually not included in practical fiber analysis. The exact characteristics of light propagation depend on the fiber size, construction, and composition, and on the nature of the light source injected.

Fiber performance and light propagation can be reasonably approximated by considering light as rays. However, more exact analysis must deal in field theory and solutions to Maxwell's electromagnetic equations. Maxwell's equations show that light does not travel randomly through a fiber; it is channeled into *modes*, which represent allowed solutions to electromagnetic field equations. In simple terms, a mode is a possible path for a light traveling down a fiber.

The characteristics of the glass fiber, in an extreme sense, can be compared to light as seen through crystal clear water, turbid water, and water containing foreign objects. These conditions are characteristics of water and have quite different effects on light traveling (propagating) through them. The glass fibers are no different, splices, breaks, boundary distortion, bubbles, core out-of-round, etc., all influence the amount of light that reaches the distant end. The main objective is to receive maximum intensity with little or no distortion.

19.4 Fiber Optics

19.4.1 Types of Fiber

Optical fibers are usually classified by their refractive index profiles and their core size. There are three main types of fibers:

1. Singlemode.
2. Multimode stepped index.
3. Multimode graded index.

Singlemode Fiber. Singlemode fiber contains a core diameter of 8–10 microns, depending on the manufacturer. A highly concentrated source such as a laser or high-efficient LED must be used to produce a singlemode for radiation into the fiber. The index of refraction in singlemode fiber is very low because the highly concentrated beam and extremely small core prevent blossoming (officially referred to as scattering) of the ray.

The small core tends to prevent the entry of extraneous modes into the fiber, as illustrated in Fig. 19-7. Loss in a singlemode fiber is very low and permits the economy of longer repeater (telephone amplifier) spacing. This optical fiber has the capability of propagating 1310nm and 1550nm wavelengths. It is well suited for intracity and intercity applications where long repeater spacing is desired.

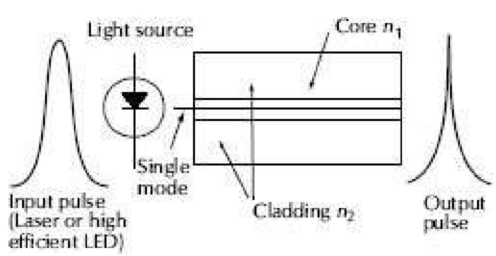


Figure 19-7. Singlemode fiber.

Multimode Step Index Fiber. The production of optical fiber includes layer deposition of core glass inside a started tube. If the glass core layers exhibit the same optical properties the fiber is classed a *step index fiber*. The core layers contain uniform transmission characteristics. The fanout of the rays and their refraction at the core-clad boundary give them the appearance of stepping through the glass, Fig. 19-8. Notice also that as the individual rays step their way through, some travel farther and take longer to reach the far end; the reason for the rounded output pulse shown. This optical fiber requires repeaters-regenerators located at short intervals.

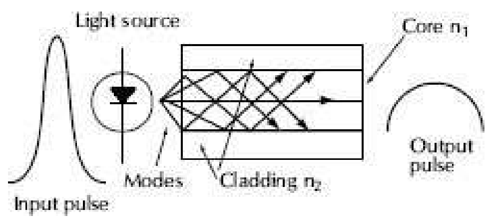


Figure 19-8. Multimode step index fiber.

All rays, or modes, arriving in unison will produce the most exact and strongest replica of the input; this is the objective. For an optical fiber to be most useful in communications, the modes must be channeled through the core in a controlled manner so they all arrive at nearly the same instant.

Multimode Graded Index Fiber. The process of manufacturing graded index fiber involves depositing different grades of glass in the starting tube to provide a core with various transmission characteristics; the outer portion does not impede the passage of modes as much as the center.

In graded index fiber, the core axis contains a higher-density glass of slow wave (ray, mode) propagation in this path for coordination with arrival of the waves in the longest path. The grades of core glass deposited from axis to perimeter are progressively less impeding to let all waves arrive in unison and greatly increase the received intensity (power).

Notice in Fig. 19-9 how each mode is bent (and slowed) in proportion to its entry point in the optical fiber, keeping them in phase. When the rays arrive in phase their powers add. This technique provides maximum signal strength over the greatest distance without regeneration because out-of-phase modes subtract from the total power.

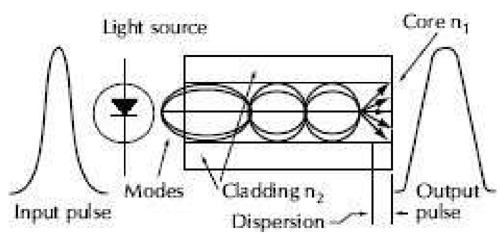


Figure 19-9. Multimode graded index fiber.

19.4.2 Characteristics of Typical Fibers

Table 19-1 gives the characteristics of typical fiber optic cable.

Table 19-1. Characteristics of Typical Cables

Type	Core Dia. (μm)	Clad-ding Dia. (μm)	Buffer Dia. (μm)	NA	Band-width MHz-km	Attenu-ation dB/km
Singlemode	8	125	250		6ps/km*	0.5
at 1300 nm	5	125	250		4ps/km*	0.4
Graded index	50	125	250	0.20	400	3
at 850nm	62.5	125	250	0.275	150	3
	85	125	250	0.26	200	3
	100	140	250	0.30	150	4
Step index	200	380	600	0.27	25	6
at 850 nm	300	440	650	0.27	20	6
PCS†	200	350	—	0.30	20	10
at 790 nm	400	450	—	0.30	15	10
	600	900	—	0.40	20	6
Plastic	—	750	—	0.50	20	400
at 650 nm	—	1000	—	0.50	20	400

* Dispersion per nanometer of source width.

† PCS (Plastic-clad silica: plastic cladding and glass core).

(Courtesy AMP Incorporated)

Dispersion. Dispersion is the spreading of a light pulse as it travels down the length of an optical fiber. Dispersion limits the bandwidth or information-carrying capacity of a fiber. In a digital modulated system, this causes the received pulse to be spread out in time. No power is actually lost due to dispersion, but the peak power is reduced as shown in Fig. 19-10. Dispersion can be canceled to zero in singlemode fibers but with multimode it often imposes the system design limit. The units for dispersion are generally given in ns/km.

Loose Tube and Tight Buffer Fiber Jackets. There are basically two types of fiber jacket protection called *loose tube* and *tight buffer*, Fig. 19-11.

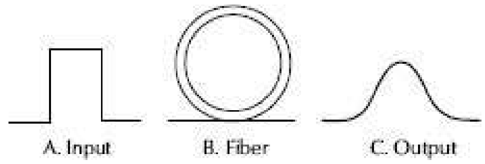


Figure 19-10. Dispersion in an optical fiber.

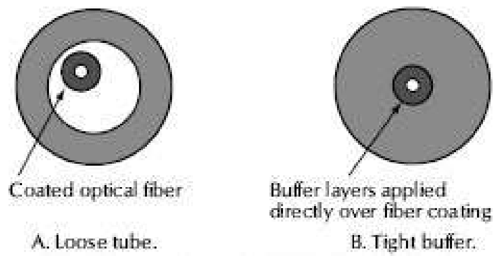


Figure 19-11. Loose tube and tight buffer fiber jackets.

The loose tube is constructed to contain the fiber in a plastic tube that has an inner diameter much larger than the fiber itself. The plastic loose tube is then filled with a gel substance. This allows the fiber to have less stress from the exterior mechanical forces due to the running or pulling of the cable. In multiple fiber loose tube or single fiber loose tube extra strength members are added to keep the fibers free of stress and to help minimize elongation and contraction. Thus, varying the amount of fibers inside the loose tube, the degree of shrinkage can be controlled due to temperature change. This allows for more consistent attenuation over temperature.

The second type, tight buffer, protects the fiber by a direct extrusion of plastic over the basic fiber coating. These tight buffer cables can withstand much greater crush and impact forces without fiber breakage. While the tight buffer has better crush capabilities and is more flexible, it lacks the better attenuation figure of the loose tube due to temperature variations which cause microbending due to sharp bends and twisting of the cable.

Strength members provide for better tensile load parameters similar to coax or electrical audio cables. An optical fiber doesn't stretch very far before it breaks, so the strength members must employ low elongation at the expected tensile loads.

A common strength member used in fiber optic cables for harsh environments is Kevlar™. Kevlar is the material used in bulletproof vests and has the best performance for optical fiber strength members. These strength members are also referred to as tactical optical fiber. They were first used for military communications and were popularized in Operation Desert-Storm in the Iraq and Kuwait war of 1991. These tactical optical fiber cables are impervious to tanks, trucks and bomb explosions. In today's audio applications involving broadcast sports events and news, tactical optical fiber cables have found a niche.

19.4.3 Signal Loss

19.4.3.1 Fiber Optic Transmission Loss (FOTL)

In addition to physical changes to the light pulse which result from frequency or bandwidth limitations, there are also reductions in level of optical power as the light pulse travels to and through the fiber. This optical power loss, or attenuation, is expressed in dB/km (decibels per kilometer). The major causes of optical attenuation in optical fiber systems are:

1. Optical fiber loss.
2. Microbending loss.
3. Connector loss.
4. Splice loss.

5. Coupling loss.

In the ANSI/IEEE Standard 812-1984 the Definition of Terms Relating to Fiber Optics defines attenuation and attenuation coefficient as follows:

Attenuation. In an optical waveguide, attenuation is the diminution of average optical power. Note: In optical waveguides, attenuation results from absorption, scattering, and other radiation. Attenuation is generally expressed in decibels (dB). However, attenuation is often used as a synonym for attenuation coefficient, expressed as dB/km. This assumes the attenuation coefficient is invariant with length. Also see—attenuation coefficient; coupling loss; differential mode attenuation; equilibrium mode distribution; extrinsic joint loss; leaky modes; macrobend loss; material scattering; microbend loss; Rayleigh scattering; spectral window; transmission loss; waveguide scattering.

Attenuation Coefficient. The rate of diminution of average optical power with respect to distance along the waveguide is defined by the equation

$$P(z) = P(0)10^{-\left(\frac{\alpha z}{10}\right)} \qquad (19\text{-}9)$$

where,

P(z) is the power at distance *z* along the guide,

P(0) is the power at *z* = 0,

α is the attenuation coefficient in dB/km if *z* is in km.

From this equation,

$$\alpha z = -10\log\left[\frac{P(z)}{P(0)}\right] \qquad (19\text{-}10)$$

This assumes that a is independent of z; if otherwise, the definition shall be given in terms of incremental attenuation as

$$P(z) = P(0)10^{-\int_0^z \frac{\alpha(z)dx}{10}} \qquad (19\text{-}11)$$

or, equivalently,

$$\alpha z = -10\frac{d}{dz}\log\left[\frac{P(z)}{P(0)}\right] \qquad (19\text{-}12)$$

19.4.3.2 Optical Fiber Loss

Attenuation varies with the wavelength of light. Windows are low-loss regions, where fibers carry light with little attenuation. The first generation of optical fibers operated in the first window, around 820nm to 850nm. The second window is the zero-dispersion region of 1300nm, and the third window is the 1550nm region. A typical 50/125 graded-index fiber offers attenuation of 4dB/km at 850nm and 2.5dB/km at 1300nm, a 30% increase in transmission efficiency. Attenuation is very high in the regions of 730nm, 950nm, 1250nm, and 1380nm; therefore, these regions should be avoided.

Evaluating loss in an optical fiber must be done with respect to the transmitted wavelength. Fig. 19-12 shows a typical attenuation curve for a low-loss multimode fiber. Fig. 19-13 does the same for a singlemode fiber; notice the high loss in the mode-transition region, where the fiber shifts from multimode to singlemode operation. Making the best use of the low-loss properties of the fiber requires that the source emit light in the low-loss regions of the fiber. Plastic fibers are best operated in the visible-light area around 650nm.

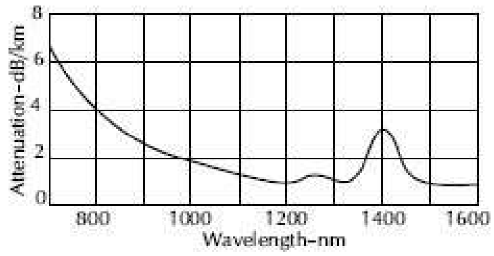


Figure 19-12. Multimode fiber spectral attenuation.

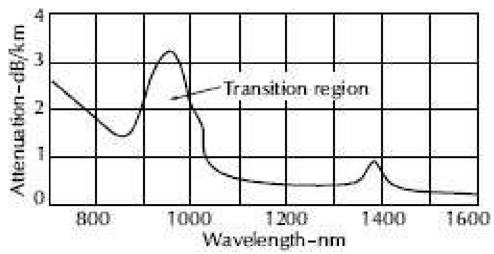


Figure 19-13. Singlemode fiber attenuation.

One important feature of attenuation in an optical fiber is that it is constant at all modulation frequencies within the bandwidth. In a copper cable, attenuation increases with the signal's frequency. The higher the frequency, the greater the attenuation. A 30MHz signal will be attenuated in a copper cable more than a 15MHz signal. As a result, signal frequency limits the distance a signal can be sent before a repeater is needed to regenerate the signal. In an optical fiber, both signals will be attenuated the same.

Attenuation in a fiber has three main causes:

1. Scattering.
2. Absorption.
3. Bending (Microbending).

Scattering. Scattering is the loss of optical energy due to imperfections in the fiber and from the basic structure of the fiber. Scattering does just what the term implies: it scatters the light in all directions. The light is no longer directional.

Rayleigh scattering is the same phenomenon that causes a red sky at sunset. The shorter blue wavelengths are scattered and absorbed while the longer red wavelengths suffer less scattering and reach our eyes, so we see a red sunset. Rayleigh scattering comes from density and compositional variations in a fiber that are natural byproducts of manufacturing. Ideally, pure glass has a perfect molecular structure and, therefore, uniform density throughout. In real glass, the density of the glass is not perfectly uniform. The result is scattering.

Since scattering is inversely proportional to the fourth power of the wavelength $(1/\lambda)^4$, it decreases rapidly at longer wavelengths. Scattering represents the theoretical lower limits of attenuation, which are as follows:

- 2.5dB at 820nm.
- 0.24dB at 1300nm.
- 0.012dB at 1550nm.

Absorption. Absorption is the process by which impurities in the fiber absorb optical energy and dissipate it as a small amount of heat. The light becomes dimmer. The high-loss regions of a fiber result from water bands, (where hydroxyl molecules significantly absorb light). Other impurities causing absorption include ions of iron, copper, cobalt, vanadium, and chromium. To maintain low losses, manufacturers must hold these ions to less than one part per billion. Fortunately, modern manufacturing techniques, including making fibers in a very clean environment, permit control of impurities to the point that absorption is not nearly as significant as it was a few years ago.

Microbend Loss. Microbend loss is that loss resulting from microbends, which are small variations or bumps in the core to cladding interface. As shown in Fig. 19-14, microbends can cause high-order modes to reflect at angles that will not allow further reflection. The light is lost.

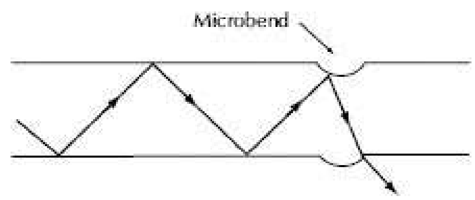


Figure 19-14. Microbend loss.

Microbends can occur during the manufacture of the fiber, or they can be caused by the cable. Manufacturing and cabling techniques have advanced to minimize microbends and their effects.

New Reduced Bend Radius Fibers. Fiber optic cable manufacturers have now significantly reduced the bend radius of the fiber. The reduced bend radius allows for more flexibility allowing installers to bend the fiber around tight corners without any discernible increase in the fiber’s attenuation. There are several names given to these optical fibers such as *bend insensitive* or *bend resistant* that can be somewhat misleading when it comes to the selection of the fiber. The user may tend to believe that the reduction of the bend radius will also eliminate any mishandling, temperature extremes, improper routing, or other external forces on the fiber. However, the user should be aware that these factors may not always be true. Selecting a reduced-bend radius-fiber really achieves the improvements of bending the fibers for tighter bends in fiber panels, frames and routing pathways like conduits, raceways and risers.

There is a common basic rule of thumb that the maximum bend radius should be ten times the outside diameter of the cable or approximately 1.5in, whichever is greater. This reduced bend radius of the fiber decreases the standard by about 50%, or to 15mm, without changing the fiber's attenuation.

There have been fiber demonstrations showing a reduced bend radius fiber patch cord and tying a tight knot within the patch cord. Then the patch cord was tested with the tight knot and revealed that no light escaped and also no increase of attenuation was present. These improvements for patch cords have been tremendous, but when it comes to using reduced bend radius for other applications such as in routing in higher densities or easy connector access they will become more critical. Thus, always consult with the manufacturer's guidelines and specifications when selecting reduced bend radius fibers.

Connector Loss. Connector loss is a function of the physical alignment of one fiber core to another fiber core. Scratches and dirt can also contaminate connector surfaces and severely reduce system performance, but most often the connector loss is due to misalignment or end separation.

Several styles of fiber optic connectors are available from major connector suppliers. Typically, each manufacturer has its own design and is generally not compatible with those of other manufacturers. However, things are constantly changing for the better so now all SMA-and ST-type connectors are compatible.

Depending on connector type, different terminating techniques are used:

- Epoxy and Polish—the fiber is epoxied in place in an alignment sleeve, then polished at the ferrule face.
- Optical and Mechanical—both lenses and rigid alignment tubes are commonly used. In addition, index matching mediums may be employed.

The optical power loss of a connector-to-connector interface typically runs between 0.1dB and 2dB, depending on the style of the connector and the quality of the preparation.

Splice Loss. Two fibers may be joined in a permanent fashion by fusion, welding, chemical bonding, or mechanical joining. A splice loss that is introduced to the system may vary from as little as 0.01dB to 0.5dB.

Coupling Loss. Loss between the fiber and the signal source or signal receiver is a function of both the device and the type of fiber used. For example, LEDs emit light in a broad spectral pattern when compared to laser diodes. Therefore, LEDs will couple more light when a larger core fiber is used, while lasers can be effective with smaller core diameters such as in singlemode systems.

Fiber core size is, therefore, a major factor in determining how much light can be collected by the fiber. Coupled optical power increases as a function of the square of the fiber core diameter.

The *numerical aperture* (NA) is the light gathering ability of a fiber. Only light injected into the fiber at angles greater than the critical angle will be propagated. The material NA relates to the refractive indices of the core and cladding

$$NA = \sqrt{n_1^2 - n_2^2} \qquad (19\text{-}13)$$

where,
NA is a unitless dimension.

We can also define the angles at which rays will be propagated by the fiber. These angles form a cone, called the acceptance cone, that gives the maximum angle of light acceptance. The acceptance cone is related to the *NA*

$$\begin{aligned} \theta &= \sin^{-1}(NA) \\ NA &= \sin\theta \end{aligned} \qquad (19\text{-}14)$$

where,
θ is the half-angle of acceptance, Fig. 19-15.

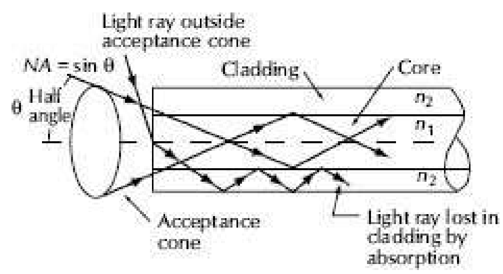


Figure 19-15. Numerical aperture (NA).

The NA of a fiber is important because it gives an indication of how the fiber accepts and propagates light. A fiber with a large NA accepts light well; a fiber with a low NA requires highly directional light.

In general, fibers with a high bandwidth have a lower NA; thus, they allow fewer modes. Fewer modes mean less dispersion and, hence, greater bandwidth. NAs range from about 0.50 for plastic fibers to 0.21 for graded-index fibers. A large NA promotes more modal dispersion, since more paths for the rays are provided.

Sources and detectors also have an NA. The NA of the source defines the angles of the exiting light. The NA of the detector defines the angles of light that will operate the detector. Especially for sources, it is important to match the NA of the source to the NA of the fiber so that all the light emitted by the source is coupled into the fiber and propagated. Mismatches in NA are sources of loss when light is coupled from a lower NA to a higher one.

19.4.3.3 Attenuation Measurement

In an optical fiber, attenuation measurements require comparison of input and output power P_{in} and P_{out}, respectively. It is measured in decibels as

$$L_{FOP} = -10\log\left(\frac{P_{out}}{P_{in}}\right) \text{ in dB} \qquad (19\text{-}15)$$

where,

the negative sign is added to give attenuation a positive value because the output power is always less than the input power for passive devices,

L_{FOP} is the level of fiber optic power expressed in dB.

Remember these are optical powers, and they are dependent on the wavelength. Optical power digital meters make their measurements readings in either dB or dBm, and also display the wavelength. The optical power level L_{OP} is computed with the equation

$$L_{OP} = 10\log\left(\frac{P_s}{P_r}\right) \text{ in dB} \qquad (19\text{-}16)$$

where,

P_s is the power of the signal,

P_r is the reference power.

If the reference power is 1mW, then the equation for the optical power level L_{OP} becomes

$$L_{OP} = 10\log\left(\frac{P_s}{1mW}\right) \text{in dB} \qquad (19\text{-}17)$$

Notice when we know the reference power is 1mW the unit of level changes to dBm. When the reference power is not specified the unit of level is in dB.

Precise fiber attenuation measurements are based on the *cut-back method test* shown in Fig. 19-16. Here a light source is used to put a signal into the optical fiber; a mode filter is used in graded index or multimode fiber to establish a consistent launch condition to allow consistency of measurements. Although modal conditioning (using mode filter) is beyond the scope of this discussion, it is a very important topic for making measurements in multimode fiber because of the effects of modal conditioning on the values one will measure in this test. Measure the amount of light that comes out at the far end, then cut the fiber back about 1 to 2m (3 to 5ft) to just past the mode filter. Measure the amount of light that comes out the new end. The difference in the light at one end and that at the other end divided by the length of the fiber gives you the loss per unit length, or the attenuation of the fiber. This is the method used by all manufacturers for testing their fiber.

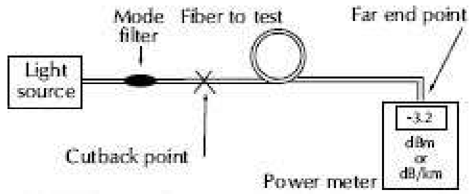


Figure 19-16. The cut-back method for fiber attenuation.

Be aware, however, that this will not accurately measure the loss that light will experience in short multimode fibers because that loss depends on propagation of high-order modes that are eliminated from measurements by adding a mode filter.

Similar measurements can be made on fiber optic cables with mounted connectors by replacing the short cut-back fiber segment with a short jumper cable (including a mode filter if desired). That approach simplifies measurements by avoiding the need to cut fibers at a modest sacrifice in accuracy. One special problem with singlemode fibers is that light can propagate short distances in the cladding, throwing off measurement results by systematically underestimating input coupling losses. To measure true singlemode transmission and coupling, fiber lengths should be at least 20 to 30m (65 to 100ft).

Testing fiber optic continuity is important for system function checks. This test for continuity is simple and doesn't require elaborate equipment. A technician on one end shines a flashlight into the fiber, and the techician on the other end looks to see if any light emerges. That quick test can be checked by measuring cable attenuation. Sites of discontinuities can be located with optical time domain reflectometers (OTDRs), as well as attenuation measurements of the cable.

The OTDR contains a high-power laser source and a sensitive photodetector coupled to a signal amplifier that has a wide dynamic range. The output signal is displayed on an integral oscilloscope. OTDRs use the fundamental reflection or backscatter properties of optical fibers by launching a well-defined optical pulse shape into the fiber and measuring and displaying the return level. However, OTDRs are more elaborate and expensive. The alternative for the audio engineer might be an optical fault finder like the one by Tektronix® (Model TOP300) in Fig. 19-17. The TOP300 is a handheld unit which weighs about one pound and incorporates easy to read LEDs. No experience is necessary for the user, just push the buttons and read the LEDs. The strong laser light shows where the fault is located. There are other test instruments for fiber optics which is beyond the scope of this chapter.

Figure 19-17. An optical fault finder. Courtesy Tektronix®.

19.4.3.4 Advancements in OTDR Testing

The optical time domain reflectometer (OTDR) is designed to troubleshoot fiber breaks and fiber losses. In the past the OTDR was very elaborate and extremely expensive. Fiber optic manufacturers finally have made the OTDR's measurement less complicated. An example of one such device is the OptiFiber® Advanced OTDR by Fluke Networks, Fig. 19-18.

Figure 19-18. OptiFiber Advanced Certifying OTDR. Courtesy of Fluke Networks.

The OptiFiber Advanced OTDR Package will test the fiber link/span, certify it, diagnose it, and document it. This is one of the first certifying OTDRs designed for network owners and installers.

The use of fiber optics in audio and broadcast networks is continually growing, and so are the requirements for testing and certifying. To insure the performance of these optical networks/LANs, network owners are demanding more information that gives them a complete picture of the fiber links. Using this type of OTDR provides a more complete picture.

The OptiFiber is the first test instrument specifically designed to keep network owners and installers on top of the latest requirements for testing and certifying fiber networks. OptiFiber integrates insertion loss and fiber length measurement, OTDR analysis, and fiber connector end face imaging to provide a higher standard of fiber certification and diagnostics. The companion PC software documents, reports, and manages all test data. OptiFiber enables audio network owners of all experience levels to certify fiber to industry and customer specifications, and troubleshoot short-haul connection links and thoroughly document their results.

19.5 Sources

Sources are transmitters of light that can be coupled into fiber optic cables. Basically the two major sources used in fiber optic communications are light emitting diodes (LEDs) and laser diodes. Both are made from semiconductor materials.

LEDs and laser diodes are created from layers of P-and N-type semiconductor materials, forming a junction. Applying a small voltage across the junction causes electrical current, consisting of electrons and holes, to flow. Light photons are emitted from the junction when the electrons and holes combine internal to the junction.

Although the LED provides less power and operates at slower speeds, it is amply suited to applications requiring speeds to several hundred megabits and transmission distances of several kilometers. It is also more reliable, less expensive, has a longer life expectancy, and is easier to use. For higher speeds or longer transmission distances, the laser diode must be considered. Table 19-2 lists the characteristics of typical sources.

Table 19-2. Characteristics of Typical Sources

Type	Output Power (μW)	Peak Wave-length (nm)	Spectral Width (nm)	Rise Time (ns)
LED	250	820	35	12
	700	820	35	6
	1500	820	35	6
LASER	4000	820	4	1
	6000	1300	2	1

Courtesy AMP Incorporated.

19.5.1 LEDs

LEDs are made from a variety of materials located in the Group III to Group V of the Periodic Table of Elements. The color or emission wavelength depends upon the material. Table 19-3 shows some common LED materials used to generate the corresponding colors and wavelengths.

You might have seen LEDs being used in VU or peak-reading meter displays, or as simple status indicators. LEDs used in fiber optics are designed somewhat differently than a simple display LED. The complexities arise from the desire to construct a source having characteristics compatible with the needs of a fiber optic system. Principal among these characteristics are the wavelength and pattern of emission. There are special packaging techniques for LEDs to couple maximum light output into a fiber, Fig. 19-19.

Table 19-3. Materials to Make LEDs and Laser Diodes

Material	Color	Wavelength
Gallium phosphide	green	560nm
Gallium arsenic phosphide	yellow-red	570-700nm
Gallium aluminum arsenide	near-infrared	800-900nm
Indium gallium arsenic phosphide	near-infrared	1300-1500nm

There are three basic types of designs for fiber optic LEDs:

- Surface emitting LED.
- Edge emitting LED.
- Microlensed LED.

Surface Emitting LED. Surface emitting LEDs, Fig. 19-20A, are the easiest and cheapest to manufacture. The result is a low-radiance output whose large emission pattern is not well suited for use with optical fibers. The problem is that only a very small portion of the light emitted can be coupled into the fiber core.

The Burrus LED, named after its inventor Charles A. Burrus of Bell Labs, is a surface-emitting LED with a hole etched to accommodate a light collecting fiber, Fig. 19-21. However, the Burrus LED is not frequently used in modern systems.

Edge Emitting LED. The edge emitting LEDs, Fig. 19-20B, use an active area having stripe geometry. Because the layers above and below the stripe have different refractive indices, carriers are confined by the waveguide effect produced. (The waveguide effect is the same phenomenon that confines and guides the light in the core of an optical fiber.) The width of the emitting area is controlled by etching an opening in the silicon oxide insulating area and depositing metal in the opening. Current through the active area is restricted to the area below the metal film. The result is a high-radiance elliptical output which couples much more light into small fibers than surface emitting LEDs.

Microlensed LED. More recently, technology has advanced such that it is possible, under production conditions, to place a microscopic glass bead that acts as a lens on top of the diode's microchip structure. This microlensed device has the advantage of direct compatibility with a very wide range of possible fibers. There are also double-lensed versions which allow light to be concentrated into the output fiber pigtail.

Dominant Wavelengths. Most LEDs will have a maximum power at a dominant wavelength lying somewhere within the range of 800 to 850nm (first window). Some LEDs are available for other wavelengths: either around 1300nm (second window) or around 1550nm (third window). The choice is dictated by:

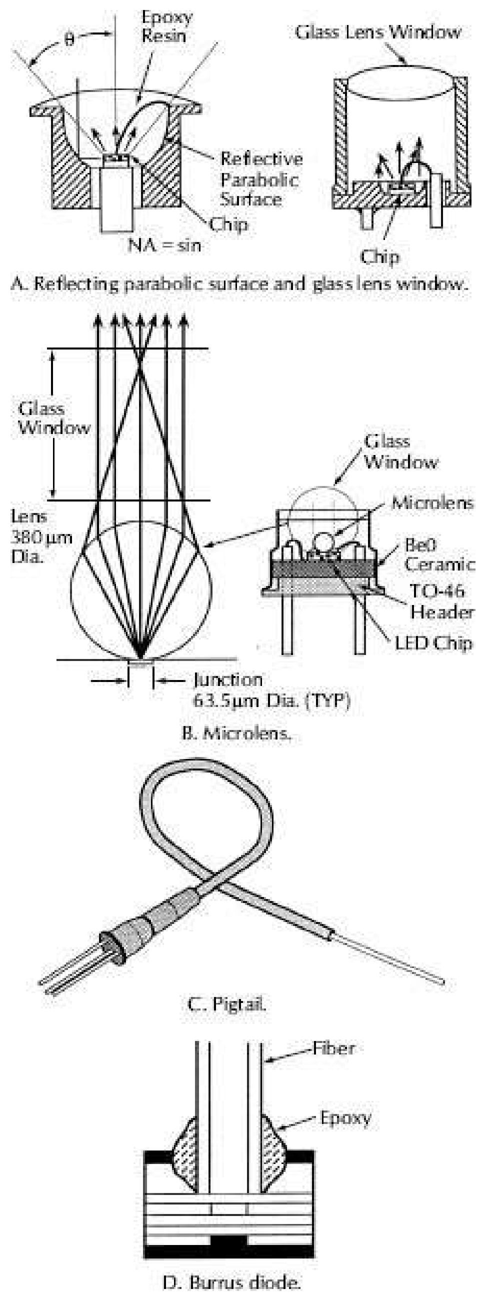


Figure 19-19. Packaging techniques attempt to couple maximum light into a fiber. Courtesy AMP Incorporated.

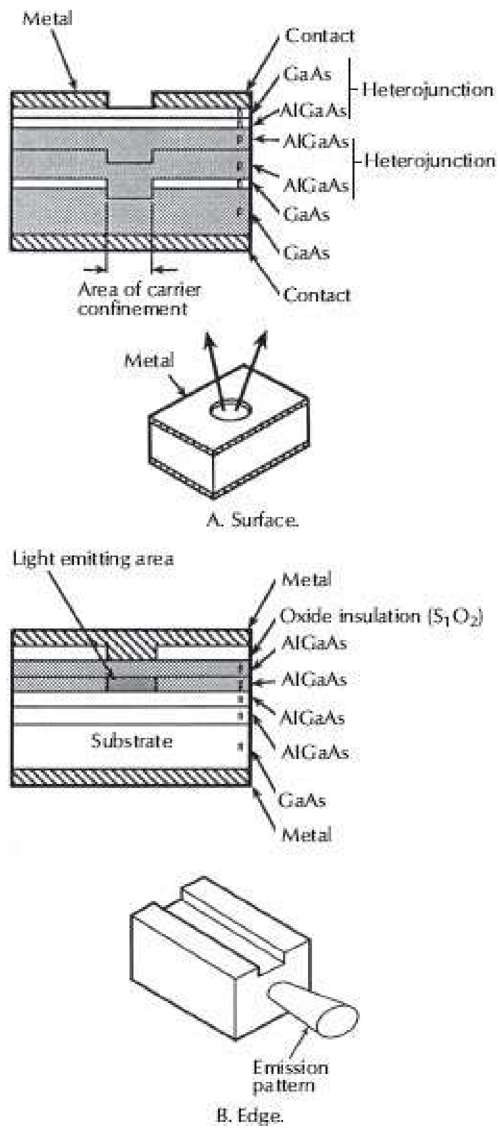


Figure 19-20. Surface and edge emitting LEDs. Courtesy AMP Incorporated.

1. Windows—i.e., loss minima, in optical fibers.
2. Availability of suitable detectors.
3. Cost.
4. Minimization of pulse spreading (dispersion) in a fiber.
5. Reliability.

Also the facility for wavelength-division multiplexing (WDM) can also be a factor influencing choice.

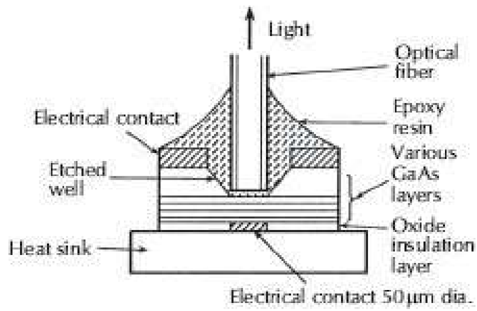


Figure 19-21. The Burrus LED double heterostructure.

19.5.2 Laser Diodes

Laser is an acronym for light amplification by the stimulated emission of radiation. The main difference between an LED and a laser is that the laser has an optical cavity required for lasing, see Fig. 19-22. This cavity, called a Fabry-Perot cavity, is formed by cleaving the opposite end of the chip to form highly parallel, reflective mirrorlike finishes.

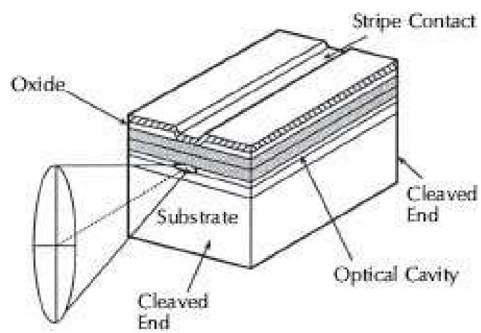


Figure 19-22. Semiconductor laser. Courtesy AMP Incorporated.

At low electrical drive currents, the laser acts like an LED and emits light spontaneously. As the drive current increases, a threshold level is reached, above which lasing action begins. A laser diode relies on high current density (many electrons in the small active area of the chip) to provide lasing action. Some of the photons emitted by the spontaneous action are trapped in the

Fabry-Perot cavity, reflecting back and forth from end mirror to end mirror. These photons have an energy level equal to the band gap of the laser material. If one of these photons influences an excited electron, the electron immediately recombines and gives off a photon. Remember that the wavelength of a photon is a measure of its energy. Since the energy of the stimulated photon is equal to the original stimulating photon, its wavelength is equal to that of the original stimulating photon. The photon created is a clone of the first photon. It has the same wavelength, phase, and direction of travel. In other words, the incident photon has stimulated the emission of another photon. Amplification has occurred, and emitted photons have stimulated further emission.

The high drive current in the chip creates opulation inversion. *Opulation inversion* is the state in which a high percentage of the atoms move from the ground state to the excited state so that a great number of free electrons and holes exist in the active area around the junction. When population inversion is present, a photon is more likely to stimulate emission than be absorbed. Only above the threshold current does population inversion exist at a level sufficient to allow lasing.

Although some of the photons remain trapped in the cavity, reflecting back and forth and stimulating further emissions, others escape through the two cleaved end faces in an intense beam of light. Since light is coupled into the fiber only from the front face, the rear face is often coated with a reflective material to reduce the amount of light emitted. Light from the rear face can also be used to monitor the output from the front face. Such monitoring can be used to adjust the drive current to maintain constant power level on the output.

Thus, the laser differs from an LED in that laser light has the following attributes:

1. Nearly monochromatic: The light emitted has a narrow band of wavelengths. It is nearly monochromatic, that is, of a single wavelength. In contrast to the LED, laser light is not continuous across the band of its special width. Several distinct wavelengths are emitted on either side of the central wavelength.
2. Coherent: The light wavelengths are in phase, rising and falling through the sine-wave cycle at the same time.
3. Highly directional: The light is emitted in a highly directional pattern with little divergence. Divergence is the spreading of a light beam as it travels from a source.

19.5.3 Superluminescent Diodes (SLDs)

A source called the superluminescent diode (SLD) is now available for use. The performance and cost of the SLD fall somewhere in between the LED and the laser. The SLD was first investigated in 1971 by the Soviet physicist, Kurbatov. The SLD may operate like a edge-emitting LED at low currents, while at high-injection currents, the output power increases superlinearly and the spectral width narrows as a result of the onset of optical gain.

19.5.4 Vertical Cavity Surface Emitting Laser (VCSEL)

A more recent source is the vertical cavity surface emitting laser (VCSEL). It is a specialized laser diode that promises to revolutionize fiber optic communications by improving efficiency and increasing data speed. The acronym VCSEL is pronounced vixel. It is typically used for the 850nm and 1300nm windows in fiber optic systems.

19.5.5 LED and Laser Characteristics

19.5.5.1 Output Power

Both LEDs and laser diodes have VI voltage versus current characteristic curves similar to regular silicon diodes. The typical forward voltage drop across LEDs and laser diodes is 1.7V.

In general, the output power of sources decreases in the following order: laser diodes, edge emitting LEDs, surface emitting LEDs. Fig. 19-23 shows some curves of relative output power versus input current for LEDs, SLDs, and laser diodes.

19.5.5.2 Output Pattern

The output or dispersion pattern of light is an important concern in fiber optics. As light leaves the chip, it spreads out. Only a portion of light actually couples into the fiber. A smaller output pattern allows more light to be coupled into the fiber. A good source should have a small emission diameter and a small NA. The emission diameter defines how large the area of emitted light is, and the NA defines at what angles the light is spreading out. If either the emitting diameter or the NA of the source is larger than those of the receiving fiber, some of the optical power will be lost. Fig. 19-24 shows typical emission patterns for the LED, SLD, and laser.

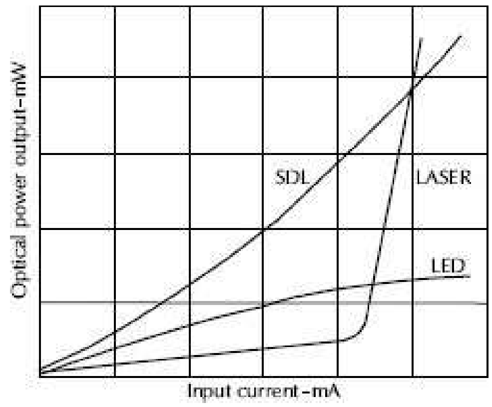


Figure 19-23. Optical power output versus input current for LEDs, SLDs, and laser diodes.

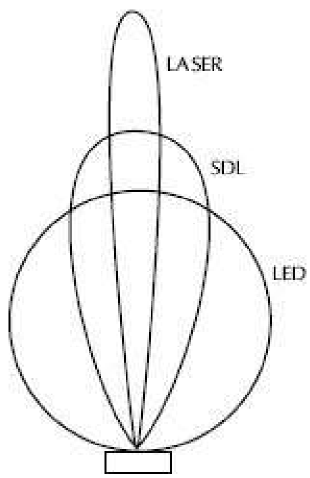


Figure 19-24. Emission patterns of sources.

19.5.5.3 Wavelength

Optical fibers are sensitive to wavelength, therefore the spectral (optical) frequency of a fiber optic source is important. LEDs and laser diodes do not emit a single wavelength; they emit a range of wavelengths. This range is known as the spectral width of the source. It is measured at 50% of the maximum amplitude of the peak wavelength. As an example, if a source has a peak wavelength of 820nm and a spectral width of 30nm, its output ranges from 805nm to 835nm from the spectral width curve specs. The spectral width of a laser diode is about 0.5nm to 6nm; the spectral width of LEDs is much wider—around 20nm to 60nm.

19.5.5.4 Speed

A source must turn on and off fast enough to meet the bandwidth requirements of the system. The source speed is specified by rise and fall times. Lasers have rise times of less than 1ns, whereas LEDs have slower rise times of about 5ns. A rough approximation of bandwidth for a given rise time is

$$BW = \frac{0.35}{t_r} \qquad (19\text{-}18)$$

where,
BW is the bandwidth in Hz,
t_r is the rise time in s.

19.5.5.5 Lifetime

The expected operating lifetime of a source runs into the millions of hours. Over time, however, the output power decreases due to increasing defects in the device's crystal-line structure. The lifetime of the source is normally considered the time where the peak output power is reduced 50% or 3dB. In general LEDs have a longer lifetime than laser diodes. As an example, an LED emitting a peak power of 1mW is considered at the end of its lifetime when its peak power becomes 500μW or 0.5mW.

19.5.5.6 Safety

There are a few main precautions to take in the field of fiber optics.

Most important is to never look directly into an LED or laser diode! Generally, the light emitted by LEDs is not intense enough to cause eye damage, however, it is best to avoid looking at all collimated beams emitted from LEDs or lasers. Be familiar with the sources used. For more safety information, you can contact the Laser Society of America or OSHA.

19.6 Detectors

The detector performs the opposite function from the source: it converts optical energy to electrical energy. The detector can be called an optoelectronic transducer. The most common detectors in fiber optics are PIN photodiodes, avalanche photodiodes (APD), and integrated detectors-preamplifiers (IDP).

The PIN photodiode is the simplest type of detector, and is useful for most applications. It is a three-layer semiconductor device having a layer of undoped (or intrinsic) material sandwiched between a layer of positively doped material and negatively doped material. The acronym PIN comes from this ordering: positive, intrinsic, negative. Light falling on the intrinsic layer causes electron-hole pairs to flow as current. In a perfect photodiode, each photon will set an electron-hole pair flowing. In real PIN photodiodes, the conversion from light to electric current is not perfect; only 60% (or less) of the photons reaching the diode causes current flow.

This ratio is the detector's responsivity. A photodiode has a responsivity of about 0.6A/W; in practical terms, an electrical current of 60μA results for every 100μW of optical energy striking the diode. Responsivity (R) is the ratio of the diode's output current to input optical power and is given in amperes/watt (A/W). The responsivity also depends on the wavelength of light. Being the simplest device, the PIN photodiode offers no amplification of the signal. Even so, it has several virtues: it is inexpensive, easy to use, and has a fast response time.

The avalanche photodiode (APD) provides some gain and is more sensitive to low-power signals than the PIN photodiode. A photon striking the APD will set a number of electron-hole pairs in motion, which in turn sets other pairs in motion, a phenomenon known as the avalanche effect. A photon initiates an avalanche of current. A typical APD has a responsivity of 15μA/μW. An additional advantage of the APD is that it is very fast, turning on and off much faster than a photodiode. The drawback to the APD is its complexity and expense. It requires high voltages for operation and is sensitive to variations in temperature. Like the laser as a source, the APD is only used where speeds and distance require it.

The integrated detector-preamplifier (IDP) is a photodetector and transimpedance amplifier in the same integrated circuit. The advantage is that the signal can be amplified or strengthened immediately, before it meets the noise associated with the load resistor. This is important since any following amplifier stages will boost not only the signal but the noise as well. The IDP amplifies the light induced current and provides a usable voltage output. The responsivity of an IDP is in volts/watt (V/W). The responsivity of a typical IDP is about 15mV/μW. Again, the device has provided gain to overcome noise and provide a suitable *SNR*.

The characteristics of typical detectors are shown in Table 19-4.

Table 19-4. Characteristics of Typical Detectors

Type	Responsivity	Response Time (ns)

PIN Photodiode	0.5μA/μW	5
	0.6μA/μW	1
	0.4μA/μW	1
APD	75.0μA/μW	2
	65.0μA/μW	0.5
IDP	4.5mV/μW	10
	35.0mV/μW	35

Courtesy AMP Incorporated.

19.6.1 Quantum Efficiency (η)

Quantum efficiency, another way of expressing a photo-diode's sensitivity, is the ratio of photons to the number of electrons set flowing in the external circuit and is expressed either as a dimensionless number or as a percentage. The responsivity can be calculated from the quantum efficiency as follows:

$$R = \frac{\eta q \lambda}{hc} \qquad (19\text{-}19)$$

where,

q is the charge of an electron,

h is Planck's constant,

c is the velocity of light.

Since q, c, and h are constants, responsivity is simply a function of quantum efficiency and wavelength.

19.6.2 Noise

Several types of noise are associated with the photodetector and with the receiver. Shot noise and thermal noise are particularly important to our understanding of photodiodes in fiber optics.

The noise current produced by a photodiode is called shot noise. Shot noise arises from the discrete nature of electrons flowing across the *p*-*n* junction of the photodiode. The shot noise can be calculated by using the following equation

$$i_{sn} = \sqrt{2qI_A(BW)} \qquad (19\text{-}20)$$

where,

q is the charge of an electron (1.6×10^{-19} coulomb),

I_a is the average current (including dark current and signal current),

BW is the receiver bandwidth.

Dark current in a photodiode is the thermally generated current. The term *dark* relates to the absence of light when in an operational circuit.

Thermal noise (i_{tn}), sometimes called Johnson or Nyquist noise, is generated from fluctuations in the load resistance of the photodiode. The following equation can be used to calculate the thermal noise

$$i_{tn} = \sqrt{\frac{4KT(BW)}{R_{eq}}} \qquad (19\text{-}21)$$

where,

K is Boltzmann's constant (1.38×10^{-23}joules/K),

T is the absolute temperature in Kelvins,

BW is the receiver's bandwidth,

R_{eq} is equivalent resistance, which can be approximated by a load resistor.

Noise in a PIN photodiode is

$$i_n = \sqrt{i_{sn}^2 + i_{TN}^2} \qquad (19\text{-}22)$$

where,

i_{TN} is the thermal noise.

For an APD, the noise associated with multiplication must also be added.

As a general rule, the optical signal should be twice the noise current to be adequately detected. More optical power may be necessary, however, to obtain the desired *SNR*.

19.6.3 Bandwidth

The bandwidth, or operating range, of a photodiode can be limited by either its rise time or its RC time constant, whichever results in the slower speed or bandwidth. The bandwidth of a circuit limited by the RC time constant is

$$BW = \frac{1}{2\pi R_{eq} C_d} \qquad (19\text{-}23)$$

where,

R_{eq} is the equivalent resistance offered by the sum of the load resistance and diode series resistance,

Cd is the diode capacitance including any contribution from the mounting.

A photodiode's response does not completely follow the exponential response of an RC circuit because changes of light frequency or intensity change the parameters. Nevertheless, considering the device equivalent to a low-pass RC filter yields an approximation of its bandwidth. Fig. 19-25 shows the equivalent circuit model of a PIN photodetector.

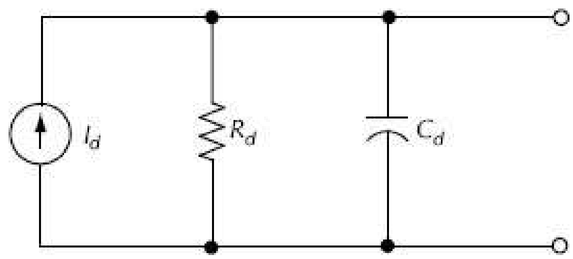


Figure 19-25. Equivalent circuit model of a PIN diode.

19.7 Transmitter/Receiver Modules

In most cases, fiber optic engineers will not design their own transmitters and receivers. They will use completed transmitter-receiver modules. A transmitter module may consist of the following elements:

1. Electronic interface analog/digital input.
2. Analog/digital converter.
3. Drive circuits (preamplifiers, etc.).
4. Optical monitoring circuit.
5. Temperature sensing and control for laser diodes.
6. LED or laser diode as light source.
7. FO connector or pigtail at output.

A receiver module may consist of the following elements:

1. PIN or APD photodiode at the input.

2. Amplification circuits.
3. Signal processor A/D.
4. Analog/digital electrical signal at the output.

Usually, the FO engineer will use a matched pair of transmitter and receiver modules as shown in Fig. 19-26. When considering transmitter/receiver modules one must consider the following requirements:

1. Type of modulation.
2. Bandwidth.
3. Noise.
4. Dynamic range.
5. Electrical and optical interface.
6. Space and cost.

19.8 Transceivers and Repeaters

Transceivers and repeaters are two important components in fiber optics. A transceiver is a transmitter and receiver both in one package to allow transmission and reception from either station. A repeater is receiver driving a transmitter. The repeater is used to boost signals when the transmission distance is so great that the signal will be too highly attenuated before it reaches the receiver. The repeater accepts the signal, amplifies and reshapes it, and sends it on its way by retransmitting the rebuilt signal.

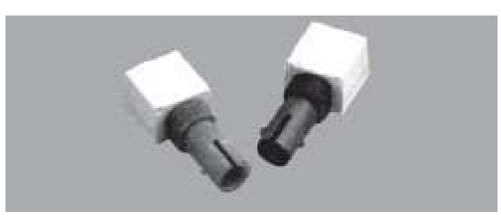


Figure 19-26. A short-wavelength lightwave data link. Courtesy Agilent Technologies.

One advantage of digital transmission is that it uses regenerative repeaters that not only amplify a signal but reshape it to its original form as well. Any pulse distortions from dispersion or other causes are removed. Analog signals use nonregenerative repeaters that amplify the signal, including any noise or distortion. Analog signals cannot be easily reshaped because the repeater does not know what the original signal looked like. For a digital signal, it does know.

Demand on Gigabit Optical Transceivers

The industry is now experiencing that the world needs more bandwidth for today's high-definition technologies. Ethernet has become a standard using both copper and fiber. Manufacturers must keep up with the demand for higher bit rates. The audio/video industry is now employing 1/2/4/10 Gigabit optical transceivers for these high bandwidth applications. Also the need to carry these audio/video signals at distances of 10km or greater has become a reality. Fiber optics can carry a signal with higher bandwidth and greater distance than their copper counterpart.

The price of copper has gone up tenfold due to the world market consumption of copper, especially in the China market. This has bought the price of fiber and fiber optic transceivers down considerably from 2 years ago. An example is the company 3Com, who manufactures fiber optic transceivers. Most of these fiber optic transceivers employ a SFP (small form factor plug-in) duplex type LC connector. Table 19-5 gives 3Com optical transceiver specification data. Fig. 19-27 is a photo of the 3Com Optical

Transceiver.

Table 19-5. 3Com Optical Transceiver—Part No. 3CSFP92

1.25 Gb Gigabit Ethemet/1.063G Fiber Channel	
Application: This 100% 3Com compliant 1000 BASE LX SFP Transceiver is hot-swappable and designed to plug directly into your SFP/GBIC interface slot in your router and switch for Ethernet and Fiber Channel network interface applications.	
Reach	10km (32,820ft)
Fiber Type	SMF (Singlemode Fiber)
Fiber Optic Connector	LC
Center Wavelength λ	1310nm
Min TX Power	−9.5dBm
Max Input Power	−3dBm
RX Sensitivity	−20dBm
Max Input Power	−3dBm
Link Budget	10.50dB
Dimensions	MSA SFP Standard Height: 0.33in (8.5mm) Width: 0.52in (13.4mm) Depth: 2.18in (55.5mm)
Power	3.3V
Operating Temperature	0°C-70°C
Standards:	IEEE 802.3 2003; ANSI X3.297-1997
Compliance	IEC-60825; FDA 21; CFR 1040.10, CFR 1040.11
Warranty	1 Year Full Replacement

Figure 19-27. 3Com Optical Transceiver. Courtesy of 3Com Corporation.

19.9 The Fiber Optic Link

A basic fiber optic link, as shown in Fig. 19-28, consists of an optical transmitter and receiver connected by a length of fiber optic cable in a point-to-point link. The optical transmitter converts an electrical signal voltage into optical power, which is launched into the fiber by either an LED or laser diode.

At the receiving point, either a PIN or APD photodiode captures the lightwave pulses for conversion back into electrical current.

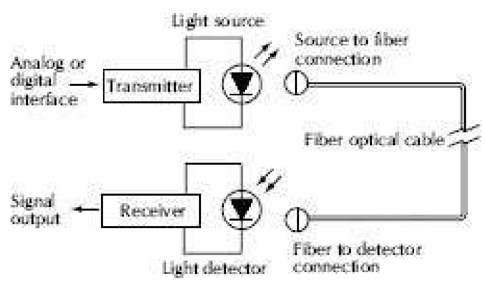


Figure 19-28. Basic fiber optic system.

It is the fiber optic system designer's job to determine the most cost- and signal-efficient means to convey this optical power, knowing the trade-offs and limits of various components. He or she must also design the physical layout of the system.

19.9.1 Fiber Optic Link Design Considerations

Fiber optic link design involves power budget analysis and rise time analysis. The power budget calculates total system losses to ensure that the detector receives sufficient power from the source to maintain the required system *SNR* or bit-error-rate (BER). Rise time analysis ensures that the link meets the bandwidth requirements of the application.

BER is the ratio of correctly transmitted bits to incorrectly transmitted bits. A typical ratio for digital systems is 10^{-9}, which means that one wrong bit is received for every one billion bits transmitted. The BER in a digital system often replaces the *SNR* in an analog system and is a measure of system quality.

19.9.2 Passive Optical Interconnections

In addition to the fiber, the interconnection system includes the means for connecting the fiber to active devices or to other fibers and hardware for efficiently packaging the system to a particular application. The three most important interconnects are FO connectors, splices, and couplers.

Interconnect losses fall into two categories—intrinsic and extrinsic.

- Intrinsic or fiber-related factors are those caused by variations in the fiber itself, such as NA (numerical aperture) mismatch, cladding mismatches, concentricity, and ellipticity, see Fig. 19-29.
- Extrinsic or connector-related factors are those contributed by the connector itself, Fig. 19-30. The four main causes of loss that a connector or splice must control are:

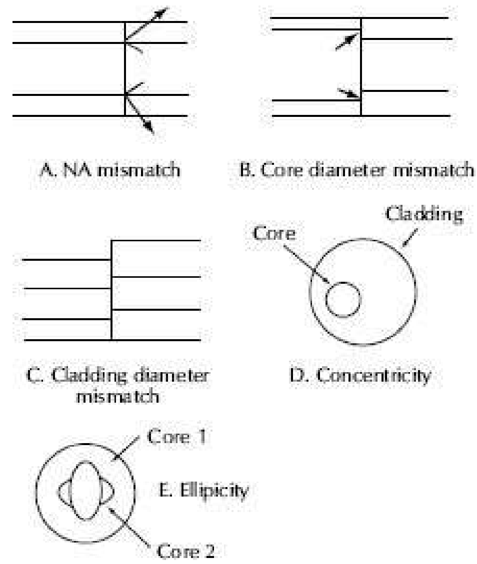


Figure 19-29. Intrinsic fiber optic losses.

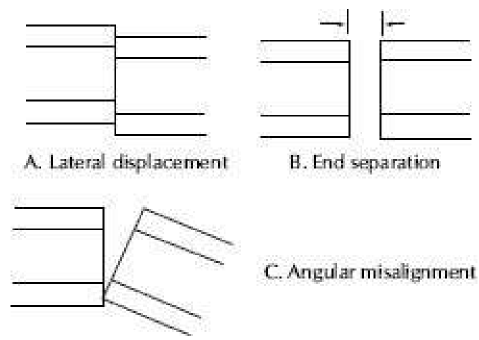


Figure 19-30. Extrinsic fiber optic losses.

1. Lateral displacement.
2. End separation.
3. Angular misalignment.
4. Surface roughness.

19.9.3 Fiber Optic Connectors

A *fiber optic connector* (FOC) is a device designed to simply and easily permit coupling, decoupling, and recoupling of optical signals or power from each optical fiber in a cable to corresponding fibers in another cable, usually without the use of tools. The connector usually consists of two mateable and demateable parts, one attached to each end of a cable or to a piece of equipment, for the purpose of providing connection and disconnection of fiber optic cables. When selecting FOCs one should look for:

1. Minimum insertion loss.
2. Consistent loss characteristics with little change after many connect/disconnect cycles.
3. Easy installation without expensive tools and special training.
4. Reliability of connection (ruggedness).
5. Low cost.

There are many different types of FOCs being used and newer types are emerging rapidly. We cannot even attempt to cover them all, but will discuss the following popular types in wide use in the communications industry, see Fig. 19-31:

1. Biconic.
2. SMA.
3. FC/PC.
4. ST (preferred for audio applications).
5. SC.
6. D4.
7. FDDI (used in audio for duplex operations).
8. Small form factor connectors:
   - LC.
   - MT-RJ.

19.9.3.1 Biconic Connector

The biconic connector was invented by AT&T Bell Laboratories. The latest in precision molding techniques are incorporated to yield fractional dB losses. It employs a conic ferrule and has a precision taper on one end that mates to a free-floating precision molded alignment sleeve within the coupling adaptor. While the biconic is still around, it has lost its popularity for the most part.

19.9.3.2 SMA Connector

The SMA was developed by Amphenol Corporation and is the oldest of all FOCs. It is similar to the SMA connector used for microwave applications. The SMA employs a ceramic ferrule and requires preparation of the fiber end for mounting. There are different versions of the SMA by other manufacturers called FSMA.

19.9.3.3 FC/PC Connector

The FC was developed by NTT Nippon Telegraph and Telephone Corp. It has a flat endface on the ferrule that provides face contact between joining connectors. A modified version of the FC called FC/PC was the first to use physical contact between fiber ends to reduce insertion loss and to increase return loss by minimizing reflections at the interfaces of the joining fibers.

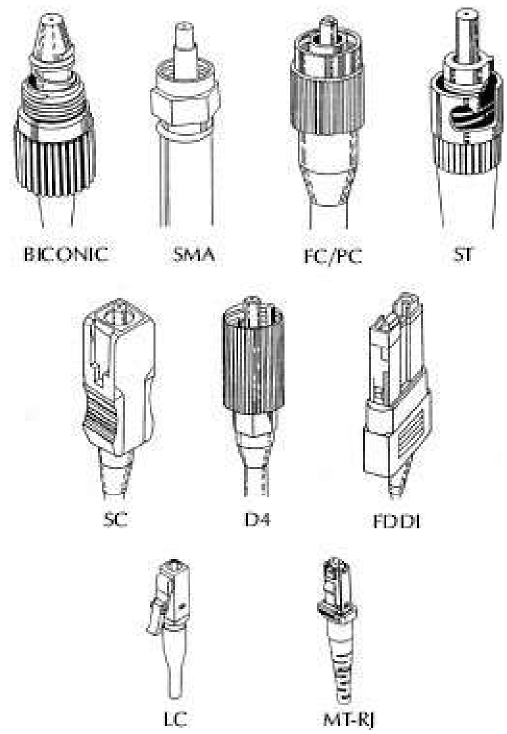


Figure 19-31. Popular fiber optic connectors. (Drawn by Ronald G. Ajemian.)

19.9.3.4 ST Connector

The ST (straight through) connector was introduced in early 1985 by AT&T Bell Laboratories. The ST connector design utilizes a spring-loaded twist and lock coupling, similar to the BNC connectors used with coax cable. The ST prevents the fiber from rotating during multiple connections. This insures more consistent insertion loss during multiple connections. The ST is becoming the most popular FOC at the present time because of performance and compatibility. There are many versions of this ST-type connector being offered by other FOC manufacturers, even some that require no epoxy, just a simple crimp.

19.9.3.5 SC Connector

Most recent on the market is the SC-type connector developed by NTT of Japan. It is a molded plastic connector which employs a rectangular cross section and push-pull instead of threaded coupling. The SC achieves a much lower insertion loss than other types of FOCs and has a greater packing density which is useful in multicable installations. Recently Hirose Electric Co. and Seiko of Japan are manufacturing an SC type that has push-pull locking and employs a zirconia ferrule.

19.9.3.6 D4 Connector

The D4 connector was designed by NEC Nippon Electric Corp., Tokyo, Japan. It is similar to the D3, which was a forerunner of the FC.

19.9.3.7 FDDI Connector

The FDDI (Fiber Data Distributed Interface) connector is another recently developed connector. This connector is described and endorsed by the FDDI standard. The IEEE 802.8 (Institute of Electrical and Electronic Engineers) committee now recommends the FDDI connector for all networks involving duplex fiber operation. However, the increasing gain of the duplex SC connector is making it more popular.

19.9.3.8 SFF (Small Form Factor) Connectors

The SFF connectors are fiber optic connectors newly designed to allow for fast, lower cost and increased density of the patch panel/cross-connect field. They are approximately half the size of the traditional ST and SC connectors.

19.9.3.9 LC Connector

The LC SFF connector by Lucent Technologies was introduced to the market in late 1996. The LC connector employs a 1.25mm ceramic ferrule with a push-pull insertion release mechanism similar to the familiar RJ-45 telephone modular plug. The LC incorporates an antisnag latch that improves durability and reduces cross-connect rearrangement effort. The LC is available in both simplex and duplex types.

19.9.3.10 MT-RJ Connector

The MT-RJ SSF connector was designed by AMP, Inc. (now TYCO) and uses the familiar RJ latching mechanism found in copper systems, but the MT-RJ latch is inherently snag-proof. The single ferrule design of the MT-RJ connector reduces the time and complication of assembly by enabling two-fiber terminations simultaneously.

19.9.3.11 OpticalCon®

Most recent on the market is the OpticalCon® connector developed and introduced in 2006 by Neutrik AG. The OpticalCon® fiber optic connection system consists of a ruggedized all-metal, dust- and dirt-protected chassis and cable connector to increase the reliability. The system is based on a standard optical LC-Duplex connection; however, the OpticalCon® improves this original design to ensure a safe and rugged connection. Due to the compatibility with conventional LC connectors, it offers the choice of utilizing a cost-effective LC connector as a permanent connection, or Neutrik's rugged OpticalCon® cable connector for mobile applications, Fig. 19-32.

Figure 19-32. OpticalCon® connector. Courtesy of Neutrik AG.

19.9.3.12 Toslink

The Toslink connector was developed by Toshiba of Japan in 1983 and is a registered trademark. This connector was originally designed for a plastic optical fiber of 1mm diameter. The actual connector/adapter is of a square construction with newer types having a protective flip cap to close the connector adapter when no plug is mated. Also this connector is referred to as JIS FO5 (JIS C5974-1993 FO5) in a simplex type and JIS FO7 for the duplex version, Fig. 19-33.

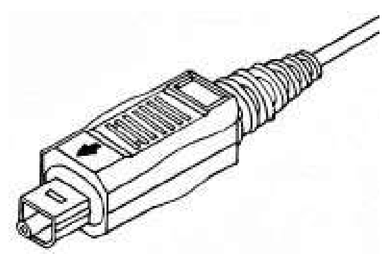


Figure 19-33. Toslink connector.

19.9.3.13 Military Grade Connectors

There are some military-grade types of FOCs that are in use for pro audio that may or may not employ a lens system. These military-grade FOCs go beyond the scope of this chapter.

19.9.4 Connector Installation

The procedure to install a fiber optic connector is similar to that of an electrical connector. However, FOCs require more care, special tools, and a little more time. But as one gains more experience, the time is significantly reduced. The following are steps in making a fiber optic connection:

1. Open the cable.
2. Remove jacketing and buffer layers to expose the fiber.
3. Cut or break the fiber.
4. Insert the fiber into the connector.
5. Attach connector to fiber with epoxy or crimp.
6. Polish or smooth the fiber end.
7. Inspect the fiber ends with a microscope.
8. Seal the connector and fiber cable.

There are presently some FOCs that do not require epoxy or polishing. Things are constantly improving for the better.

19.9.4.1 Current Fiber Optic Connectors for Glass Optical Fibers (GOF)

The LC connector is becoming a de facto standard for pro-audio and video applications. Audio equipment manufacturers are now seeing the benefits of this connector along with its harsh environment type by Neutrik called OpticalCon®. Ongoing work is still in progress for fiber optic cables and connectors by the AES standard group on fiber optic connectors and cables.

19.9.4.2 LC Versus Other Types of Fiber Optic Connectors (ST, SC, etc.)

The LC-type connector over other types of connectors is made for more consistent performance and reliability. The benefits of using the LC type connector are:

1. It is square in shape and keyed, allowing for antirotation, which in turn increases the life expectancy of the connector when mated frequently.
2. It provides for quicker access to patch panel applications where the ST connector (for example) has to be turned to lock.
3. It uses a push-pull insertion release mechanism similar to the familiar RJ-45 telephone plug.
4. It allows tightly spaced patch panels, because it need not be turned to be engaged or disengaged.
5. The LC is called a small form factor (SFF) connector, which is about half the size of the SC connector and provides for high-density patch panels.
6. It offers better axial load and side pull features than the ST connector, thus eliminating disturbances caused by the user touching the cable or boot.
7. Users feel comfortable with LC because of its operational resemblance to an RJ-45 electrical connector.
8. The LC type is universally available throughout the world.
9. It eliminates optical discontinuities resulting from pulling on the cable.
10. It is cost effective.

NOTE: For manufacturers with a large base of existing ST or SC type connectors installed, there are hybrid adapters to mate ST or SC connectors to an LC connector, or vice versa, if needed.

19.9.4.3 Fiber Optic Connector Termination Advancements

Fiber connector manufacturers have now improved the termination process in putting a connector together in a few easy steps. One such device is made by Corning, Inc., called the UniCam® connector system. The UniCam® can be best described as a mini pigtail. It incorporates a factory-installed fiber stub that is fully bonded into the connector's ferrule. The other end is precisely cleaved and placed into the patented alignment mechanism of Corning's mechanical splice. Both the field fiber and fiber stub are fully protected from environmental factors. Unlike other no-epoxy, field-installable connectors, the UniCam® connector requires no polishing, which cuts down the time and cost to install and terminate fiber optic connectors. Now it takes about one minute to terminate a fiber optic LC, ST, or SC, which is much faster than the time to solder an XLR audio connector. Fig. 19-34 shows a UniCam® connector system tool.

Figure 19-34. UniCam connector system. Courtesy of Corning,

Inc.

19.9.4.4 Fiber Optic Splices

A splice is two fibers joined in a permanent fashion by fusion, welding, chemical bonding, or mechanical joining. The three main concerns in a splice are:

1. Splicing loss.
2. Physical durability of the splice.
3. Simplicity of making the splice.

The losses in a fiber optic splice are the same as for FOCs, intrinsic and extrinsic. However the tolerances for a splice are much tighter; therefore, lower attenuation figures are produced.

There are far too many splicing types available to mention; therefore, the following discussion is on splices useful for audio applications. One type by Lucent is called the *CSL LightSplice System*. It provides a fast, easy cleave/mechanical splice for permanent and restoration splicing of singlemode and multimode fibers. The CSL LightSplice System features low loss and reflection and unlimited shelf life, and it does not require polishing or the use of adhesives. The splice, Fig. 19-35, also enables the user to visually verify the splicing process.

Figure 19-35. Lucent CSL LightSplice. Courtesy Lucent Technologies.

Another splice type is Fibrlok™ optical fiber splice by 3M TelComm Products Division. After cable preparation, the two fibers are inserted into a Fibrlok splice. The assembly tool is then used to close the cap, which forces the three locating surfaces against the fibers. This aligns the fibers precisely and permanently clamps them inside the splice. Fibrlok is for both single- or multimode fibers. The splice, Fig. 19-36, can be performed in about 30 seconds after preparing the fiber ends.

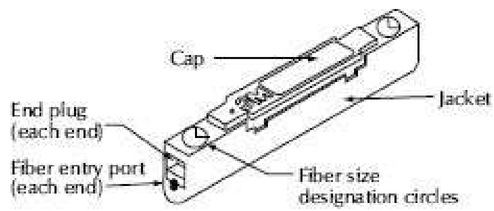


Figure 19-36. Fibrlok optical fiber splice. **19.10 Couplers**

19.10 Couplers

A fiber optic coupler is used to connect three or more fibers together. The coupler is different from a connector or splice which joins only two entities. The fiber optic coupler is more important in fiber optics than for electrical signal transmission because the way optical fibers transmit light makes it difficult to connect more than two points. Fiber optic couplers or splitters were designed to solve that problem.

There are five main concerns when selecting coupler:

1. Type of fiber used (single- or multimode).
2. Number of input or output ports.
3. Sensitivity to direction.
4. Wavelength selectivity.

5. Cost.

There are two types of passive couplers called the *T* and the *Star* couplers, Fig. 19-37. The T coupler has three ports connected to resemble the letter T. The star coupler can employ multiple input and output ports and the number of inputs can be different from the number of outputs.

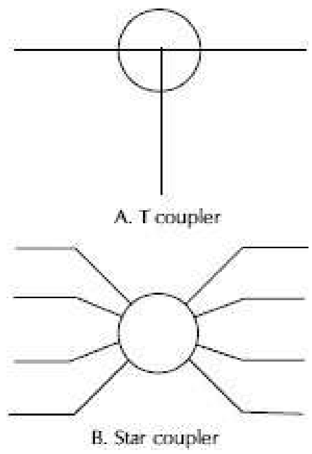


Figure 19-37. T and star couplers.

Couplers are quite simple to use. The following calculation must be made:

Excess loss: The losses that are internal to the coupler from scattering, absorption, reflections, misalignment, and poor isolation. Excess loss is the ratio of the sum of all the output power at the output ports to the input power at the input ports. Usually it is expressed in dB.

Insertion loss: This loss is the ratio of the power appearing at a given output port to that of an input port. Thus, insertion loss varies inversely with the number of terminals.

19.11 Fiber Optic System Design

19.11.1 System Specifications

When designing an FO system, it is often best to order all the component parts from one manufacturer; this way you can be sure of the parts being compatible. Many manufacturers have developed complete systems and have components available for the asking. The following are important things to consider when selecting parts and designing a system:

1. If system is analog:
   - A. Bandwidth in hertz (Hz) or megahertz (MHz).
   - B. Distortion in decibels (dB).
   - C. Operating temperature range in degrees Celsius (°C).
2. If system is digital:
   - A. Required BER. Upper BER is usually in megabits per second (Mbps). Lower BER is usually in bits per second (bps).
   - B. Operating temperature range in degrees Celsius (°C).
3. If system is audio/video:
   - A. Bandwidth in hertz (Hz) or megahertz (MHz).
   - B. Distortion in decibels (dB).
   - C. Crosstalk in decibels (dB) (for multiple channels).
   - D. Operating temperature in degrees Celsius (°C).

19.11.1.1 Transmitter Specifications

1. Input impedance in ohms (Ω).
2. Maximum input signal in dc volts (Vdc), rms or effective volts

(Vrms), peak-to-peak volts (Vp-p).

3. Optical wavelength in micrometers (μm) or nanometers (nm).
4. Optical power output in microwatts (μ,W) or (dBm).
5. Optical output rise time in nanoseconds (ns).
6. Required power supply dc voltage, usually 5 ±0.25Vdc or 15 ±1Vdc.

19.11.1.2 Light Source Specifications

1. Continuous forward current in milliamps (mA).
2. Pulsed forward current in milliamps (mA).
3. Peak emission wavelength in nanometers (nm).
4. Spectral width in nanometers (nm).
5. Peak forward voltage in dc volts (Vdc).
6. Reverse voltage in dc volts (Vdc).
7. Operating temperature range in degrees Celsius (°C).
8. Total optical power output in microwatts (μ,W).
9. Rise/fall times in nanoseconds (ns).

19.11.1.3 Fiber Specifications

1. Mode—single or multimode.
2. Index—step or graded.
3. Attenuation in decibels per kilometer (dB/km).
4. Numerical aperture (NA) (a sine value).
5. Intermodal dispersion in nanoseconds per kilometer (ns/km).
6. Core and cladding diameters in micrometers (μm).
7. Core and cladding index of refraction (a ratio).
8. Bend radius of fiber in centimeters (cm).
9. Tensile strength in pounds per square inch (psi).

19.11.1.4 Cable Specifications

1. Number of fibers (a unit).
2. Core and cladding diameters in micrometers (μm).
3. Cable diameter in millimeters (mm).
4. Minimum bend radius in centimeters (cm).
5. Weight in kilograms per kilometer (kg/km).

19.11.1.5 Detector Specifications

1. Continuous forward current in milliamps (mA).
2. Pulsed forward current in milliamps (mA).
3. Peak reverse voltage in dc volts (Vdc).
4. Capacitance in picofarads (pF).
5. Wavelength in micrometers (μm) or nanometers (nm).
6. Quantum efficiency (η) in percent (%).
7. Responsivity in amps per watt (A/W).
8. Rise/fall time in nanoseconds (ns).
9. Monitor dark current in nanoamperes (nA).
10. Active area diameter in micrometers (μm).
11. Gain coefficient in volts (V) (for APD).
12. Operating temperature in degrees Celsius (°C).

19.11.1.6 Receiver Specifications

1. Output impedance in ohms (Ω).
2. Output signal in dc volts dc (Vdc), rms or effective volts (Vrms), peak-to-peak volts (Vp-p).
3. Optical sensitivity in microwatts (μW), nanowatts (nW), decibels referenced to 1mW (dBm), or megabits per second (Mbps).

4. Optical wavelength for rated sensitivity in nanometers (nm).
5. Maximum optical input power (peak) in microwatts (μW) or (dBm).
6. Analog/digital rise and fall time in nanoseconds (ns).
7. Propagation delay in nanoseconds (ns).
8. Required power supply in dc volts (Vdc).
9. TTL compatibility.
10. Optical dynamic range in decibels (dB).
11. Operating temperature in degrees Celsius (°C).

19.11.2 Design Considerations

Before designing a fiber optic system, certain factors must be realized.

1. What type of signal information is it?
2. Is signal analog or digital?
3. What is the information bandwidth?
4. What power is required?
5. What is the total length of the fiber optic cable?
6. What is the distance between transmitter and receiver?
7. Are there any physical obstacles that the cable must go through?
8. What are the tolerable signal parameters?
9. What is the acceptable *SNR* if system is analog?
10. What is the acceptable BER and rise/fall time if system is digital?

Once these parameters are established, the fiber optic system can be designed.

19.11.3 Design Procedures

The procedures for designing a fiber optic system are as follows:

1. Determine the signal bandwidth.
2. If the system is *analog*, determine the SNR. This is the ratio of output signal voltage to noise voltage, the larger the ratio the better. The *SNR* is expressed in decibels (dB). *SNR* curves are provided on detector data sheets.
3. If the system is *digital*, determine the BER. A typical good BER is 10^{-9}. BER curves are provided on detector data sheets.
4. Determine the link distance between the transmitter and the receiver.
5. Select a fiber based on attenuation.
6. Calculate the fiber bandwidth for the system. This is done by dividing the bandwidth factor in megahertz per kilometer by the link distance. The bandwidth factor is found on fiber data sheets.
7. Determine the power margin. This is the difference between the light source power output and the receiver sensitivity.
8. Determine the total fiber loss by multiplying the fiber loss in dB/km by the length of the link in kilometers (km).
9. Count the number of FO connectors. Multiply the connector loss (provided by manufacturer data) by the number of connectors.
10. Count the number of splices. Multiply the splice loss (provided by manufacturer data) by the number of splices.
11. Allow 1dB for source/detector coupling loss.
12. Allow 3dB for temperature degradation.
13. Allow 3dB for time degradation.
14. Sum the fiber loss, connector loss, splice loss, source/detector coupling loss, temperature degradation loss, time degradation

loss (add values of Steps 8 through 13) to find the total system attenuation.

15. Subtract the total system attenuation from the power margin. If the difference is negative, the light source power receiver sensitivity must be changed to create a larger power margin. A fiber with a lower loss may be chosen or the use of fewer connectors and splices may be an alternative if it is possible to do so without degrading the system.
16. Determine the rise time. To find the total rise time, add the rise time of all critical components, such as the light source, intermodal dispersion, intramodal dispersion, and detector. Square the rise times. Then take the square root of the sum of the total squares and multiply it by a factor of 110%, or 1.1, as in the following equation:

$$\text{System rise time} = 1.1\sqrt{T_1^2 T_2^2 T_3^2 \ldots + T_N^2} \quad (19\text{-}24)$$

19.11.3.1 Fiber Optic System Attenuation

The total attenuation of a fiber optic system is the difference between the power leaving the light source/transmitter and the power entering the detector/receiver. In Fig. 19-38, power entering the fiber is designated as P_S or source power. L_{C1} is the power loss at the source to fiber coupling, usually 1dB per coupling. The power is of that signal launched into the fiber from the light source at the fiber coupling. L_{F1} represents the loss in the fiber between the source and the splice. Fiber optic cable losses are listed in manufacturer's spec sheets and are in dB/km. L_{SP} represents the power loss at the splice. A typical power loss of a splice is 0.3 to 0.5dB. L_{F2} represents the power loss in the second length of fiber.

L_{C2} is the power loss at the fiber to detector coupling. Finally, P_D is the power transmitted into the detector. Other power losses due to temperature and time degradation are generally around 3 dB loss each. Power at the detector is then generalized as

$$P_D = P_S - (L_{C1} + L_{F1} + L_{SP} + L_{F2} + L_{C2}) \qquad (19\text{-}25)$$

Note: All power and losses are expressed in decibels (dB).

19.11.3.2 Additional Losses

If the core of the receiving fiber is smaller than that of the transmitting fiber, loss is introduced. The following equation can be used to determine the coupling loss from fiber to fiber:

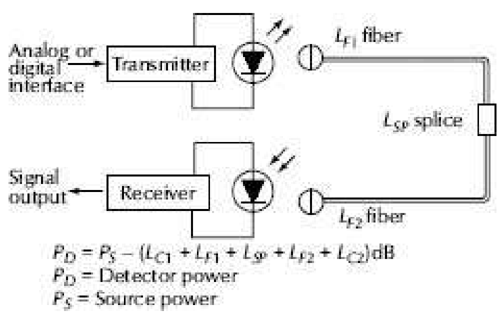


Figure 19-38. Fiber optic system attenuation.

$$L_{dia} = -10\log\left(\frac{dia_r}{dia_t}\right)^2 \qquad (19\text{-}26)$$

where,

L_{dia} equals the loss level of the core's diameter,

dia_r is the diameter of the receiving fiber core in μm,

dia_t is the diameter of the transmitting fiber core in μm.

No diameter mismatch loss occurs when light passes from a smaller core to a larger core.

Differences in NA also contribute loss when the input NA of the receiving fiber is less than that of the output NA of the transmitting fiber:

$$L_{NA} = -10\log\left(\frac{NA_r}{NA_t}\right)^2 \qquad (19\text{-}27)$$

where,

L_{NA} is the loss level of the numerical aperture,

NA_r is the receiving numerical aperture,

NA_t is the transmitting numerical aperture.

Calculation of the NA loss requires that the output NA of the transmitting fiber be known. Since the actual output NA varies with source, fiber length, and modal patterns, using the material NA yields misleading results. No NA mismatch loss occurs when the receiving fiber has an NA greater than that of the transmitting fiber.

The loss of optical power from mismatches in NA and diameter between the source and the core of multimode fiber is as follows:

- When the diameter of the source is greater than the core diameter of the fiber, the mismatch loss is

$$L_{dia} = -10\log\left(\frac{dia_{fiber}}{dia_{source}}\right)^2 \text{ in dB} \qquad (19\text{-}28)$$

where,

L_{dia} is the level of core diameter mismatch loss.

- No loss occurs when the core diameter of the fiber is larger. When the NA of the source is larger than the NA of the fiber, the mismatch loss is

$$L_{NA} = -10\log\left(\frac{NA_{fiber}}{NA_{source}}\right)^2 \text{ in dB} \qquad (19\text{-}29)$$

where,

L_{NA} is the numerical aperture mismatch loss.

- No loss occurs when the fiber NA is the larger. Area or diameter loss occurs when a source's area or diameter of emitted light is larger than the core of the fiber. (Area is often used instead of diameter because of the elliptical beam pattern of edge emitters and lasers.) Area or diameter loss is equal to

$$L_{area} = -10\log\left(\frac{area_{fiber}}{area_{source}}\right) \text{ in dB} \qquad (19\text{-}30)$$

where,

L_{area} is the loss level of the area.

Data sheets for sources often give the area and NA of the output. Although some may not, they may be calculated from information such as polar graphs that are often provided. Calculation of the NA loss and area loss yields an estimate of loss resulting from optical differences between source and fiber. Additional interconnection loss comes from connector related loss, which includes Fresnel reflections and misalignment contributed by a connector.

As with sources, two main causes of loss in coupling light from a fiber into the detector results from mismatches in diameter and NA. When $diad_{et} < diaf_{iher}$, then

$$L_{dia} = -10\log\left(\frac{dia_{det}}{dia_{fiber}}\right)^2 \text{ in dB} \qquad (19\text{-}31)$$

When $NA_{det} < NA_{fiber}$, then (19-32)

$$L_{NA} = -10\log\left(\frac{NA_{det}}{NA_{fiber}}\right)^2 \text{ in dB} \qquad (19\text{-}33)$$

where,

L_{dia} is the loss level of the diameter,

L_{NA} is the loss level of the numerical aperture.

Since detectors can be easily manufactured with large active diameters and wide angles of view, such mismatches are less common than with sources. Other losses occur from Fresnel reflections and mechanical misalignment between the connector and the diode package.

19.12 To Multimode or Singlemode, that is the Question?

First, let's review the two basic fiber optic cable types multimode and singlemode.

Multimode graded index fibers have a larger core diameter than singlemode fiber. The two core diameters are 62.5µm and 50µm which have been pretty well standardized for glass optical fibers. The 62.5µm optical fiber was popularized in the late 1980s by the telecommunication companies. These multimode optical fibers have been typically suited for LANs with moderate distances up to about 2km.

The 50µm multimode core diameter was first used in the 1970s for short haul applications. Later the larger 62.5µm core diameter was being used to allow the light from lost cost LEDs to be injected into this fiber core. As distance and bandwidth increased a breakthrough of the introduction of VCSELs (Vertical Cavity Surface Emitting Lasers) allowed the smaller 50µm core diameter to regain its popularity for better efficiency than the 62.5µm. Therefore, the 62.5µm core diameter is being phased out as time goes on.

Singlemode optical fibers have a much smaller core diameter of about 8µm to 10µm. This allows the light transmitted in the core to carry only a singlemode thus increasing the distance to 50km and greater bandwidth.

Over the years of about 2009 to 2013 we have seen new standards emerged. A new laser optimized multimode and singlemode fiber standard has risen. These optimized multimode and singlemode are designated with the nomenclature of OMx and OSx for Optimized Multimode and Optimized Singlemode respectively. The × designation is a number after the OM and OS to represent the grade of optimization. The higher the number the better the optimization. For instance OM4 is better than OM3, etc. Table 19-6 gives the generic OMx and OSx designation.

The OSx is now also being standardized in the same way as the OMx types. This now allows the designer or user to mix and match for optimized performance using these performance standards. Before these standards came to be there was no consistency with the types of optical fiber chosen. Older systems used the OM1 and OM2 types which now have become more incompatible with today's systems.

Note: OMx and OSx Fiber Types are optimized for Laser/LED sources at optimized wavelengths in nanometers. Please always consult the optical fiber manufacturer's data when using these different optical fiber types. In general, the higher the OMx/OSx numerical number, the better the system performance shall be from that optical fiber type. Therefore, OM4 is better than OM3, etc.

Table 19-6. Industry Standard Designations for MMF and SMF

Core Diameter	**IEC-60793-2-10**	**OM Fiber Type ISO/IEC 11801**	**Common Fiber Type**
62.5mm	Type A1b	OM1	MMF
50 mm	Type A1a.1	OM2	MMF
50 mm	Type A1a.2	OM3	MMF
50 mm	Type A1a.3	OM4	MMF
Core Diameter	**IEC-60793-2-50**	**OM Fiber Type ISO/IEC 11801**	**Common Fiber Type**
8 – 10mm	Type B1.1	OS1	SMF
8 – 10mm	Type B1.3	OS2	SMF

19.13 Fiber Optic Deployment for Future Audio/Video Networks

As network designers and engineers move towards higher data rates of 40-100Gbps there are more choices now available. We are no longer seeing one or two optical fibers in a system. Systems used in multichannel audio/video now deploy as many as 4 to 24 ribbon fibers. To accommodate these ribbon fibers a new fiber optic connector has emerged.

The MPO/MTP was introduced by NTT of Japan and USConec and its acronym stands for "multi-fiber push on/pull off". The MTP connector is a high performance MPO connector with multiple engineered product enhancements to improve optical and mechanical performance when compared to generic MPO connectors. The MTP connector, Fig. 19-39, has a removable housing which allows the user to do the following:

- Re-work and re-polish the MT ferrule
- Change the gender after assembly or even in the field
- Scan the ferrule interferometrically after assembly
- Eliminates lost pins
- Centers spring force
- Eliminates fiber damage from spring

The MTO/MTP is available in both MMF and SMF with a 2.5 × 6.4mm ferrule. These connectors are optioned to be made with 4, 8, 12, 24, 48 and 72 multiple fiber counts.

Figure 19-39. MPT connector.

Neutrik AG, developed the opticalCON MPT® cable connector that accommodates 12 optical fiber channels (PC or APC) based on conventional and proven MTP® connectivity protected by a ruggedized and durable all-metal housing, Fig. 19-40. It features a spring loaded push-pull locking mechanism and an excellent cable retention utilizing aramid yarn. The optical connection is exceptional and well protected against dirt and dust by an automatically operated sealing cover. The opticalCON MPT® has been very popular for large/small broadcast and live sports installations.

The cable connector comes pre-assembled and is not available as a single component. The cable is available in different lengths and is packed either in a case, on drum or airspool.

Features & Benefits

- Offers a cut-proof and rodent-resistant double-jacket, glass-yarn armored cable construction.
- Rugged 12-channel fiber optic connection system.
- For Point-to-Point multichannel routing.
- Innovative spherical shutter guarantees low maintenance.
- Dust and water resistant according to IP65 in mated condition.

Figure 19-40. opticalCON MPT® cable connector. Courtesy Neutrik, AG.

19.14 Fiber Optic Considerations

The professional audio engineer, technician or personnel is now facing many new challenges of distributing audio signals. The use of fiber optics is becoming easier, more efficient and cost effective over its copper counterpart. The many breakthroughs in fiber optic technology are leading the way into the future. Glass optical fiber cables are more robust and cost effective enough to use for longer runs exceeding 2km. Plastic fibers (POFs) are very good at shorter distances (25ft or less), but they do not meet the fire codes for most building structures. Jitter still seems to be problematic with POFs even at 15ft in some cases. However, plastic fiber is improving by mixing combinations of glass and plastic which is referred to as plastic-clad silica (PCS) (plastic cladding and glass core). These PCS are being used in industrial applications as well as some telecommunication areas. There are many types of fiber optic system link designs. Usually, the designer is far better off designing and buying components from one or two vendors, which took the guess work out of system compatibility. The advancements of tools for connecting and splicing optical fibers has now become simple and time efficient enough to easily integrate in any audio system. As bandwidths keep increasing, and greater integration between audio and video, the only thing that will keep up with it is fiber optics. The integrity of the audio/video signals will not be altered, while keeping the quality at high levels. We are now experiencing fiber to the home FTTH and, to coin a phrase, fiber to the studio/set FTTS. The audio community is seeing many technological breakthroughs and these fiber optic cables, connectors, and opto-chips are becoming an integral part of pro-audio/video systems.

19.15 Glossary of Fiber Optic Terms

Absorption: Together with scattering, absorption forms the principal cause of the attenuation of an optical waveguide. It results from unwanted impurities in the waveguide material and has an effect only at certain wavelengths.

Angle of Incidence: The angle between an incident ray and the normal to a reflecting surface.

Attenuation: The reduction of average optical power in an optical waveguide, expressed in dB. The main causes are scattering and absorption, as well as optical losses in connectors and splices. Attenuation or loss is expressed by

$$\mu = -10\log\left(\frac{P_o}{P_i}\right)\text{dB}$$

Attenuator: An optical element that reduces intensity of an optical signal passing through it (i.e., attenuates it). Example: AT&T makes attenuators built into connectors that incorporate a biconic sleeve consisting of a carbon-coated mylar filter. They come in steps of 6dB, 12dB, 16dB, and 22dB values.

Avalanche Photodiode (APD): A photodiode designed to take advantage of avalanche multiplication of photo-current. As the reverse-bias voltage approaches the breakdown voltage, hole-electron pairs created by absorbed photons acquire sufficient energy to create additional hole-electron pairs when they collide with ions; thus, a multiplication or signal gain is achieved.

Axial Ray: A light ray that travels along the axis of an optical fiber.

Backscattering: A small fraction of light that is deflected out of the original direction of propagation by scattering suffers a reversal of direction. In other words, it propagates in the optical waveguide towards the transmitter.

Bandwidth: The lowest frequency at which the magnitude of the waveguide transfer function decreases to 3dB (optical power) below its zero frequency value. The bandwidth is a function of the length of the waveguide, but may not be directly proportional to the length.

Bandwidth Distance Product (BDP): The bandwidth distance product is a figure of merit that is normalized for a distance of 1km and is equal to the product of the optical fiber's length and the 3dB bandwidth of the optical signal. The bandwidth distance product is usually expressed in megahertz*kilometer (MHz*km) or gigahertz*kilometer (GHz*km). For example, a common multimode fiber with bandwidth-distance product of 500MHz*km could carry a 500MHz signal for 1km. Therefore, a 1000MHz or 1GHz signal for 0.5km. Thus, as the distance increases, for 2km, the BDP would be 250MHz etc.

Beamsplitter: A device used to divide or split an optical beam into two or more separate beams.

Beamwidth: The distance between two diametrically opposed points at which the irradiance is a specified fraction of the beam's peak irradiance; Beamwidth is most often applied to beams that are circular in cross section.

BER (Bit Error Rate): In digital applications, the ratio of bits received in error to bits sent. *BER*s of one errored bit per billion (1×10^{-9}) sent are typical.

Buffer: Material used to protect optical fiber from physical damage, providing mechanical isolation and/or protection. Fabrication techniques include tight or loose tube buffering, as well as multiple buffer layers.

Burrus LED: A surface-emitting LED with a hole etched to accommodate a light-collecting fiber. Named after its inventor, Charles A. Burrus of Bell Labs.

Chromatic Dispersion: Spreading of a light pulse caused by the difference in refractive indices at different wavelengths.

Cladding: The dielectric material surrounding the core of an optical fiber.

Coarse Wavelength Division Multiplexing (CWDM): CWDM is a cost-effective solution to dense wavelength division modulation (DWDM) that was developed to have channel spacing by the International Telecommunication Union (ITU) in 2002. This standard allows for a 20nm spacing of channels using wavelengths between 1270nm and 1610nm.

Coherent: Light source (laser) in which the amplitude of all waves is exactly equivalent and rise and fall together.

Core: The central region of an optical fiber through which light is transmitted.

Coupler: An optical component used to split or combine optical signals. Also known as a “Splitter,” “T-coupler,” “2 × 2,” or “1 × 2” coupler.

Coupling Loss: The power loss suffered when coupling light from one optical device to another.

Critical Angle: The smallest angle from the fiber axis at which a ray may be totally reflected at the core-cladding interface.

Cutoff Wavelength: The shortest wavelength at which only the fundamental mode of an optical waveguide is capable of propagation.

Dark Current: The external current that, under specified biasing conditions, flows in a photodetector when there is no incident radiation.

Data Rate: The maximum number of bits of information that can be transmitted per second, as in a data transmission link. Typically expressed as megabits per second (Mb/s).

Decibel (dB): The standard unit of level used to express gain or loss of optical or electrical power.

Dense Wavelength Division Multiplexing (DWDM): An enhancement of WDM (see Wavelength Division Multiplexing) that uses many wavelengths in the 1550nm window (ranges 1530nm to 1560nm) for transmitting multiple signals, and often uses fiber optic amplification. Many narrowband transmitters send signals to a DWDM Optical Multiplexer (Mux), which combines all of the signals onto a single fiber. At the other end a DWDM Optical Demultiplexer (Demux) separates the signals out to the many receivers.

Detector: A transducer that provides an electrical output signal in response to an incident optical signal. The current is dependent on the amount of light received and the type of device.

Dispersion: Spread of the signal delay in an optical waveguide. It consists of various components: modal dispersion, material dispersion, and waveguide dispersion. As a result of its dispersion, an optical waveguide acts as a low-pass filter for the transmitted signals.

Ferrule: A component of a fiber optic connector that holds a fiber in place and aids in its accurate alignment.

Fiber Data Distributed Interface (FDDI): An emerging standard developed by AT&T, Hewlett-Packard Co, and Siemens Corp., using a 100Mbps token ring network that employs dual optical fibers.

Fiber Optic: Any filament or fiber made of dielectric materials, that guides light.

Fiber Optic Link: A fiber optic cable with connectors attached to a transmitter (source) and receiver (detector).

Fresnel Reflection: The reflection of a portion of the light incident on a planar surface between two homogeneous media having different refractive indices. Fresnel reflection occurs at the air–glass interfaces at entrance and exit ends of an optical fiber.

Fundamental Mode: The lowest order mode of a waveguide.

Graded Index Fiber: An optical fiber with a variable refractive index that is a function of the radial distance from the fiber axis.

Incoherent: An LED light source that emits incoherent light as opposed to the laser which emits coherent light. (See Coherent.)

Index Matching Material: A material, often a liquid or cement, whose refractive index is nearly equal to the core index, used to reduce Fresnel reflections from a fiber end face.

Index of Refraction: See Refractive Index.

Injection Laser Diode (ILD): Laser diode.

Insertion Loss: The attenuation caused by the insertion of an optical component. In other words, a connector or coupler in an optical transmission system.

Intensity: Irradiance.

Integrated Optical Components (IOCs): Optical devices (singly or in combination) that use light transmission in waveguides. The waveguides structure and confine the propagating light to a region with one or two very small dimensions of the order of the wavelength of light. A common material used in the fabrication process of an IOC is Lithium Niobate (LiNbO).

Intermodal Distortion: Multimode distortion.

Irradiance: Power density at a surface through which radiation passes at the radiating surface of a light source or at the cross section of an optical waveguide. The normal unit is watts per centimeters squared, or W/cm^2.

Laser Diode (LD): Semiconductor diode that emits coherent light above a threshold current.

Launch Angle: Angle between the propagation direction of the incident light and the optical axis of an optical waveguide.

Launching Fiber: A fiber used in conjunction with a source to excite the modes of another fiber in a particular way. Launching fibers are most often used in test systems to improve the precision of measurements.

Light: In the laser and optical communication fields, the portion of the electromagnetic spectrum that can be handled by the basic optical techniques used for the visible spectrum extending from the near ultraviolet region of approximately 0.3micron, through the visible region and into the mid infrared region of about 30 microns.

Light Emitting Diode (LED): A semiconductor device that emits incoherent light from a p-n junction when biased with an electrical current in the forward direction. Light may exit from the junction strip edge or from its surface, depending on the device's structure.

Lightwaves: Electromagnetic waves in the region of optical frequencies. The term light was originally restricted to radiation visible to the human eye, with wavelengths between 400nm and 700nm. However, it has become customary to refer to radiation in the spectral regions adjacent to visible light (in the near infrared from 700nm to about 2000nm) as light to emphasize the physical and technical characteristics they have in common with visible light.

Macrobending: Macroscopic axial deviations of a fiber from a straight line, in contrast to microbending.

Microbending: Curvatures of the fiber that involve axial displacements of a few micrometers and spatial wavelengths of a few millimeters. Microbends cause loss of light and consequently increase the attenuation of the fiber.

Micron: Micrometer (μm). One millionth of a meter (1×10^{-6}m).

Modal Dispersion: Pulse spreading due to multiple light rays traveling different distances and speeds through an optical fiber.

Modal Noise: Disturbance in multimode fibers fed by laser diodes. It occurs when the fibers contain elements with mode-dependent attenuation, such as imperfect splices, and is more severe the better the coherence of the laser light.

Modes: Discrete optical waves that can propagate in optical waveguides. They are eigenvalue solutions to the differential equations that characterize the waveguide. In a singlemode fiber, only one mode, the fundamental mode, can propagate. There are several hundred modes in a multimode fiber that differ in field pattern and propagation velocity. The upper limit to the number of modes is determined by the core diameter and the numerical aperture of the waveguide.

Modified Chemical Vapor Deposition (MCVD) Technique: A process in which deposits are produced by heterogeneous gas/solid and gas/liquid chemical reactions at the surface of a substrate. The MCVD method is often used in fabricating optical waveguide preforms by causing gaseous material to react and deposit glass oxides. Typical starting chemicals include volatile compounds of silicon, germanium, phosphorus, and boron, which form corresponding oxides after heating with oxygen or other gases. Depending on its type, the preform may be processed further in preparation for pulling into an optical fiber.

Monochromatic: Consisting of a single wavelength. In practice, radiation is never perfectly monochromatic but, at best, displays a narrow band of wavelengths.

Multimode Distortion: The signal distortion in an optical waveguide resulting from the superposition of modes with differing delays.

Multimode Fiber: Optical waveguide whose core diameter is large compared with the optical wavelength and in which, consequently, a large number of modes are capable of propagation.

Nanometer (nm): One billionth of a meter (1×10^{-9}m).

Noise Equivalent Power (NEP): The rms value of optical power that is required to produce an rms *SNR* of 1; an indication of noise level that defines the minimum detectable signal level.

Numerical Aperture: A measure of the range of angles of incident light transmitted through a fiber. Depends on the differences in index of refraction between the core and the cladding.

Optimized Fiber Class (OM1, OM2, OM3, and OS1, OS2 designations in accordance with ISO11801: Bandwidth and the maximum transmission distance of different optical/optimized fiber classes for 10G Ethernet application, Table 19-7.

Table 19-7. Bandwidth and the Maximum Transmission Distance of Different Optimized Fiber Classes for 10G Ethernet Application

Fiber Type	Bandwidth 850nm MHz*km	Bandwidth 1300nm MHz*km	1Gbps Transmission Distance		10Gbps Transmission Distance		Fiber Class
Multimode			**@850nm**	**@1300nm**	**@850nm**	**@1300nm**	
Traditional 62.5/125 µm	200	500	275m	550m	33m	300m	OM1
Traditional 50/125 µm	400	800	500m	1000m	66m	450m	OM1
Traditional 50/125/62.5 µm	500	500	550m	550m	82m	300m	OM2
50/125 µm-110	600	1200	750m	2000m	110m	850m	OM2+
50/125 µm-150	700	500	750m	550m	150m	300m	OM2
50/125 µm-300	1500	500	1000m	550m	300m	300m	OM3
50/125 µm-550	3500	500	1000m	550m	550m	550m	NA
Singlemode			**@1310m**	**@1550nm**	**1310/1383/1550nm**		
Traditional 9/125 µm			5000m		5000m		OS1
8-10/125 µm			5000m		5000m–10,000m		OS2

Note: OS1 constructed with tight buffer and OS2 constructed with loose tube buffer fiber. Always consult with the optical cable manufacturer for the latest Optimized performance.

Optical Fiber Zero Water Peak: A peak in attenuation in optical fibers caused by contamination from hydroxyl (OH) ions that are residuals of the manufacturing process. Water peak causes wavelength attenuation and pulse dispersion in the general regions of 1360nm to 1460nm. Low-water-peak fiber (LWPF) and zero-water-peak fiber (ZWPF) resolves water peak issues in the 1360nm to 1460nm region thereby opening the entire spectrum from 1260 to 1625nm for high-performance optical transmission technologies employing coarse wavelength division multiplexing (CWDM).

Optical Time Domain Reflectometer (OTDR): A method for characterizing a fiber wherein an optical pulse is transmitted through the fiber and the resulting back-scatter and reflections to the input are measured as a function of time. Useful in estimating the attenuation coefficient as a function of distance and identifying defects and other localized losses.

Optoelectronic: Any device that functions as an electrical-to-optical or optical-to-electrical transducer.

Optoelectronic Integrated Circuits (OEICs): Combination of electronic and optical functions in a single chip.

Peak Wavelength: The wavelength at which the optical power of a source is at a maximum.

Photocurrent: The current that flows through a photosensitive device, such as a photodiode, as the result of exposure to radiant power.

Photodiode: A diode designed to produce photocurrent by absorbing light. Photodiodes are used for the detection of optical power and for the conversion of optical power into electrical power.

Photon: A quantum of electromagnetic energy.

Pigtail: A short length of optical fiber for coupling optical components. It is usually permanently fixed to the components.

PIN-FET Receiver: An optical receiver with a PIN photodiode and low noise amplifier with a high impedance input, whose first stage incorporates a field-effect transistor (FET).

PIN Photodiode: A diode with a large intrinsic region sandwiched between *p*-doped and *n*-doped semiconducting regions. Photons in this region create electron hole pairs that are separated by an electric field thus generating an electric current in the load circuit.

Plastic Optical Fiber (POF): An optical fiber composed of plastic instead of glass. POFs are used for short distances of typically 25ft or less.

Preform: A glass structure from which an optical fiber waveguide may be drawn.

Pre-terminated Optical Fiber: An optical fiber customized with any type OMx/OSx, specified length and fiber optic connector type. They are pre-made and tested to the requirements of your system. No need to attach/terminate a connector to an optical fiber in the field. It is done for the customer. This is truly Plug and Play.

Primary Coating: The plastic coating applied directly to the cladding surface of the fiber during manufacture to preserve the integrity of the surface.

Ray: A geometric representation of a light path through an optical medium; a line normal to the wavefront indicating the direction of radiant energy flow.

Rayleigh Scattering: Scattering by refractive index fluctuations (inhomogeneities in material density or composition) that are small with respect to wavelength.

Receiver: A detector and electronic circuitry to change optical signals into electrical signals.

Receiver Sensitivity: The optical power required by a receiver for low error signal transmission. In the case of digital signal transmission, the mean optical power is usually quoted in watts or dBm (decibels referenced to 1mW).

Reflection: The abrupt change in direction of a light beam at an interface between two dissimilar media so that the light beam returns to the media from which it originated.

Refraction: The bending of a beam of light at an interface between two dissimilar media or in a medium whose refractive index is a continuous function of position (graded index medium).

Refractive Index: The ratio of the velocity of light in a vacuum to that in an optically dense medium.

Repeater: In a lightwave system, an optoelectronic device or module that receives an optical signal, converts it to electrical form, amplifies or reconstructs it, and retransmits it in optical form.

Responsivity: The ratio of detector output to input, usually measured in units of amperes per watt (or microamperes per microwatt).

Singlemode Fiber: Optical fiber with a small core diameter in which only a singlemode—the fundamental mode—is capable of propagation. This type of fiber is particularly suitable for wideband transmission over large distances, since its bandwidth is limited only by chromatic dispersion.

Source: The means (usually LED or laser) used to convert an electrical information carrying signal into a corresponding optical signal for transmission by an optical waveguide.

Splice: A permanent joint between two optical waveguides.

Spontaneous Emission: This occurs when there are too many electrons in the conduction band of a semiconductor. These electrons drop spontaneously into vacant locations in the valence band, a photon being emitted for each electron. The emitted light is incoherent.

ST Connector: A type of connector used on fiber optic cable utilizing a spring-loaded twist and lock coupling similar to the BNC connectors used with coax cable.

Star Coupler: An optical component used to distribute light signals to a multiplicity of output ports. Usually the number of input and output ports are identical.

Step Index Fiber: A fiber having a uniform refractive index within the core and a sharp decrease in refractive index at the core-cladding interface.

Stimulated Emission: A phenomenon that occurs when photons in a semiconductor stimulate available excess charge carriers to the emission of photons. The emitted light is identical in wavelength and phase with the incident coherent light.

Superluminescent Diodes (SLDs): Superluminescent diodes (SLDs) are distinguished from both laser diodes and LEDs in that the emitted light consists of amplified spontaneous emission having a spectrum much narrower than that of LEDs but wider than that of lasers.

T (or Tee) Coupler: A coupler with three ports.

Threshold Current: The driving current above which the amplification of the light-wave in a laser diode becomes greater than the optical losses, so that stimulated emission commences. The threshold current is strongly temperature dependent.

Total Internal Reflection: The total reflection that occurs when light strikes an interface at angles of incidence greater than the critical angle.

Transmission Loss: Total loss encountered in transmission through a system.

Transmitter: A driver and a source used to change electrical signals into optical signals.

Tree Coupler: An optical component used to distribute light signals to a multiplicity of output ports. Usually the number of output ports is greater than the number of input ports.

Vertical Cavity Surface Emitting Laser (VCSEL): A specialized laser diode that promises to revolutionize fiber optic communications by improving efficiency and increasing data speed. The acronym VCSEL is pronounced vixel. Typically used for the 850nm and 1300nm windows.

Y Coupler: A variation on the T coupler in which input light is split between two channels (typically planar waveguide) that branch out like a Y from the input.

Wavelength Division Multiplexing (WDM): Simultaneous transmission of several signals in an optical waveguide at differing wavelengths.

Window: Refers to ranges of wavelengths matched to the properties of the optical fiber. The window ranges for fiber optics are:

- First window, 820nm to 850nm.
- Second window, 1300nm to 1310nm.
- Third window, 1360 nm to 1460 nm.

- Forth window, 1550nm.

Bibliography

R. G. Ajemian, "Audio Rides the Light of Fiber Optics: A Tutorial," *J. Audio Eng. Soc.*, Vol. 57, no. 6, June 2009.

R. G. Ajemian, and A. Grundy, "Fiber Optics: The New Medium for Audio," *J. Audio Eng. Soc.*, Vol. 38, no. 3, March 1990.

R. G. Ajemian, "Fiber Optic Connector Considerations for Professional Audio," *J. Audio Eng. Soc.*, Vol. 40, no. 6, June 1992.

R. G. Ajemian, "A Tutorial: Fiber Optic Cables and Connectors for Pro-Audio," *Audio Engineering Society*, 123rd Convention Paper 7210, October 2007.

R. G. Ajemian, *Handbook of Fiber Optic Connectors*, abc Teletraining, Inc. (AVO), 2000.

M. K. Barnoski, *Fundamentals of Optical Fiber Communications*, New York: Academic Press, 1981.

W. B. Bielowski, "Low Loss Waveguide Communication Technology," Corning, NY: Corning Glass Works. Corning, Inc., *Just the Facts magazine*, New York.

P. De Lancie, "Optic Link," *Mix Magazine*, August 1989.

"Guide to Fiber Optics," Richmond, IN: Belden Wire and Cable Company.

Jeff Hecht, *Understanding Fiber Optics*, Carmel, IN: SAMS, A division of Macmillan Computer Publishing, 1998. C. Hentschel, *Fiber Optics Handbook*, Palo Alto, CA: Hewlett Packard, 1988.

"IEEE Standard Definitions of Terms Relating to Fiber Optics ANSI/IEEE Std 812-1984," New York: The Institute of Electrical and Electronics Engineers, Inc.

Gerd Keiser, *Optical Fiber Communications*, New York: McGraw-Hill, 1999.

M. Lambert, "Making The Digital Connection," *Mix Magazine*, May 1989.

"Lightguide Catalog," Berkeley Heights, NJ: AT&T, 1990.

Gunther Mahlke, and Peter Gossing, *Fiber Optic Cables*, New York: John Wiley & Sons, 1987.

Neutrik USA, http://www.neutrik.com, Product Catalog (OpticalCon), 4115 Taggart Creek Road, Charlotte, NC.

"News of the Week (Scientists create smallest laser)," *Telephony*, July 24, 1989.

"Oak Link Catalog," Germany: Klotz & Co., 1989.

"Optimate Catalog 86-792," Harrisburg, PA: AMP Inc., May 1987.

Owl Fiber Optics, http://www.owlfo.com, Fiber Optic Glossary, Flushing, NY

J. C. Palais, *Fiber Optic Communications*, Englewood Cliffs, NJ: Prentice Hall, 1998.

"Practical Testing of Fiber Optic Systems and Components Seminar Workbook," Boston, MA: Fotec, Inc.

R. G. Seippel, *Optoelectronics for Technology & Engineering*, Englewood Cliffs, NJ: Prentice Hall, 1989.

D. J. Sterling, Jr., *Technician's Guide to Fiber Optics*, New York: Delmar Publishers Inc., 1999.

"Technical Data for Fiber Optic Link Toslink," Burlington, MA: Toshiba America, Inc., 2006.

"Effective Integration of Reduced Bend Radius Fiber into the Network," ADC Telecommunications, Inc., September 2007, Minneapolis, MN 55440-1101.

"Optical Transceiver Installation Guide", November 2003, 3Com Corporation, Southborough, MA USA 01772-1886.

Corning Cable Systems LLC, "UniCam Pretium Connector Catalog", June 2007, Hickory, NC 28603-0489 USA

"OptiFiber® Certifying OTDR", *Fluke Networks*, Everett WA 98203, USA

Part 4

Electroacoustic Devices

Chapter 20

Microphones

by Glen Ballou and Joe Ciaudelli

20.1 Introduction

All sound sources have different characteristics; their waveform varies, their phase characteristics vary, their dynamic range and attack time vary and their frequency response varies, just to name a few. No one microphone will reproduce all of these characteristics equally well. In fact, each sound source will sound better or more natural with one type or brand of microphone than all others. For this reason we have and always will have many types and brands of microphones.

Microphones are electroacoustic devices that convert acoustical energy into electrical energy. All microphones have a diaphragm or moving surface that is excited by the acoustic wave. The corresponding output is an electrical signal that represents the acoustic input.

Microphones fall into two classes: *pressure* and *velocity*. In a *pressure* microphone the diaphragm has only one surface exposed to the sound source so the output corresponds to the instantaneous sound pressure of the impressed sound waves. A pressure microphone is a *zero-order gradient microphone*, and is associated with omnidirectional characteristics.

The second class of microphone is the *velocity* microphone, also called a *first-order gradient microphone*, where the effect of the sound wave is the difference or gradient between the sound wave that hits the front and the rear of the diaphragm. The electrical output corresponds to the instantaneous particle velocity in the impressed sound wave. Ribbon microphones as well as pressure microphones that are altered to produce front-to-back discrimination are of the velocity type.

Microphones are also classified by their pickup pattern or how they discriminate between the various directions the sound source arrives from, see section 20.4 Microphone Pickup Patterns. These classifications are:

- *Omnidirectional*—pickup is equal in all directions.
- *Bidirectional*—pickup is equal from the two opposite directions (180°) apart and zero from the two directions that are 90° from the first.
- *Unidirectional*—pickup is from one direction only, the pickup appearing cardioid or heart-shaped.

The air particle relationships of the air particle displacement, velocity, and acceleration that a microphone sees as a plane wave in the far field, are shown in Fig. 20-1.

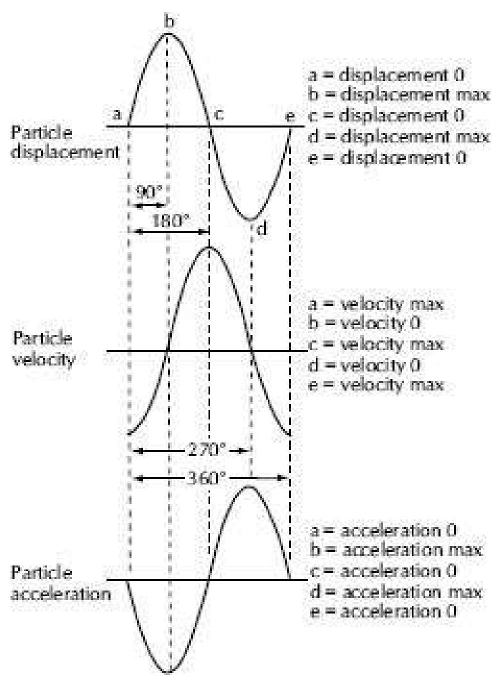


Figure 20-1. Air particle motion in a sound field, showing relationship to velocity and acceleration.

20.2 Microphone Specifications

20.2.1 Microphone Sensitivity

Microphone sensitivity[1] is the measure of the electrical output of a microphone with respect to the acoustic sound pressure level input.

Sensitivity is measured in one of three methods:

Open-circuit voltage	0 dB = 1 V/μbar
Maximum power output	0 dB = 1 mW/10 μbar = 1 mW/Pa
Electronic Industries Association (EIA) sensitivity	0 dB = EIA standard SE-105

The common sound pressure levels used for measuring microphone sensitivity are:

94dB SPL 10dyn/cm^2 SPL	10μbar or 1Pa
74dB SPL 1dyn/cm^2 SPL	1μbar or 0.1Pa
0dB SPL 0.0002dyn/cm^2 SPL	0.0002Pa or 20μPa—threshold of hearing

94dB SPL is recommended since 74dB SPL is too close to typical noise levels.

20.2.2 Open-Circuit Voltage Sensitivity

There are several good reasons for measuring the open-circuit voltage:

- If the open-circuit voltage and the microphone impedance are known, the microphone performance can be calculated for any condition of loading.
- It corresponds to an effective condition of use. A microphone should be connected to a high impedance to yield maximum *SNR*. A 150–250Ω microphone should be connected to 2kΩ or greater.
- When the microphone is connected to a high impedance

compared to its own, variations in microphone impedance do not cause variations in response.

The open-circuit voltage sensitivity (S_v) can be calculated by exposing the microphone to a known SPL, measuring the voltage output, and using the following equation:

$$S_v = 20\log E_o - dB_{SPL} + 94 \qquad (20\text{-}1)$$

where,
S_v is the open-circuit voltage sensitivity in dB re 1V for a 10dyn/cm^2 SPL (94dB SPL) acoustic input to the microphone,
E_o is the output of the microphone in V,
dB_{SPL} is the level of the actual acoustic input.

The microphone measurement system can be setup as shown in Fig. 20-2. The setup requires a random-noise generator, a microvoltmeter, a high-pass and a low-pass filter set, a power amplifier, a test-loudspeaker, and a sound level meter (SLM). The SLM is placed a specific measuring distance (about 5–6ft or 1.5–2m) in front of the loudspeaker. The system is adjusted until the SLM reads 94dB SPL (a band of pink noise from 250 to 5000Hz is excellent for this purpose). The microphone to be tested is now substituted for the SLM.

It is often necessary to know the voltage output of the microphone for various SPLs to determine whether the microphone will overload the preamplifier circuit or the *SNR* will be inadequate. To determine this, use

$$E_o = 10\left(\frac{S_v + dB_{SPL} - 94}{20}\right) \qquad (20\text{-}2)$$

where,
E_o is the voltage output of microphone,
S_v is the open-circuit voltage sensitivity,
dB_{SPL} is the sound pressure level at the microphone.

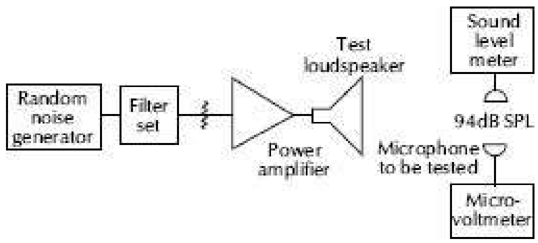


Figure 20-2. Method of determining open-circuit voltage sensitivity of a microphone. (From Reference 1.)

20.2.3 Maximum Power Output Sensitivity

The *maximum power output sensitivity*[1] from of specification gives the maximum power output in decibels available from the microphone for a given sound pressure and power reference. Such a specification can be calculated from the internal impedance and the open-circuit voltage of the microphone. This specification also indicates the ability of a microphone to convert sound energy into electrical power. The equation is

$$S_p = 10\log\frac{V_o^2}{R_o} + 44\,\text{dB} \qquad (20\text{-}3)$$

where,
S_p is the power level microphone sensitivity in dB,
V_o is the open-circuit voltage produced by a 1μbar (0.1Pa) sound pressure,
R_o is the internal impedance of the microphone.

The form of this specification is similar to the voltage specification except that a power as opposed to a voltage reference is given with the sound pressure reference. A 1mW power reference and a 10 μbar (1 Pa) pressure reference are commonly used. This form of microphone specification is quite meaningful because it takes into account both the voltage output and the internal impedance of the microphone.

S_p can also be calculated easily from the open-circuit voltage sensitivity

$$S_p = S_v - 10\log Z + 44\,\mathrm{dB} \qquad (20\text{-}4)$$

where,

S_p is the dB rating for an acoustical input of 94dB SPL (10dyn/cm^2) or 1Pa,

Z is the measured impedance of the microphone (the specifications of most manufacturers use the rated impedance).

The output level can also be determined directly from the open-circuit voltage

$$S_p = 10\log\frac{E_o^2}{0.001Z} - 6\,\mathrm{dB} \qquad (20\text{-}5)$$

where,

E_o is the open-circuit voltage,

Z is the microphone impedance.

Because the quantity 10log(E^2/0.001Z) treats the open-circuit voltage as if it appears across a load, it is necessary to subtract 6dB. (The reading is 6dB higher than it would have been had a load been present.)

20.2.4 Electronic Industries Association (EIA) Output Sensitivity

The Electronic Industries Association (EIA) Standard SE-105, August 1949, defines the system rating (G_M) as the ratio of the maximum electrical output from the microphone to the square of the undisturbed sound field pressure in a plane progressive wave at the microphone in decibels relative to 1mW/0.0002 dyn/cm^2. Expressed mathematically,

$$G_M = 20\log\frac{E_o}{P} - 10\log Z_o - 50\text{ dB} \qquad (20\text{-}6)$$

where,
E_o is the open-circuit voltage of the microphone,
P is the undisturbed sound field pressure in dyn/cm^2,
Z_o is the microphone-rated output impedance in Ω.

For all practical purposes, the output level of the microphone can be obtained by adding the sound pressure level relative to 0.0002dyn/cm^2 to G_M.

Because G_M, S_V, and S_P are compatible, G_M can also be calculated

$$G_M = S_v - 10\log R_{MR} - 50\text{ dB} \qquad (20\text{-}7)$$

where,
G_M is the EIA rating,
R_{MR} is the EIA center value of the nominal impedance range shown below.

Ranges (Ω)		Values Used (Ω)
20–80	=	38
80–300	=	150
300–1250	=	600
1250–4500	=	2400
4500–20,000	=	9600
20,000–70,000	=	40,000

The EIA rating can also be determined from the chart in Fig. 20-3.

20.2.5 Various Microphone Sensitivities

Microphones are subjected to sound pressure levels anywhere from 40dB SPL when distant micing to 150dB SPL when extremely close micing, i.e., 1/4 in (6mm) from the rock singer's mouth or inside a drum or horn).

Various types of microphones have different sensitivities, which is important to know if different types of microphones are intermixed since gain settings, *SNR*, and preamplifier overload will vary. Table 20-1 gives the sensitivities of a variety of different types of microphones.

20.2.6 Microphone Thermal Noise

Since a microphone has an impedance, it generates thermal noise. Even without an acoustic signal, the microphone will still produce a minute output voltage. The thermal noise voltage, E_n, produced by the electrical resistance of a sound source is dependent on the frequency bandwidth under consideration, the magnitude of the resistance, and the temperature existing at the time of the measurement. This voltage is

$$E_n = 4ktR(bw) \tag{20-8}$$

where,
k is the Boltzmann's constant, 1.38×10^{-23} J/K,
t is the absolute temperature, 273° + room temperature, both in °C,
R is the resistance in Ω,
bw is the bandwidth in Hz.

To change this to *dBv* use

$$EIN_{dBv} = 20\log\frac{E_n}{0.775} \tag{20-9}$$

The thermal noise relative to 1V is −198dB for a 1Hz bandwidth and 1Ω impedance. Therefore,

$$\frac{TN}{1\,\mathrm{V}} = -198\,\mathrm{dB} + 10\log(bw) + 10\log Z \tag{20-10}$$

where,
TN is the thermal noise relative to 1V,
bw is the bandwidth in Hz,
Z is the microphone impedance in Ω.

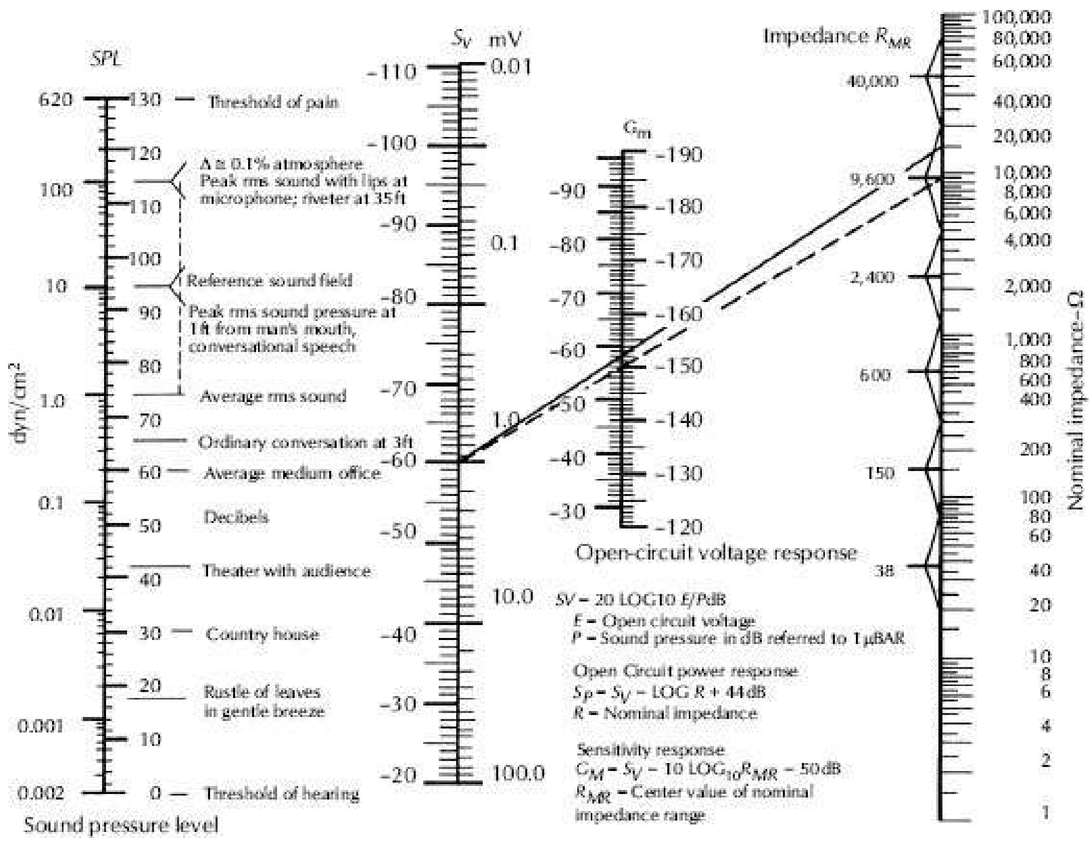


Figure 20-3. Microphone sensitivity conversion chart.

Table 20-1. Sensitivities of Various Types of Microphones

Type of Microphone	*Sp*	*Sv*
Carbon-button	−60 to −50dB	
Crystal		−50 to −40dB
Ceramic		−50 to −40dB
Dynamic (moving coil)	−60 to −52dB	−85 to −70dB
Capacitor	−60 to −37dB	−85 to −45dB
Ribbon-velocity	−60 to −50dB	−85 to −70dB
Transistor	−60 to −40dB	
Sound power	−32 to −20dB	
Line level	−40 to 0dB	−20 to 0dB
Wireless	−60 to 0dB	−85 to 0dB

Thermal noise relative to IV can be converted to equivalent input noise (EIN) by

$$EIN_{dBm} = -198\,\text{dB} + 10\log(bw) + 10\log Z - 6 - 20\log 0.775\ \text{V} \qquad (20\text{-}11)$$

Since the *EIN* is in dBm and dBm is referenced to 600Ω, the impedance *Z* is 600Ω.

20.3 Microphone Practices

20.3.1 Placement

Microphones are placed in various relationships to the sound source to obtain various sounds. Whatever position gives the desired effect that is wanted is the correct position. There are no exact rules that must be followed, however, certain recommendations should be followed to assure a good sound.

20.3.1.1 Microphone-to-Source Distance

Microphones are normally used in the direct field. Under this condition, inverse square law attenuation prevails, meaning that each time the distance is doubled, the microphone output is reduced 6dB. For instance, moving from a microphone-to-source distance of 2.5 to 5cm (1 to 2in) has the same effect as moving from 15 to 30cm (6 to 12in), 1 to 2 ft (30 to 60cm), or 5 to 10ft (1.5 to 3m).

Distance has many effects on the system. In a reinforcement system, doubling the distance reduces gain before feedback 6dB; in all systems it reduces the effect of microphone-to-source variations.

Using the inverse-square-law equation for attenuation,

$$attenuation_{dB} = 20\log\frac{D_1}{D_2} \quad (20\text{-}12)$$

it can be seen, at a microphone-to-source distance of 2.5 cm (1 in), moving the microphone only 1.25cm (½in) closer will increase the signal 6dB and 1.25cm (½in) farther away will decrease the signal 3.5dB for a total signal variation of 9.5dB for only 2.5cm (1in) of total movement! At a source-to-microphone distance of 30cm (12in), a movement of 2.5cm (1in) will cause a signal variation of only 0.72dB. Both conditions can be used advantageously; for instance, close micing is useful in feedback-prone areas, high noise level areas (rock groups), or where the talent wants to use the source to microphone variations to create an effect.

The farther distances are most useful where lecterns and table microphones are used or where the talker wants movement without level change.

The microphone-to-source distance also has an effect on the sound of a microphone, particularly one with a cardioid pattern. As the distance decreases, the proximity effect increases creating a bassy sound (see section 20.4.3.1 Proximity Effects). Closing in on the microphone also increases breath noise and pop noise.

20.3.1.2 Distance from Large Surfaces

When a microphone is placed next to a large surface such as the floor, 6dB of gain can be realized, which can be a help when far micing.

As the microphone is moved away from the large surface but still in proximity of it, cancellation of some specific frequencies will occur, creating a notch of up to 30dB, Fig. 20-4. The notch is created by the cancellation of a frequency that, after reflecting off the surface, reaches the microphone diaphragm 180° out of polarity with the direct sound.

The frequency of cancellation, f_c, can be calculated from the equation

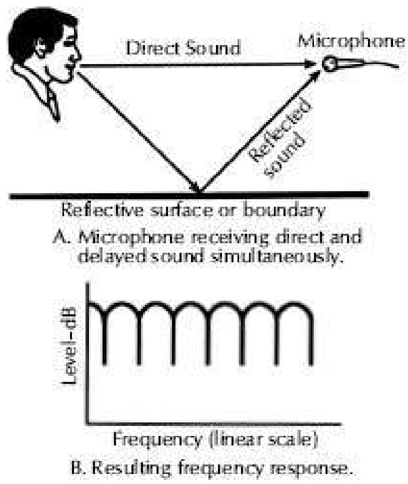


Figure 20-4. Effects of cancellation caused by near reflections (comb filters).

$$f_c = \frac{0.5c}{D_{r1} + D_{r2} - D_d} \qquad (20\text{-}13)$$

where,

c is the speed of sound, 1130ft/s or 344m/s,

0.5 is the out-of-polarity frequency ratio,

D_{r1} is the reflected path from the source to the surface in ft or m,

D_{r2} is the reflected path from the surface to the microphone in ft or m,

D_d is the direct path from the source to the microphone in ft or m.

If the microphone is 10ft from the source and both are 5ft above the floor, the canceled frequency is

$$f_c = \frac{1130 \times 0.5}{7.07 + 7.07 - 10}$$
$$= 136.47\,\text{Hz}$$

If the microphone is moved to 2ft (61cm) above the floor, the canceled frequency is 319.20Hz. If the microphone is 6in (15.2cm) from the floor, the canceled frequency is 1266.6Hz. If the microphone is 1in from the floor, the canceled frequency is 7239.7Hz.

20.3.1.3 Behind Objects

Sound, like light, does not go through solid or acoustically opaque objects. It does, however, go through objects of various density. The transmission loss or ability of sound to go through this type of material is frequency dependent; so if an object of this type is placed between the sound source and the microphone, the pickup will be attenuated according to the transmission characteristics of the object.

Low-frequency sound bends around objects smaller than their wavelength, which affects the frequency response of the signal. The normal effect of placing the microphone behind an object is an overall reduction of level, a low-frequency boost, and a high-frequency roll-off.

20.3.1.4 Above the Source

When the microphone is placed above or to the side of a directional sound source (i.e., horn or trumpet), the high-end frequency response will roll off because high frequencies are more directional than low frequencies, so less high-frequency *SPL* will reach the microphone than low-frequency *SPL*.

20.3.1.5 Direct versus Reverberant Field

Micing in the reverberant field picks up the characteristic of the room because the microphone is picking up as much or more of the room, as it is the direct sound from the source. When micing in the reverberant field, only two microphones are required for stereo since isolation of the individual sound sources is impossible. When in the reverberant field, a directional microphone will lose much of its directivity. Therefore, it is often advantageous to use an omnidirectional microphone that has smoother frequency response. To mic sources individually, you must be in the direct field and usually very close to the source to eliminate cross-feed.

20.3.2 Grounding

The grounding of microphones and their interconnecting cables is of extreme importance since any hum or noise picked up by the cables will be amplified along with the audio signal. Professional systems generally use the method shown in Fig. 20-5. Here the signal is passed through a two-conductor shielded cable to the balanced input of a preamplifier. The cable shield is connected to pin number 1, and the audio signal is carried by the two conductors and pins 2 and 3 of the XLR-type connector. The actual physical ground is connected at the preamplifier chassis only and carried to the microphone case. In no instance is a second ground ever connected to the far end of the cable, because this will cause the flow of ground currents between two points of grounding.

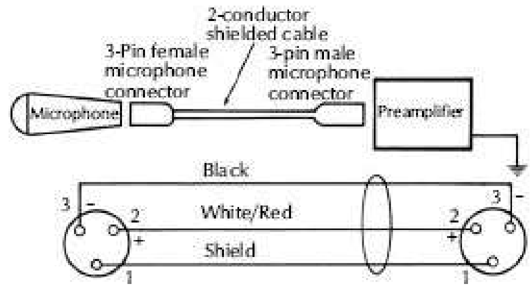


Figure 20-5. Typical low-impedance microphone to preamplifier wiring.

In systems designed for semiprofessional and home use, the method in Fig. 20-6 is often used. Note that one side of the audio signal is carried over the cable shield to a pin-type connector. The bodies of both the male and female connector are grounded: the female to the amplifier case and the male to the cable shield. The microphone end is connected in a similar manner; here again the physical ground is connected only at the preamplifier chassis. Hum picked up on the shield and not on the center conductor is added to the signal and amplified through the system.

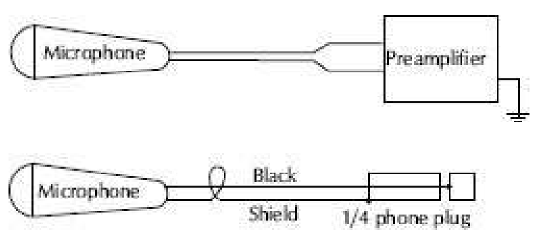


Figure 20-6. Typical semiprofessional, hi-fi microphone to preamplifier wiring.

20.3.3 Polarity

Microphone polarity, or *phase* as it is often called, is important especially when multiple microphones are used. When they are in polarity they add to each other rather than have canceling effects. If multiple microphones are used and one is out of polarity, it will cause comb filters, reducing quality and stereo enhancement. The EIA standard RS-221.A, October 1979, states “Polarity of a microphone or a microphone transducer element refers to in-phase or out-of-phase condition of voltage developed at its terminals with respect to the sound pressure of a sound wave causing the voltage.”

Note: *Exact in-phase relationship* can be taken to mean that the voltage is coincident with the phase of the sound pressure wave causing the voltage. In practical microphones, this perfect relationship may not always be obtainable.

The positive or in-phase terminal is that terminal that has a positive potential and a phase angle less than 90° with respect to a positive sound pressure at the front of the diaphragm.

When connected to a *three-pin XLR connector* as per EIA standard RS-297, the polarity shall be as follows:

- Out-of-phase—terminal 3 (black).
- In-phase—terminal 2 (red or any color other than black).
- Ground—terminal 1 (shield).

Fig. 20-7 shows the proper polarity for three-pin and five-pin XLR connectors and for three-pin and five-pin DIN connectors.

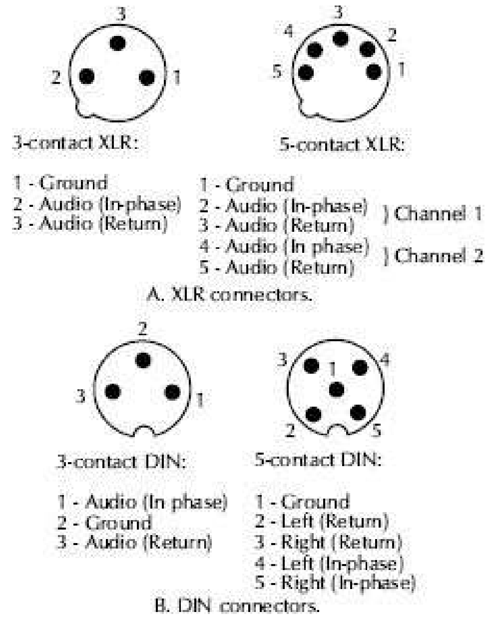


Figure 20-7. Microphone connector polarity.

A simple method of determining microphone polarity is as follows:

Place two microphones with the same frequency response and sensitivity next to each other and connect them to separate inputs of the same mixer. With a single acoustic source into the microphones (pink noise is a good source), one mixer volume control is adjusted for a normal output level as indicated on a VU meter. Note the volume control setting and turn it off. Make the same adjustment for the second microphone, and note the setting of this volume control.

Now open both controls to the same settings. If the microphones are out of polarity, the quality of reproduction will be distorted and the total output could be down 40–50dB from the output of only one microphone. Reversing the electrical connections to one microphone will bring them into polarity, making the quality about the same as one microphone operating and the output level will be higher.

If the microphones are of the bidirectional type, one may be turned 180° to bring it into polarity and later corrected electrically. If the microphones are of the directional type, only the output or cable connections can be reversed. After polarizing a bidirectional microphone, the rear should be marked with a white stripe for future reference.

20.3.4 Balanced or Unbalanced

Microphones can be connected either *balanced* or *unbalanced*. All professional installations use a balanced system for the following reasons:

- Reduced pickup of hum.
- Reduced pickup of electrical noise and transients.
- Reduced pickup of electrical signals from adjacent wires.

These reductions are realized because the two signal conductors shown in Fig. 20-8 pick up the same stray signal with equal intensity and polarity, so the noise is impressed evenly on each end of the transformer primary, eliminating a potential across the transformer and canceling any input noise. Because the balanced wires are in a shielded cable, the signal to each conductor is also greatly reduced.

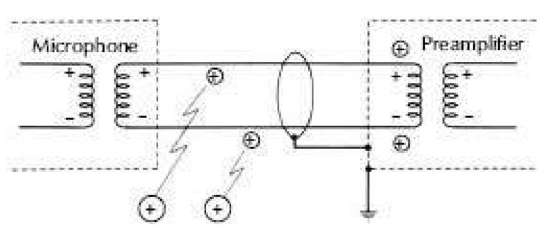


Figure 20-8. Noise cancellation on balanced, shielded microphone cables.

When installing microphones into an unbalanced system, any noise that gets to the inner unbalanced conductor is not canceled by the noise in the shield, so the noise is transmitted into the preamplifier. In fact, noise impressed on the microphone end of the shield adds to the signal because of the resistance of the shield between the noise and the amplifier.

Balanced low-impedance microphone lines can be as long as 500ft (150m) but unbalanced microphone lines should never exceed 15ft (4.5m).

20.3.5 Impedance

Most professional microphones are low impedance, 200Ω, and are designed to work into a load of 2000Ω. High-impedance microphones are 50,000Ω and are designed to work into an impedance of 1-10MΩ. The low-impedance microphone has the following advantages:

- Less susceptible to noise. A noise source of relatively high impedance cannot "drive" into a source of relatively low impedance (i.e., the microphone cable).
- Capable of being connected to long microphone lines without noise pickup and high-frequency loss.

All microphone cable has inductance and capacitance. The capacitance is about 40pF (40×10^{-12}F)/ft (30cm). If a cable is 100ft long (30m), the capacitance would be (40×10^{-12}) × 100ft or 4×10^{-9}F or 0.004μF. This is equivalent to a 3978.9Ω impedance at 10,000Hz and is found with the equation

$$X_c = \frac{1}{2\pi f C} \tag{20-14}$$

This has little effect on a microphone with an impedance of 200Ω as it does not reduce the impedance appreciably as determined by

$$Z_T = \frac{X_c Z_m}{X_c + Z_m} \tag{20-15}$$

For a microphone impedance of 200Ω, the total impedance Z_T=190Ω or less than 0.5dB.

If this same cable were used with a high-impedance microphone of 50,000Ω, 10,000Hz would be down more than 20dB.

Making the load impedance equal to the microphone impedance will reduce the microphone sensitivity 6dB, which reduces the overall *SNR* by 6dB. For the best *SNR*, the input impedance of low-impedance microphone preamplifiers is always 2000Ω or greater.

If the load impedance is reduced to less than the microphone impedance, or the load impedance is not resistive, the microphone frequency response and output voltage will be affected.

Changing the load of a high-impedance or ceramic microphone from 10MΩ to 100kΩ reduces the output at 100Hz by 27dB.

20.4 Microphone Pickup Patterns

Microphones are made with single- or multiple-pickup patterns and are named by the pickup pattern they employ. The pickup patterns and directional response characteristics of the various types of microphones are shown in Fig. 20-9.

20.4.1 Omnidirectional Microphones

The omnidirectional, or spherical polar response of the pressure microphones, Fig. 20-10 is created because the diaphragm is only exposed to the acoustic wave on the front side. No cancellations are produced by having sound waves hitting both the front and rear of the diaphragm at the same time.

Omnidirectional microphones become increasingly directional as the diameter of the microphone reaches the wavelength of the frequency in question, as shown in Fig. 20-11. The microphone should have the smallest diameter possible if omnidirectional characteristics are required at high frequencies. The characteristic that allows waves to bend around objects is known as *diffraction* and happens when the wavelength is long compared to the size of the object. As the wavelength approaches the size of the object, the wave cannot bend sharply enough and, therefore, passes by the object. The various responses start to diverge at the frequency at which the diameter of the diaphragm of the microphone, *D*, is approximately one-tenth the wavelength, λ, of the sound arriving

$$D = \frac{\lambda}{10} \tag{20-16}$$

The frequency, *f*, at which the variation begins is

$$f = \frac{v}{10D} \tag{20-17}$$

where,
v is the velocity of sound in ft/s, or m/s,
D is the diameter of the diaphragm in ft or m.

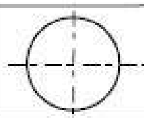
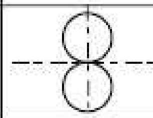
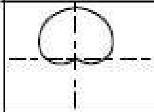
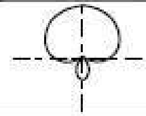
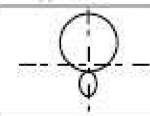
Microphone	Omnidirectional	Bidirectional	Directional	Supercardioid	Hypercardioid
Directional Response Characteristics					
Voltage output	$E = E_0$	$E = E_0 \cos\theta$	$E = \frac{E_0}{2}(1 + \cos\theta)$	$E = \frac{E_0}{2}[(\sqrt{3} - 1) + (3 - \sqrt{3})\cos\theta]$	$E = \frac{E_0}{4}(1 + 3\cos\theta)$
Random energy Efficiency (%)	100	33	33	27	25
$\frac{\text{Front response}}{\text{Back response}}$	1	1	∞	3.8	2
$\frac{\text{Front random response}}{\text{Total random response}}$	0.5	0.5	0.67	0.93	0.87
$\frac{\text{Front random response}}{\text{Back random response}}$	1	1	7	14	7
Equivalent distance	1	1.7	1.7	1.9	2
Pickup angle (2θ) for 3dB attenuation	-	90°	130°	116°	100°
Pickup angle (2θ) for 6dB attenuation	-	120°	180°	156°	140°

Figure 20-9. Performance characteristics of various microphones.

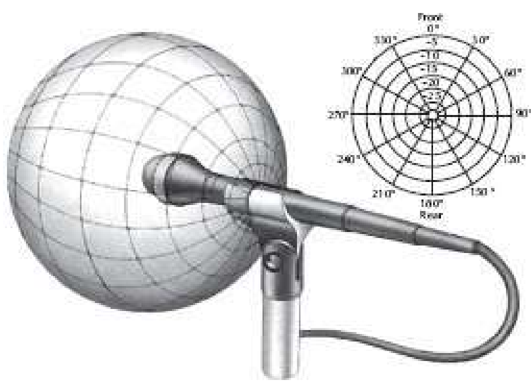


Figure 20-10. Omnidirectional pickup pattern. Courtesy Shure Incorporated.

For example, a ½ in (1.27cm) microphone will begin to vary from omnidirectional, though only slightly, at

$$f = \frac{1130}{\left(10 \times \frac{0.5}{12}\right)}$$
$$= 2712\,\text{Hz}$$

and will be down approximately 3dB at 10,000Hz.

Omnidirectional microphones are capable of having a very flat, smooth frequency response over the entire audio spectrum because only the front of the diaphragm is exposed to the source, eliminating phase cancellations found in unidirectional microphones.

For smoothness of response the smaller they are, the better. The problem usually revolves around the smallest diaphragm possible versus the lowest signal-to-noise ratio, *SNR*, or the smaller the diaphragm, the lower the microphone sensitivity, therefore poorer *SNR*.

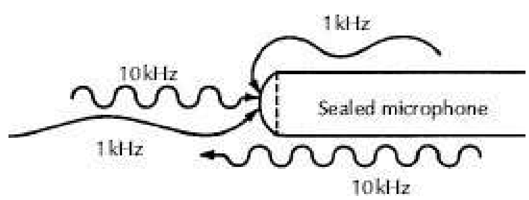


Figure 20-11. High-frequency directivity of an omnidirectional microphone.

Omnidirectional microphones have very little proximity effect. See section 20.4.3.1 Proximity Effects.

Because the pickup pattern is spherical, the random energy efficiency is 100%, and the ratio of front response to back or side is 1:1, therefore signals from the sides or rear will have the same pickup sensitivity as from the front, giving a directivity index of 0dB. This can be helpful in picking up wanted room characteristics or conversations around a table as when recording a symphony. However, it can be detrimental when in a noisy environment.

Omnidirectional microphones are relatively free from mechanical shock because the output at all frequencies is high so the diaphragm can be stiff. This allows the diaphragm to follow the magnet or stationary system it operates against when subjected to mechanical motion.

20.4.2 Bidirectional Microphones

A *bidirectional microphone* is one that picks up from the front and back equally well with little or no pickup from the sides. The field pattern, Fig. 20-12, is called a *figure eight*.

Because the microphone discriminates between the front, back, and sides, random energy efficiency is 33%. In other words, background noise, if it is in a reverberant field, will be 67% lower than with an omnidirectional microphone. The front-to-back response will still remain one; however, the front-to-side response will approach infinity, producing a directivity index of 4.8. This can be extremely useful when picking up two conversations on opposite sides of a table. The increased directional capabilities of the microphone means the pickup distance is 1.7 times greater before feedback in the direct field than for an omnidirectional microphone. The included pickup cone angle shown in Fig. 20-13 for 6dB attenuation on a perfect bidirectional microphone is 120° off the front of the microphone and 120° off the rear of the microphone. Diffraction causes this angle to vary with frequency, becoming narrower as the frequency increases.

Figure 20-12. Bidirectional pickup pattern. Courtesy Sennheiser Electronic Corporation.

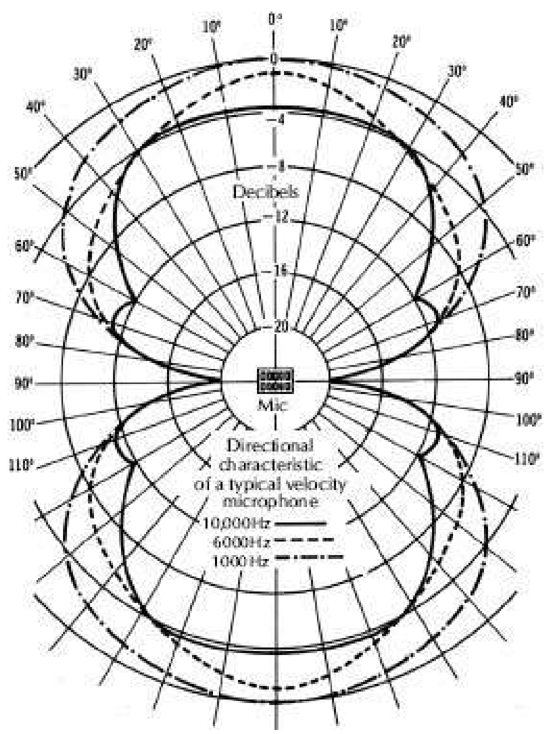


Figure 20-13. Polar pattern of a typical bidirectional ribbon velocity microphone showing the narrowing pattern at high frequencies.

20.4.3 Unidirectional Microphones

Unidirectional microphones have a greater sensitivity to sound pickup from the front than any other direction. The average unidirectional microphone has a front-to-back ratio of 20–30dB; that is, it has 20–30dB greater sensitivity to sound waves approaching from the front than from the rear.

Unidirectional microphones are usually listed as *cardioid* or *directional*, Fig. 20-14, *supercardioid*, Fig. 20-15, or *hypercardioid*, Fig. 20-16. The pickup pattern is called *cardioid* because it is heart shaped. Unidirectional microphones are the most commonly used microphones because they discriminate between signal and random unwanted noise. This has many advantages including:

- Less background noise,
- More gain before feedback especially when used in the direct field,
- Discrimination between sound sources.

Figure 20-14. Cardioid pickup pattern. Courtesy Shure Incorporated.

Figure 20-15. Supercardioid pickup pattern. Courtesy Shure Incorporated.

The cardioid pattern can be produced by one of two methods:

1. The first method combines the output of a pressure diaphragm and a pressure-gradient diaphragm, as shown in Fig. 20-17. Since the pressure-gradient diaphragm has a bidirectional pickup pattern and the pressure diaphragm has an omnidirectional pickup pattern, the wave hitting the front of the diaphragms add, while the wave hitting the rear of the diaphragms cancel as it is 180° out-of-phase with the rear pickup pattern of the pressure diaphragm. This method is expensive and seldom used for sound reinforcement or general-purpose microphones.

Figure 20-16. Hypercardioid pickup pattern. Courtesy Sennheiser Electronic Corporation.

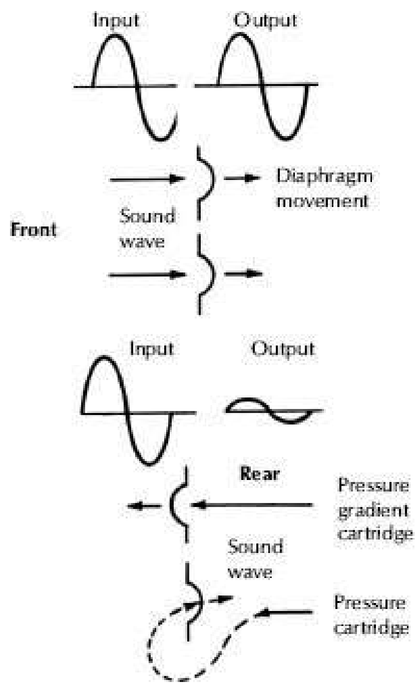


Figure 20-17. Two-diaphragm cardioid microphone.

2. The second and most widely used method of producing a cardioid pattern is to use a single diaphragm and acoustically delay the wave reaching the rear of the diaphragm. When a wave approaches from the front of the diaphragm, it first hits the front and then the rear of the diaphragm after traveling through an acoustical delay circuit, as shown in Fig. 20-18A. The pressure on the front of the diaphragm is at 0° while on the rear of the diaphragm it is some angle between 0° and 180°, as shown in Fig. 20-18B. If the rear pressure was at 0°, the output would be 0. It would be ideal if the rear pressure were at 180° so that it could add to the input, doubling the output.

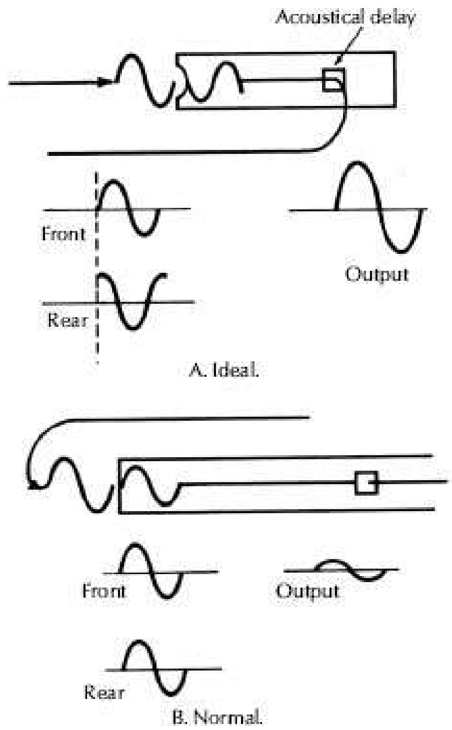


Figure 20-18. Cardioid microphone employing acoustical delay.

The phase inversion is caused by the extra distance the wave has to travel to reach the back of the diaphragm. When the wave is coming from the rear of the microphone, it hits the front and back of the diaphragm at the same time and with the same polarity, therefore canceling the output.

The frequency response of cardioid microphones is rougher than an omnidirectional microphone due to the acoustical impedance path and its effects on the front wave response. The front and rear responses of a cardioid microphone are not the same. Although the front pattern may be essentially flat over the audio spectrum, the back response usually increases at low and high frequencies, Fig. 20-19.

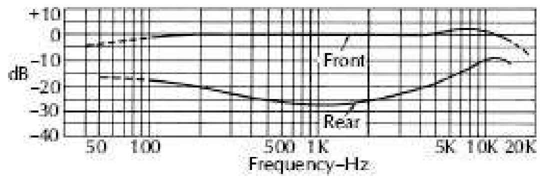


Figure 20-19. Frequency response of a typical cardioid microphone.

Discrimination between the front and back response is between 15 and 30dB in the mid frequencies and could be as little as 5–10dB at the extreme ends, Fig. 20-19.

20.4.3.1 Proximity Effects

As the source is moved closer to the diaphragm, the low-frequency response increases due to the proximity effect, Fig. 20-20. The proximity effect[2] is created because at close source-to-microphone distance, the magnitude of the sound pressure on the front is appreciably greater than the sound pressure on the rear. In the vector diagram shown in Fig. 20-21A, the sound source was a distance greater than 2ft from the microphone. The angle *2KD* is found from D, which is the acoustic distance from the front to the rear of the diaphragm and $K = 2\pi/\lambda$. Fig. 20-21B shows the vector diagram when used less than 4inches to the sound source.

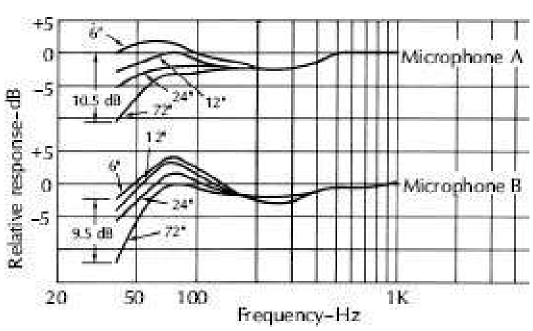


Figure 20-20. Proximity effect variations in response with distance between source and microphone for cardioid microphones.

Courtesy Telex Electro-Voice.

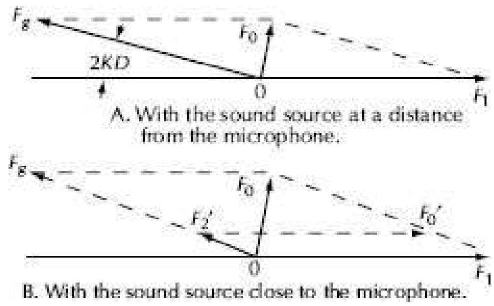


Figure 20-21. Vector diagram of a unidirectional microphone. Courtesy Telex Electro-Voice.

In both cases, force Fj, the sound pressure on the front of each diaphragm, is the same. Force F_2 is the force on the back of the diaphragm when the microphone is used at a distance from the sound source, and F_0 is the resultant force. The force F_2' on the back of the diaphragm is created by a close sound source. Laterally, the vector sum F_0' is considerably larger in magnitude than F_0 and therefore produces greater output from the microphone at low frequencies. This can be advantageous or disadvantageous. It is particularly useful when vocalists want to add low frequency to their voice or an instrumentalist to add low frequencies to the instrument. This is accomplished by varying the distance between the microphone and the sound source, increasing bass as the distance decreases.

Unidirectional microphones are much more sensitive to vibration relative to their acoustic sensitivity than omnidirectional types. Fig. 20-22 shows a plot of vibration sensitivity versus frequency for a typical omnidirectional and unidirectional microphone with the levels normalized with respect to acoustical sensitivity.

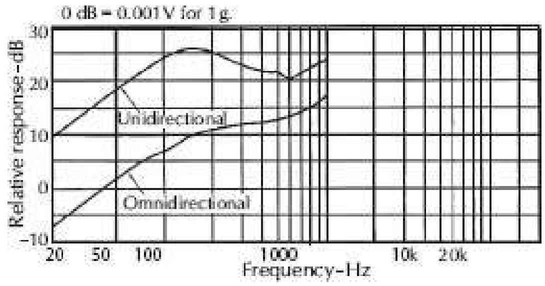


Figure 20-22. Vibration sensitivity of microphone cartridge.

The vibration sensitivity of the unidirectional microphone is about 15dB higher than the omnidirectional and has a peak at about 150Hz. The peak gives a clue to help explain the difference.

Unidirectional microphones are usually differential microphones; that is, the diaphragm responds to a pressure differential between its front and back surfaces. The oncoming sound wave is not only allowed to reach the front of a diaphragm but, through one or more openings and appropriate acoustical phase-shift networks, reaches the rear of the diaphragm. At low frequencies, the net instantaneous pressure differential causing the diaphragm to move is small compared to the absolute sound pressure, Fig. 20-23. Curve A is the pressure wave that arrives at the front of the diaphragm. Curve B is the pressure wave that reaches the rear of the diaphragm after a slight delay due to the greater distance the sound had to travel to reach the rear entry and some additional phase shift it encounters after entering. The net pressure actuating the diaphragm is curve C, which is the instantaneous difference between the two upper curves. In a typical unidirectional microphone, the differential pressure at 100Hz will be about one-tenth of the absolute pressure or 20dB down from the pressure an omnidirectional microphone would experience.

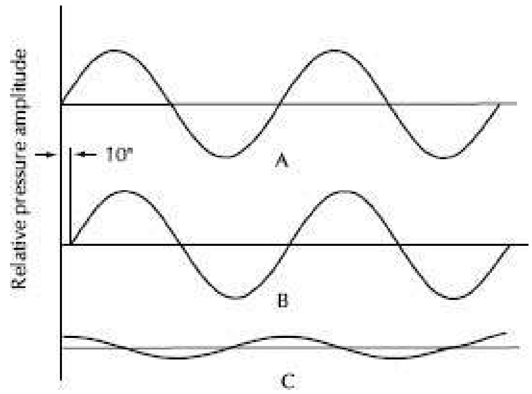


Figure 20-23. Differential pressure at low frequencies on unidirectional microphones.

To obtain good low-frequency response, a reasonable low-frequency electrical output is required from a unidirectional microphone. To accomplish this, the diaphragm must move more easily for a given sound pressure. Some of this is accomplished by reducing the damping resistance to less than one-tenth used in an omnidirectional microphone. This reduction in damping increases the motion of the mechanical resonant frequency of the diaphragm and voice coil, around 150Hz in Fig. 20-22, making the microphone much more acceptable to structure-borne vibrations. Since the diaphragm of an omnidirectional microphone is much more heavily damped, it will respond less to inertial or mechanical vibration forces.

To eliminate unwanted external low-frequency noise from effecting a unidirectional microphone, some kind of isolation such as a microphone shock mount is required to prevent the microphone cartridge from experiencing mechanical shock and vibration.

20.4.3.2 Frequency Response

Frequency response is an important specification of unidirectional microphones and must be carefully analyzed and interpreted in terms of the way the microphone is to be used. If a judgment as to the sound quality of the microphone is made strictly from a single on-axis response, the influence of the proximity effect and off-axis response would probably be overlooked. A comparison of frequency response as a function of microphone-to-source distance will reveal that *all* unidirectional microphones experience a certain amount of proximity effect. In order to evaluate a microphone, this variation with distance is important.

When using a unidirectional microphone[3] in a handheld or stand-mounted configuration, it is conceivable that the performer will not always remain exactly on axis. Variations of ±45° often occur, and so a knowledge of the uniformity of response over such a range is important. The nature of these response variations is shown in Fig. 20-24. Response curves such as these give a better indication of this type of off-axis performance than polar response curves. The polar response curves are limited in that they are usually given for only a few frequencies, therefore, the complete spectrum is difficult to visualize.

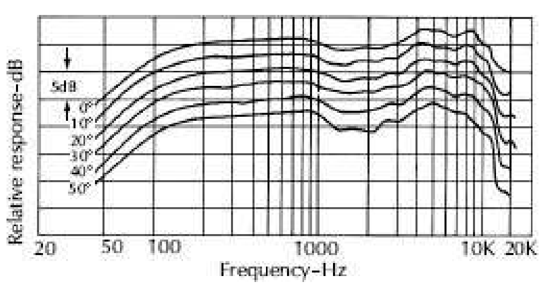


Figure 20-24. Variations in front response versus angular position. Note: Curves have been displaced by 2.5dB for comparison purposes.

For applications involving feedback control or noise rejection, the polar response or particular off-axis response curves, such as at 135° or 180°, are important. These curves can often be misleading due to the acoustic conditions and excitation signals used. Such measurements are usually made under anechoic conditions at various distances with sine-wave excitation. Looking solely at a rear response curve as a function of frequency is misleading since such a curve does not indicate the polar characteristic at any particular frequency, but only the level at one angle. Such curves also tend to give the impression of a rapidly fluctuating high-frequency discrimination. This sort of performance is to be expected since it is virtually impossible to design a microphone of practical size with a constant angle of best discrimination at high frequencies, Fig. 20-25. The principal factor influencing this variation in rear response is diffraction, which is caused by the physical presence of the microphone in the sound field. This diffraction effect is frequency dependent and tends to disrupt the ideal performance of the unidirectional phase-shift elements.

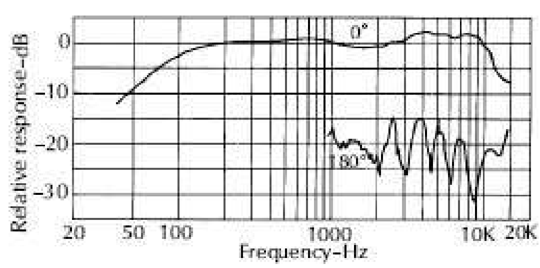


Figure 20-25. Typical fluctuations in high-frequency rear response for a cardioid microphone. Courtesy Shure Incorporated.

To properly represent this high-frequency off-axis performance, a polar response curve is of value, but it, too, can be confusing at high frequencies. As can be seen in Fig. 20-26, where two polar response curves only 20Hz apart are shown. The question that arises then is how can such performance be properly analyzed? A possible solution is to run polar response curves with bands of random noise such as 1/3-octaves of pink noise. Random noise is useful because of its averaging ability and because its amplitude distribution closely resembles program material.

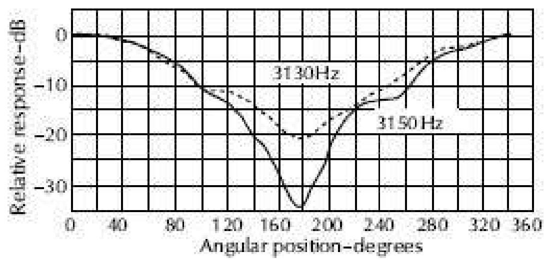


Figure 20-26. An example of rapid variations in high-frequency polar response for single-frequency excitation. Courtesy Shure Incorporated.

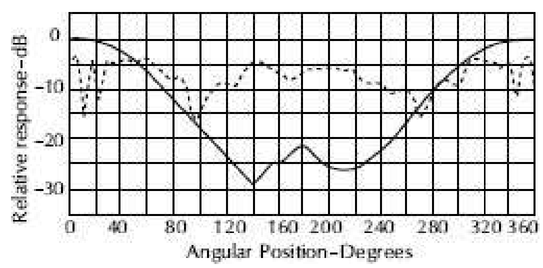


Figure 20-27. An example of a head obstacle on a polar response. Courtesy Shure Incorporated.

Anechoic measurements are only meaningful as long as no large objects are in close proximity to the microphone. The presence of the human head in front of a microphone will seriously degrade the effective high-frequency discrimination. An example of such degradation can be seen in Fig. 20-27 where a head object was placed 2in (5cm) in front of the microphone. (The two curves have not been normalized.) This sort of performance results from the head as a reflector and is a common cause of feedback as one approaches a microphone. This should not be considered as a shortcoming of the microphone, but rather as an unavoidable result of the sound field in which it is being used. At 180°, for example, the microphone sees, in addition to the source it is trying to reject, a reflection of that source 2 in (5cm) in front of its diaphragm. This phenomenon is greatly reduced at low frequencies because the head is no longer an appreciable obstacle to the sound field. It is thus clear that the effective discrimination of any unidirectional microphone is greatly influenced by the sound field in which it is used.

20.5 Cardioid Microphones

Cardioid microphones are named by the way sound enters the rear cavity. The sound normally enters the rear of the microphone's cavity through single or multiple holes in the microphone housing, Fig. 20-28.

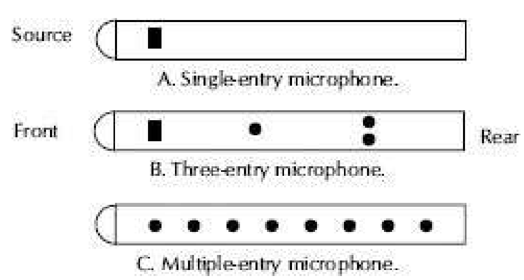


Figure 20-28. Three types of cardioid microphones.

20.5.1 Single-Entry Cardioid Microphones

All single-entrant cardioid microphones have the rear entrance port located at one distance from the rear of the diaphragm. The port location is usually within 1½in (3.8cm) of the diaphragm and can cause a large proximity effect. The Electro-Voice DS35 is an example of a single-entrant cardioid microphone, Fig. 20-29.

Figure 20-29. Electro-Voice DS35 single-entrant microphone. Courtesy Electro-Voice, Inc.

The low-frequency response of the DS35 varies as the distance from the sound source to the microphone decreases, Fig. 20-30. Maximum bass response is produced in close-up use with the microphone 1½ in (3.8cm) from the sound source. Minimum bass response is experienced at distances greater than 2ft (0.6m). Useful effects can be created by imaginative application of the variable low-frequency response.

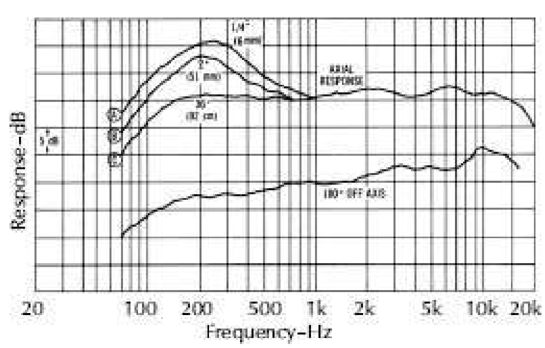


Figure 20-30. Frequency response versus distance for an Electro-Voice DS35 single-entrant cardioid microphone. Courtesy Electro-Voice, Inc.

Another single-entrant microphone is the Shure SM-81.[4] The acoustical system of the microphone operates as a first-order gradient microphone with two sound openings. Fig. 20-31 shows a simplified cross-sectional view of the transducer, and Fig. 20-32 indicates the corresponding electrical analog circuit of the transducer and preamplifier.

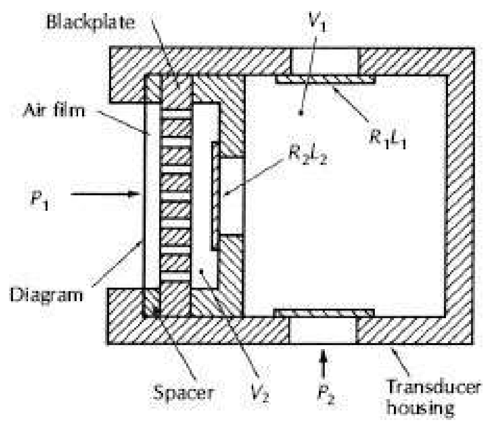


Figure 20-31. Simplified cross-sectional view of the Shure SM81 condenser transducer. Courtesy Shure Incorporated.

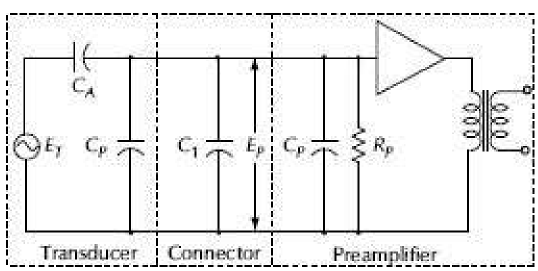


Figure 20-32. Electrical equivalent circuit of the Shure SM81 condenser transducer and preamplifier. Courtesy Shure Incorporated.

Referring to Fig. 20-31, one sound opening, which is exposed to the sound pressure p_1, is represented by the front surface of the diaphragm. The other sound opening, or rear entry, consists of a number of windows in the side of the transducer housing where the sound pressure p_2 prevails. The diaphragm has an acoustical impedance Z_0, which also includes the impedance of the thin air film between the diaphragm and backplate. The sound pressure p_2 exerts its influence on the rear surface of the diaphragm via a screen mounted in the side windows of the transducer housing, having a resistance R_1 and inertance L_1, through the cavity V_1 with compliance C_1. A second screen has a resistance R_2 and inertance L_2, through a second cavity V_2 with compliance C_2, and finally through the perforations in the backplate.

The combination of circuit elements L_1, R_1, C_1, L_2, R_2, C_2 form a ladder network with lossy inertances, and is called a *lossy ladder network*. The transfer characteristic of this network enforces a signal delay on the pressure p_2 imparting directional (cardioid) characteristics for low and medium frequencies. At high frequencies the attenuation caused by the network is large, and the resulting pressure arriving at the back of the diaphragm due to p_2 is small. The microphone then operates much like an omnidirectional system under the predominant influence of p_1. At these frequencies directional characteristics are attained by diffraction of the sound around a suitably shaped transducer housing.

A rotary low-frequency response shaping switch allows the user to select between a flat and a 6dB/octave roll-off at 100Hz or an 18dB/octave cutoff at 80Hz. The 100Hz roll-off compensates for the proximity effect associated with a 6in (15cm) source to microphone distance, while the 80Hz cutoff significantly reduces most low-frequency disturbances with minimal effect on voice material. In the flat position the microphone has a 6dB/octave electronic infrasonic roll-off, with −3dB at 10Hz to reduce the effects of inaudible low-frequency disturbances on microphone preamplifier inputs. Attenuation is provided for operation at high sound pressure levels (to 145dB SPL) by means of a rotary capacitive switch.

A final example of a single entry cardioid microphone is the Shure SM57 supercardioid dynamic microphone, Fig. 20-33, which uses a neodymium magnet for hotter output and incorporates an improved shock mount.

Figure 20-33. Shure SM57 dynamic microphone. Courtesy Shure Incorporated.

20.5.2 Three-Entry Cardioid Microphones

The Sennheiser MD441 is an example of a three-entry cardioid microphone, Fig. 20-34. The low-frequency rear entry has a *d* (distance from center of diaphragm to the entry port) of about 2.8in (7cm), the mid-frequency entry *d* is about 2.2in (5.6cm) and the high-frequency entry *d* is about 1.5in (3.8cm), with the transition in frequency occurring between 800Hz and 1kHz. Each entry consists of several holes around the microphone case rather than a single hole.

Figure 20-34. Sennheiser MD441 three-entry cardioid microphone. Courtesy Sennheiser Electronic Corporation.

This configuration is used for three reasons. By using a multiple arrangement of entry holes around the circumference of the microphone case into the low-frequency system, optimum front response and polar performance can be maintained, even though most of the entries may be accidentally covered when the microphone is handheld or stand mounted. The microphone has good proximity performance because the low-frequency entry ports are far from the diaphragm (4.75in) as well as the high-frequency entry having very little proximity influence at low frequencies. The three-entry configuration has a cardioid polar response pattern that provides a wide front working angle as well as excellent noise rejection and feedback control.

20.5.3 Multiple-Entry Cardioid Microphones

The Electro-Voice RE20 Continuously Variable-D microphone, Fig. 20-35, is an example of multiple-entry microphones. Multiple-entry microphones have many rear entrance ports. They can be constructed as single ports, all at a different distance from the diaphragm, or as a single continuous opening port. Each entrance is tuned to a different band of frequencies, the port closest to the diaphragm being tuned to the high frequencies, and the port farthest from the diaphragm being tuned to the low-frequency band. The greatest advantage of this arrangement is reduced-proximity effect because of the large distance between the source and the rear entry low-frequency port and mechanical crossovers are not as sharp and can be more precise for the frequencies in question.

Figure 20-35. Electro-Voice RE20 multiple-entry (variable-D cardioid microphone). Courtesy Telex Electro-Voice.

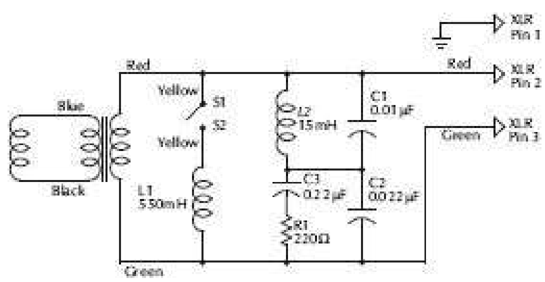


Figure 20-36. Electro-Voice RE20 cardioid wiring diagram. Note "bass tilt" switch circuit. Courtesy Electro-Voice.

As in many cardioid microphones, the RE20 has a low-frequency roll-off switch to reduce the proximity effect when close micing. Fig. 20-36 shows the wiring diagram of the RE20. When the "bass tilt" switch is closed, the low-frequency response will tilt down 4.5dB from 400Hz to 100Hz.

20.5.4 A Four Element Cardioid Microphone

The Audio-Technica AT5040 is a large-diaphragm side-address electret condenser vocal microphone with a cardioid polar pattern. The microphone incorporates a four-part rectangular element, Fig. 20-37. The four matched diaphragms function together (with outputs proprietarily summed) as a single high-performance element providing a combined surface area twice that of a standard one-inch circular diaphragm without the increased weight and decreased transient response that are the expected limitations of a single large diaphragm.

The diaphragms are 2 microns thick, vapor-deposited gold and aged so the characteristics remain constant over years of use. The microphone includes an internal shock mounting that decouples the capsule from the microphone body. For additional isolation, it is provided with a AT8480 shock mount.

The AT5040 was designed as a vocal microphone with extremely smooth top end and controlled sibilance, Fig. 20-38. The large-diaphragm characteristics and fast transient response also make it useful for recording acoustic instruments such as piano, guitar, strings, and saxophone.

Figure 20-37. Audio-Technica AT5040 microphone. Courtesy Audio-Technica U. S., Inc.

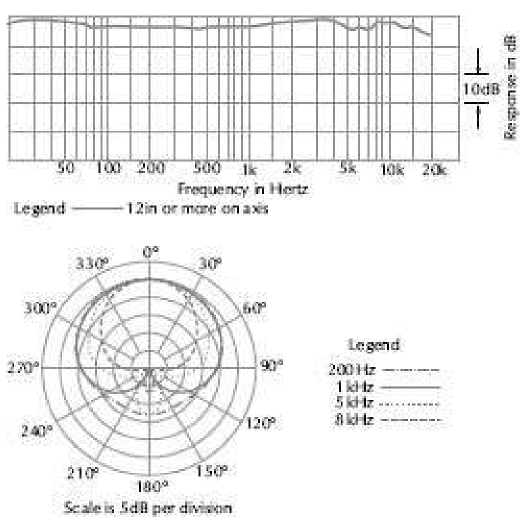


Figure 20-38. Audio-Technica AT5040 polar and frequency response. Courtesy Audio-Technica U. S., Inc.

20.5.5 Dual Capsule Microphone

The Sennheiser MKH 800 TWIN, Fig. 20-39, is a condenser microphone which incorporates a dual capsule consisting of two symmetrical push-pull transducers with high linearity. It is a “side fire" microphone with the two cardioid pick-up patterns of the transducer aligned back-to-back across the axis of the microphone. The signals of both transducers are not combined in the microphone in order to generate differing pick-up patterns but are available separately as two channels at the microphone output allowing the pick-up patterns to be remotely adjusted. The signals can be combined in any desired way in the mixing console in order to create all pick-up patterns from omnidirectional to Figure-8 with an infinite number of intermediate stages.

Figure 20-39. Sennheiser MKH 800 TWIN. Courtesy Sennheiser Electronic Corporation.

The two transducer systems of the KS 80 capsule feature back-to-back pick-up patterns. Both systems are linked acoustically by a common air chamber in the housing. Each system is coupled backwards to the sound field through the other system. In order to prevent the resulting transition to an omnidirectional pick-up pattern that is typical for double diaphragm capsules at low frequencies, both systems are fixed over air gaps on the capsule housing, which has four small openings. The air gaps and openings form an additional frequency dependent lateral acoustic input that stabilizes the cardioid pick-up pattern at low frequencies.

The two signals of the MKH 800 TWIN allow the remote adjustment of the pick-up patterns at the mixing console. The two microphone signals (front and rear) are routed to separate channels and summed together. The sum signal is then distributed over the stereo channels as usual using the pan control. The pan control of both channels has to be aligned identically for correct operation, Figs. 20-40 and 20-41.

Wide cardioid. The wide cardioid pattern is the result of the rear channel having 10dB less amplification than the front channel. The pick-up pattern becomes more omnidirectional at higher amplification and more cardioid at lower amplification.

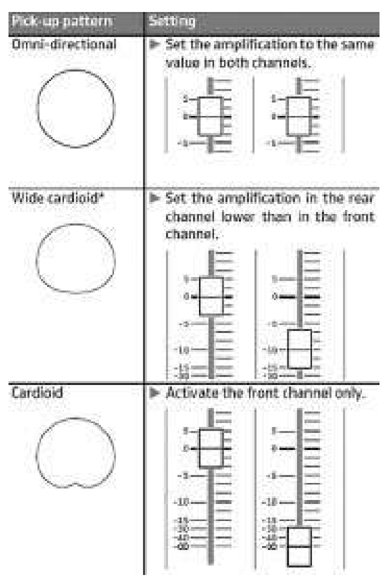


Figure 20-40. Setup for omni to wide cardioid pickup. Courtesy Sennheiser Electronic Corporation.

At the same time, the rear attenuation (180° attenuation) of the microphone changes. It is the direct result of the amplification ratio between the front and the rear channel, i.e., 10dB in the example of the wide cardioid pattern.

Super-cardioid. The super-cardioid pattern is the result of the amplification of the rear channel being 10dB lower than that of the front channel and the phase of the rear channel being inverted. At higher amplification, the pick-up pattern tends towards the figure-8 pattern, otherwise the pattern becomes more cardioid.

The cancellation angle at which the microphone is especially insensitive also changes. It is 180° in the case of the cardioid pattern, 120° for the super-cardioid pattern and 90° for the figure-8 pattern. If the MKH 800 TWIN is used as a supporting microphone, the attenuation between different groups of instruments in an orchestra can be optimized in this way. Here too, the rear attenuation is the result of the amplification ratio between the front and the rear channel, i.e., 10dB in the case of the super-cardioid pattern.

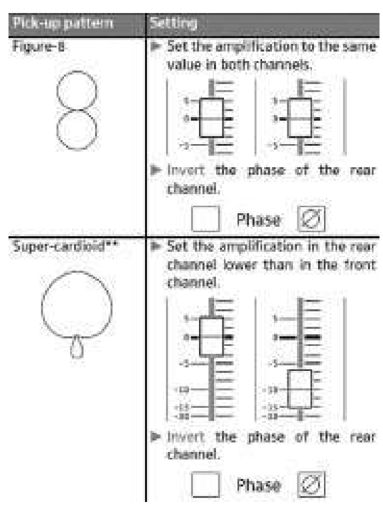


Figure 20-41. Setup for Figure-8 to Supercardioid pickup. Courtesy Sennheiser Electronic Corporation.

Surround applications. As a result of the symmetry of the microphone, it is also possible to create any desired rear pick-up pattern. For this purpose the microphone signals are additionally routed to two other channels, whereby front and rear channel exchange roles. The settings are then made in the same way as described above and the rear pick-up patterns can be freely selected. Both pick-up patterns are then available simultaneously, for example for a surround front channel and a surround rear channel. With two MKH 800 TWIN, four surround channels can be created in this way.

If only the cardioid pick-up pattern is required for the front and rear channels, the microphone signals can also be used directly. This minimizes the necessary effort. With a MKH 800 TWIN and a figure-8 microphone full surround-sound recording using the double MS technique (MSM) can be accomplished. As is common with the MS technique, the figure-8 microphone is positioned above the MKH 800 TWIN and directed to the left. The front and rear left and right surround channels are produced by matrixing the figure-8 signal and the front and rear signals of the MKH 800 TWIN.

By combining the front and rear signals of the MKH 800 TWIN, a center channel with any pick-up pattern can be achieved and even a center rear channel is possible. In spite of the wide range of options, postproduction still only requires the original microphone signals to be saved, i.e., only three signals for five or six surround channels.

20.6 Digital Microphones

Despite continuing improvements, integrated circuits available on the market still constitute a limiting factor in the conversion of audio data from analog to digital form.

For instance, the best delta-sigma A/D converters currently available as integrated circuits provide a dynamic range of 115–120dB (A-weighted), for a theoretical word length of 24 bits.

Compare this to a high-quality analog condenser microphone's dynamic range of up to 130dB. In order to prevent the addition of noise to the signal, a significantly improved A/D converter is required. The conversion process must simultaneously be optimized for the signal levels and source impedance in the microphone.

When A/D conversion is carried out in a mixing console or other equipment, as a rule, loss of signal quality is to be expected, since the conversion occurs after level matching has already been performed. As a result, the dynamic range is affected by headroom considerations and by the characteristics of the microphone preamplifier and the A/D converter.

Digital microphone technology digitizes the capsule signal directly in the microphone, permitting level matching and other processing steps to be performed digitally, preserving the signal quality that is generated by the microphone.

Neumann Solution-D Digital Microphone System

The Solution-D digital microphone system is comprised of the following components:

- A digital microphone, e.g., the D-01 large-diaphragm microphone or the KM D small diaphragm microphone.
- The DMI-2 or DMI-8 interface.
- The RCS remote control software which operates and remotely controls the microphone.

The signal and data transmission of the microphone conform to the AES42 standard which includes the balanced digital microphone output signal, the phantom power supply of 10V, and a remote data stream that contains a signal for synchronizing the microphone with a master clock. A standard balanced 3-pin XLR cable is used for transmission of the signal. The AES42 signal carries data in both directions, Fig. 20-42. The AES3 signal data from the microphone can include information about the microphone that is connected, for instance, make, model number, and serial number. 10V phantom power and remote control data for adjusting the microphone is sent to the microphone. This can include polar pattern adjustment, low-cut filter, pre-attenuation, gain adjustment, and peak limiter.

Figure 20-42. Basic elements of AES42.

The D-01 microphone appears similar to an analog microphone, Fig. 20-43, however, an A/D converter is optimized for capsule signal conditions receives the output signal directly from the capsule. The signal is converted into a digital signal, generating an internal 28-bit signal with a dynamic range of more than 130dB, A-weighted, including capsule characteristics. The digital signal is next processed in the microphone, Fig. 20-44. Parameters such as the directional characteristics, pre-attenuation, low-cut filter, gain, and various switching functions can be set digitally and controlled remotely. External components such as analog preamplifiers and A/D converters are no longer required.

Figure 20-43. Neumann D-01 digital microphone. Courtesy Sennheiser Electronic Corporation.

The AES42 standard describes two modes of synchronizing the microphone with the receiver, e.g., a mixing console or DMI-2 digital interface.

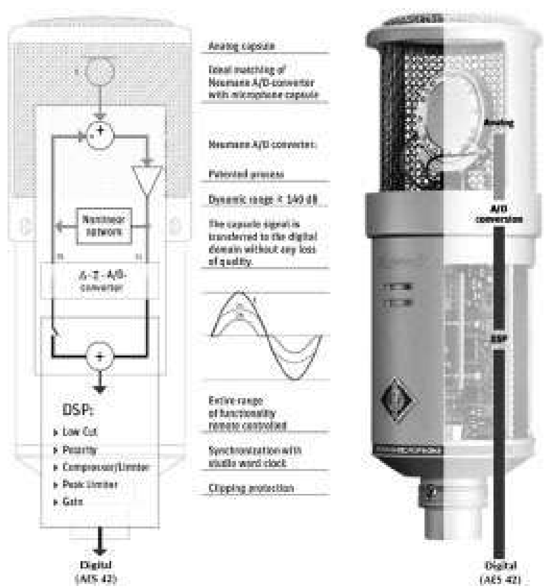


Figure 20-44. Neumann D-01 internal circuitry. Courtesy

Sennheiser Electronic Corporation.

Mode 1: The microphone operates asynchronously, using the sampling rate of its internal quartz oscillator. In this case, a sample rate converter is required at the receiver. However, this mode of operation should be used only if mode 2 synchronization is not possible, since conventional sample rate converters will impair quality, for instance in terms of dynamic range and latency time.

Mode 2: The microphone is synchronized with a master word clock, which can be either an external word clock or the internal word clock of the AES42 receiver (Neumann DMI-2). A control signal is generated that is transmitted via the remote control data stream to the microphone, where it controls the frequency of the internal quartz oscillator.

Equipment that supports AES42 standard can process the output signals of the Solution D microphones directly. In all other cases, a separate two channel device that converts AES42 data format from the microphone to an AES/EBU signal, the DMI-2, Fig. 20-45, or DMI-8 digital microphone interface is used. The DMI-2 is operated via the Neumann RCS remote control software, which is installed on a computer. The computer is connects to the DMI-2 via USB port and USB to RS 485 interface converter. If a large number of microphones are used, several DMI-2's can be cascaded.

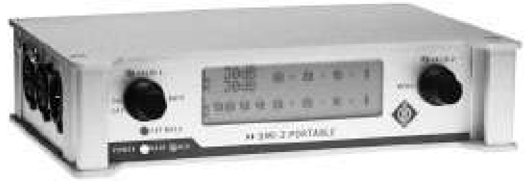


Figure 20-45. Neumann DMI-2 portable controller. Courtesy Sennheiser Electronic Corporation.

In addition to a word clock input and output, the DMI-2 also has an internal work clock generator. If no master clock signal, e.g., from a mixing console is present at the input, the DMI-2 internal word clock is used automatically to synchronize the two microphone channels.

Control of the digital microphones is accomplished via the RCS remote control software which runs as an independent program on a computer. All-important parameters are displayed on the screen and can be changed anytime. During production the audio engineer can monitor the operating status and parameters of all connected microphones, and make changes as needed, Fig. 20-46.

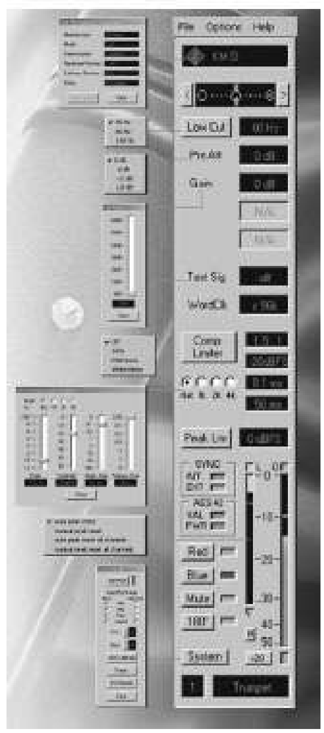


Figure 20-46. Neumann RCS microphone control software. Courtesy Sennheiser Electronic Corporation.

20.7 Carbon Microphones

One of the earliest types of microphones, the *carbon microphone*, is still found in old telephone handsets. It had very limited frequency response, was very noisy, had high distortion, and required a hefty dc power supply. A carbon microphone,[5] Fig. 20-47, operates in the following manner.

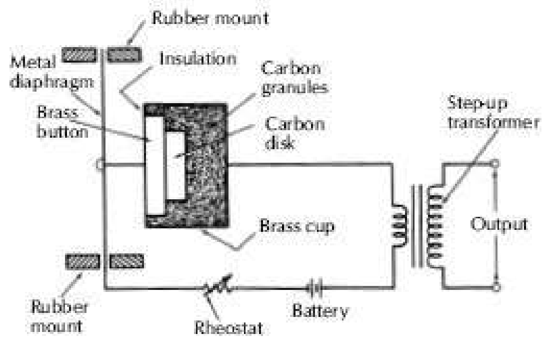


Figure 20-47. Connection and construction of a single-button carbon microphone.

Several hundred small carbon granules are held in close contact in a brass cup called a *button* that is attached to the center of a metallic diaphragm. Sound waves striking the surface of the diaphragm disturb the carbon granules, changing the contact resistance between their surfaces. A battery or dc power source is connected in series with the carbon button and the primary of an audio impedance-matching transformer. The change in contact resistance causes the current from the power supply to vary in amplitude, resulting in a current waveform similar to the acoustic waveform striking the diaphragm. The impedance of the carbon button is low so a step-up transformer is used to increase the impedance and voltage output of the microphone and to eliminate dc from the output circuit.

A modern carbon microphone is the handheld Shure 104C, Fig. 20-48. It is extremely rugged and recommended for mobile and fixed stations for police, fire, etc. The frequency response is limited to the voice range, Fig. 20-49. The output impedance is 50Ω and it draws 50mA of current. The output sensitivity is 5dB below 1V for a 100microbar signal.

20.8 Crystal and Ceramic Microphones

Figure 20-48. Shure model 104C carbon microphone. Courtesy Shure Incorporated.

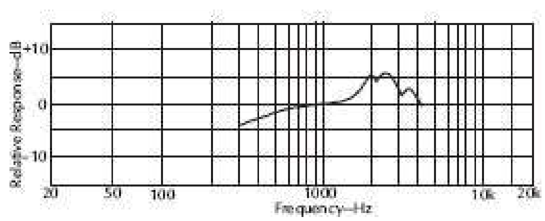


Figure 20-49. Frequency response of the Shure model 104C carbon microphone. Courtesy Shure Incorporated.

Crystal and ceramic microphones[6] operate as follows: piezoelectricity is "pressure electricity" and is a property of certain crystals such as Rochelle salt, tourmaline, barium titanate, and quartz. When pressure is applied to these crystals, electricity is generated. Present-day commercial materials such as ammonium dihydrogen phosphate (ADP), lithium sulfate (LN), dipotassium tartrate (DKT), potassium dihydrogen phosphate (KDP), lead zirconate, and lead titanate (PZT) have been developed for their piezoelectric qualities. Ceramics do not have piezoelectric characteristics in their original state, but the characteristics are introduced in the materials by a polarizing process. In piezoelectric ceramic materials the direction of the electrical and mechanical axes depends on the direction of the original dc polarizing potential. During polarization a ceramic element experiences a permanent increase in dimensions between the poling electrodes and a permanent decrease in dimension parallel to the electrodes.

The crystal element can be cut as a bender element that is only affected by a bending motion or as a twister element that is only affected by a twisting motion, Fig. 20-50.

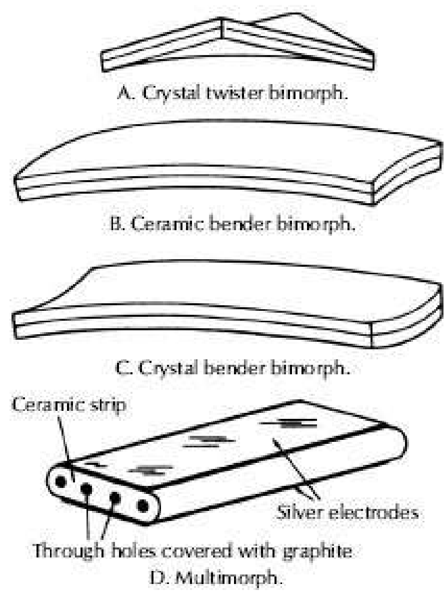


Figure 20-50. Curvatures of bimorphs and multimorph. Courtesy Clevite Corp., Piezoelectric Division.

The internal capacitance of a crystal microphone is about 0.03μF for the diaphragm-actuated type and 0.0005–0.015μF for the sound-cell type.

The *ceramic* microphone operates like a crystal microphone except it employs a barium titanate slab in the form of a ceramic, giving it better temperature and humidity characteristics.

Crystal and ceramic microphones normally have a frequency response from 80 to 6500Hz but can be made to have a flat response to 16kHz. Their output impedance is about 100kΩ, and they require a minimum load of 1-5MΩ to produce a level of about −30dB re 1V/Pa.

20.9 Dynamic Microphones

The *dynamic microphone* is also referred to as a *pressure* or *moving-coil* microphone. It employs a small diaphragm and a voice coil, moving in a permanent magnetic field. Sound waves striking the surface of the diaphragm cause the coil to move in the magnetic field, generating a voltage proportional to the sound pressure at the surface of the diaphragm.

In a dynamic pressure unit, Fig. 20-51, the magnet and its associated parts (magnetic return, pole piece, and pole plate) produce a concentrated magnetic flux of approximately 10,000G in the small gap.

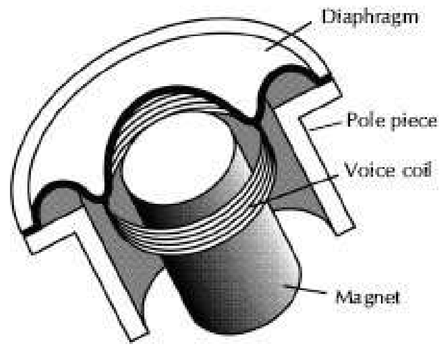


Figure 20-51. A simplified drawing of a dynamic microphone. Courtesy Shure Incorporated.

The diaphragm, a key item in the performance of a microphone, supports the voice coil centrally in the magnetic gap, with only 0.006in clearance.

An omnidirectional diaphragm and voice-coil assembly is shown in Fig. 20-52. The compliance section has two hinge points with the section between them made up of tangential corrugated triangular sections that stiffen this portion and allow the diaphragm to move in and out with a slight rotating motion. The hinge points are designed to permit high-compliance action. A spacer supports the moving part of the diaphragm away from the top pole plate to provide room for its movement. The cementing flat is bonded to the face plate. A stiff hemispherical dome is designed to provide adequate acoustical capacitance. The coil seat is a small step where the voice coil is mounted, centered, and bonded on the diaphragm.

Mylar™, a polyester film manufactured by the DuPont Company, is commonly used for diaphragms. Mylar is a unique plastic. Extremely tough, it has high tensile strength, high resistance to wear, and outstanding flex life. Since Mylar™ is extremely stable, its properties do not change within the temperature and humidity range in which microphones are used.

The specific gravity of Mylar™ is approximately 1.3 as compared to 2.7 for aluminum so a Mylar™ diaphragm may be made considerably thicker without upsetting the relationship of the diaphragm mass to the voice-coil mass.

The voice coil weighs more than the diaphragm so it is the controlling part of the mass in the diaphragm voice-coil assembly. The voice coil and diaphragm mass (analogous to inductance in an electrical circuit) and compliance (analogous to capacitance), make the assembly resonate at a given frequency as any tuned electrical circuit. The free-cone resonance of a typical undamped unit is in the region of 350Hz.

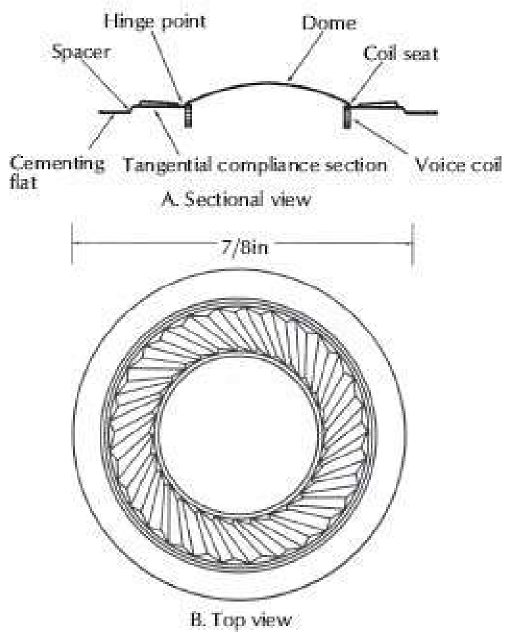


Figure 20-52. Omnidirectional diaphragm and voice coil assembly.

If the voice coil were left undamped, the response of the assembly would peak at 350Hz, Fig. 20-53. The resonant characteristic is damped out by the use of an acoustic resistor, a felt ring that covers the openings in the centering ring behind the diaphragm. This is analogous to electrical resistance in a tuned circuit. While this reduces the peak at 350Hz, it does not fix the droop below 200Hz. Additional acoustical resonant devices are used inside the microphone case to correct the drooping. A cavity behind the unit (analogous to capacitance) helps resonate at the low frequencies with the mass (inductance) of the diaphragm and voice-coil assembly.

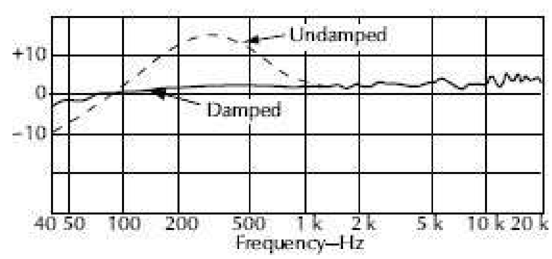


Figure 20-53. Diaphragm and voice-coil assembly response curve.

Another tuned resonant circuit is added to extend the response down to 35Hz. This circuit, tuned to about 50Hz, is often a tube that couples the inside cavity of the microphone housing to the outside.

The curvature of the diaphragm dome provides stiffness, and the air cavity between it and the dome of the pole piece form an acoustic capacitance. This capacitance resonates with the mass (inductance) of the assembly to extend the response up to 20kHz.

Fig. 20-54 illustrates the effect of a varying sound pressure on a moving-coil microphone. For this simplified explanation, assume that a massless diaphragm voice-coil assembly is used. The acoustic waveform, Fig. 20-54A, is one cycle of an acoustic waveform, where *a* indicates atmospheric pressure *AT*; and *b* represents atmospheric pressure plus a slight overpressure increment Δ or *AT* + Δ.

The electrical waveform output from the moving-coil microphone, Fig. 20-54B, does not follow the phase of the acoustic waveform because at maximum pressure, $AT + \Delta$ or b, the diaphragm is at rest (no velocity). Further, the diaphragm and its attached coil reach maximum velocity, hence maximum electrical amplitude at point c on the acoustic waveform. This is of no consequence unless another microphone is being used along with the moving-coil microphone where the other microphone does not see the same 90° displacement. Due to this phase displacement, condenser microphones should not be mixed with moving-coil or ribbon microphones when micing the same source at the same distance. (Sound pressure can be proportional to velocity in many practical cases.)[7]

A steady overpressure which can be considered an acoustic square wave, Fig. 20-54C, would result in the output shown in Fig. 20-54D. As the acoustic pressure rises from a to b, it has velocity creating a voltage output from the microphone. Once the diaphragm reaches its maximum displacement at b, and stays there during the time interval represented by the distance between b and c, voice-coil velocity is zero so electrical output voltage ceases and the output returns to zero. The same situation repeats itself from c to e and from e to f on the acoustic waveform. As can be seen, a moving-coil microphone cannot reproduce a square wave.

Another interesting theoretical consideration of the moving-coil microphone mechanism is shown in Fig. 20-55. Assume a sudden transient condition. Starting at a on the acoustic waveform, the normal atmospheric pressure is suddenly increased by the first wavefront of a new signal and proceeds to the first overpressure peak, $AT + \Delta$ or b. The diaphragm will reach a maximum velocity halfway to b and then return to zero velocity at b. This will result in a peak, a', in the electrical waveform. From b on, the acoustic waveform and the electrical waveform will proceed as before, cycle for cycle, but 90° apart.

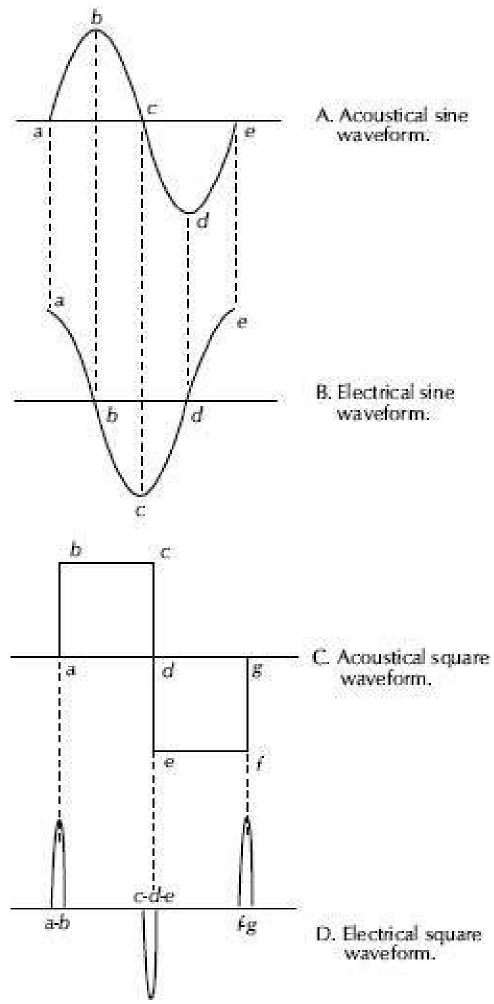


Figure 20-54. Effect of a varying sound pressure on a moving-coil microphone.

In this special case, peak a' does not follow the input precisely so it is something extra. It will probably be swamped out by other problems (especially mass) encountered in a practical moving-coil microphone. It does illustrate that even with a "perfect," massless, moving-coil microphone, "perfect" electrical waveforms will not be produced.

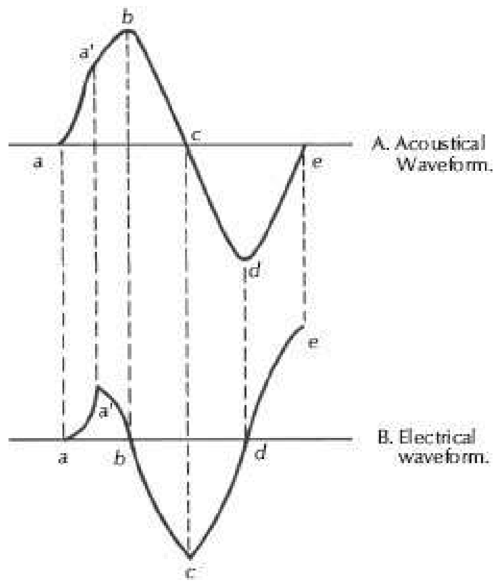


Figure 20-55. Effect of a transient condition on a moving-coil microphone.

When sound waves vibrate the diaphragm, the voice coil has a voltage induced in it proportional to the magnitude and at the frequency of the vibrations. The voice coil and diaphragm have some finite mass and any mass has inertia that causes it to want to stay in the condition it is in—namely, in motion or at rest. If the stationary part of the diaphragm-magnet structure is moved in space, the inertia of the diaphragm and coil causes them to try to remain fixed in space. Therefore, there will be relative motion between the two parts with a resultant electrical output. An electrical output can be obtained in two ways, by motion of the diaphragm from airborne acoustical energy or by motion of the magnet circuit by structure-borne vibration. The diaphragm motion is the desired output, while the structure-borne vibration is undesired.

To reduce the undesired output the mass of the diaphragm and voice coil may be reduced, but there are practical limits, or the frequency response may be limited mechanically with stiffer diaphragms or electronically with filter circuits. However limited response makes the microphone unsuitable for broad-range applications.

20.10 Capacitor Microphones

In a *capacitor* or *condenser* microphone, Fig. 20-56, the sound pressure level varies the head capacitance of the microphone by deflecting one or two plates of the capacitor, causing an electrical signal that varies with the acoustical signal. The varying capacitance can be used to modulate an RF signal that is later demodulated or can be used as one leg of a voltage divider, Fig. 20-57, where *R* and *C* form the voltage divider of the power supply ++ to –.

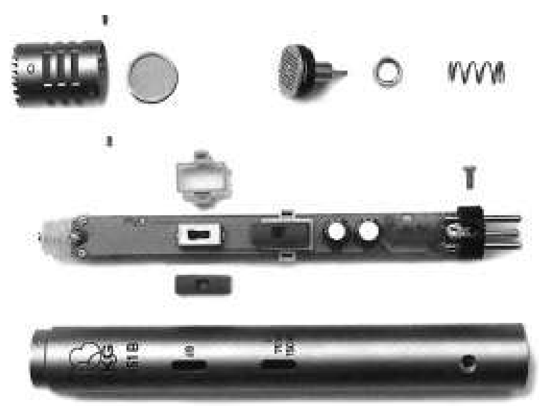


Figure 20-56. AKG C451 microphone. Courtesy AKG Acoustics.

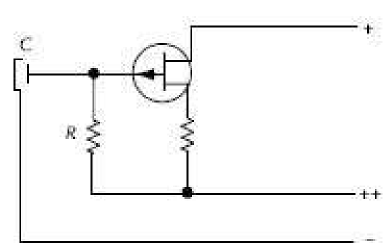


Figure 20-57. Voltage divider type of capacitor microphone.

The head of most capacitor microphones consists of a small two-plate 40–50pF capacitor. One of the two plates is a stretched diaphragm; the other is a heavy back plate or center terminal, Fig. 20-57. The back plate is insulated from the diaphragm and spaced approximately 1mil (0.001in) from, and parallel to, the rear surface of the diaphragm. Mathematically the output from the head may be calculated as

$$E_O = \frac{E_p a^2 P}{8dt} \qquad (20\text{-}18)$$

where,
E_O is the output in V,
E_p is the dc polarizing voltage in V,
a is the radius of active area of the diaphragm in cm,
P is the pressure in dyn/cm^2,
d is the spacing between the back plate and diaphragm in cm,
t is the diaphragm tension in dyn/cm.

Some inexpensive or older capacitor microphones operate with an equivalent noise level of 15–30dB SPL. Although a 20–30dB SPL is in the range of a well-constructed studio, a 20–30dB microphone equivalent noise is not masked by room noise because room noise occurs primarily at low frequencies and microphone noise at high frequency as hiss.

A feature of digital audio is the enlarged dynamic range and reduced noise floor. Unfortunately, due to this improvement, the inherent noise of the microphones may become audible, because it is no longer covered up by the noise of the recording medium.

The capacitor microphone has a much faster rise time than the dynamic microphone because of the significantly lower mass of the moving parts (diaphragm versus diaphragm/coil assembly). The capacitor rise time rises from 10% of its rise time to 90% in approximately 15μs, while the rise time for the dynamic microphone is in the order of 40μs.

Capacitor microphones generate an output electrical waveform in step or phase with the acoustical waveform and can be adapted to measure essentially dc overpressures, Fig. 20-58. To produce a true square wave, the microphone would have to be a radio-frequency capacitor microphone. A dc voltage divider type would have droop in the top and bottom of the square wave.

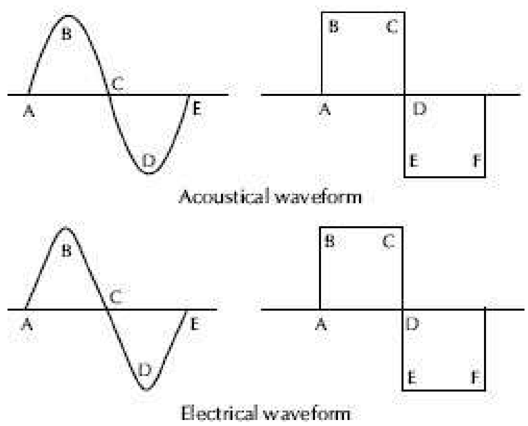


Figure 20-58. Capacitor microphone acoustic wave and electrical signals. Note the in-phase condition.

Some advantages of capacitor microphones are:

- Small, low-mass rigid diaphragms that reduce vibration pickup.
- Smooth, extended-range frequency response.
- Rugged—capable of measuring very high sound pressure levels (rocket launches).
- Low noise (which is partially canceled by the need for electronics).

- Small head size, which provides low diffraction interference.

20.10.1 Voltage Divider Capacitor Microphone

Voltage divider-type capacitor microphones require a preamplifier as an integral part of the housing and a source of polarizing voltage for the head, and a source of power.

A high-quality capacitor microphone, the Sennheiser K6 Modular Condenser Microphone Series is suitable for recording studios, television and radio broadcast, motion picture studios, and stage and concert hall applications, as well as high-quality commercial sound installations.

The K6/ME62 series is a *capacitor microphone system*, Fig. 20-59, that uses AF circuitry with field-effect transistors with a low noise level (15dB per DIN IEC 651), and lifelong stability. Low current consumption at low voltage and phantom circuit powering permit feeding the microphone supply voltage via a standard two-conductor shielded audio cable or an internal AA battery.

Figure 20-59. Modular microphone system with an omnidirectional cartridge utilizing a voltage divider circuit. Courtesy Sennheiser Electronic Corporation.

The K6 offers interchangeable capsules, allowing the selection of different response characteristics from omnidirectional to cardioid to shotgun to adapt the microphone to various types of environments and recording applications.

Because of the new PCM recorders, signal-to-noise ratio (*SNR*) has reached a level of 90dB, requiring capacitor microphones to increase their *SNR* level to match the recorder. The shotgun K6 series microphone, Fig. 20-60, has an equivalent noise level of 16dB (DIN IEC 651).

Figure 20-60. The same microphone shown in Fig. 20-59 with a shotgun cartridge. Courtesy Sennheiser Electronic Corporation.

As in most circuitry, the input stage of a voltage divider-type capacitor microphone has the most effect on noise, Fig. 20-61. It is important that the voltage on the transducer does not change. This is normally accomplished by controlling the input current. In the circuit of Fig. 20-61, the voltages V_{in}, V_o, and V_D are within 0.1% of each other. Noise, which might come into the circuit as V_{in} through the operational amplifier, is only 1.1% of the voltage V_o.

Preattenuation, i.e., attenuation between the capacitor and the amplifier, can be achieved by connecting parallel capacitors to the input, by reducing the input stage gain by means of capacitors in the negative feedback circuit, or by reducing the polarizing voltage to one-third its normal value and by using a resistive voltage divider in the audio line.

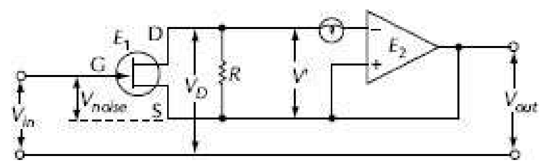


Figure 20-61. Simplified schematic of a microphone input circuit. Courtesy AKG Acoustics.

20.10.2 Phantom Power for Capacitor Microphones

A common way to supply power for capacitor microphones is with a phantom circuit. Phantom or simplex powering is supplying power to the microphone from the input of the following device such as a preamplifier, mixer, or console.

Most capacitor microphone preamplifiers can operate on any voltage between 9Vdc and 52Vdc because they incorporate an internal voltage regulator. The preamplifier supplies the proper polarizing voltage for the capacitor capsule plus impedance matches the capsule to the balanced low-impedance output.

Standard low-impedance, balanced microphone input receptacles are easily modified to simplex both operating voltage and audio output signal, offering the following advantages in reduced cost and ease of capacitor microphone operation:

- Special external power supplies and separate multi-conductor cables formerly required with capacitor microphones are eliminated.
- The B+ supply in associated recorders, audio consoles, and commercial sound amplifiers can be used to power the microphone directly.
- Dynamic, ribbon, and capacitor microphones can be used interchangeably on standard, low-impedance, balanced microphone circuits.
- Dynamic, ribbon, and self-powered capacitor microphones may be connected to the modified amplifier input without defeating the microphone operating voltage.
- Any recording, broadcast, and commercial installation can be inexpensively upgraded to capacitor microphone operation using existing, two-conductor microphone cables and electronics.

Phantom circuit use requires only that the microphone operating voltage be applied equally to pins 2 and 3 of the amplifier low-impedance (normally an XLR input) receptacle. Pin 1 remains ground and circuit voltage minus. The polarity of standard microphone cable wiring is not important except for the usual audio polarity requirement (see section 20.5.3). Two equally effective methods of amplifier powering can be used:

1. Connect an amplifier B+ supply of 9–12V directly to the ungrounded center tap of the microphone input transformer, as shown in Fig. 20-62. A series-dropping resistor is required for voltages between 12 and 52V. Fig. 20-63 is a typical resistor value chart. A chart can be made for any microphone if the

current is known for a particular voltage.

2. A two-resistor, artificial center powering circuit is required when the microphone input transformer is not center-tapped, or when input attenuation networks are used across the input transformer primary. Connect a B+ supply of 9–12V directly to the artificial center of two 332Ω, 1% tolerance precision resistors, as shown in Fig. 20-64. Any transformer center tap should not be grounded. For voltages between 12 and 52V, double the chart resistor value of Fig. 20-63.

Any number of capacitor microphones may be powered by either method from a single B+ source according to the current available. Use the largest resistor value shown (R_v *max*) for various voltages in Fig. 20-63 for minimum current consumption.

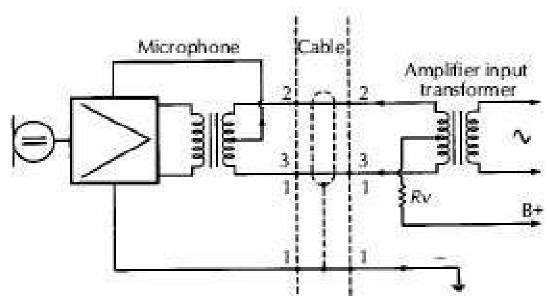


Figure 20-62. Direct center-tap connection method of phantom powering capacitor microphones. Courtesy AKG Acoustics.

20.10.3 Capacitor Radio-Frequency, Frequency-Modulated Microphones

A *frequency-modulated microphone* is a capacitor microphone that is connected to a radio-frequency (RF) oscillator. Pressure waves striking the diaphragm cause variations in the capacity of the microphone head that frequency modulates the oscillator. The output of the modulated oscillator is passed to a discriminator and amplified in the usual manner.

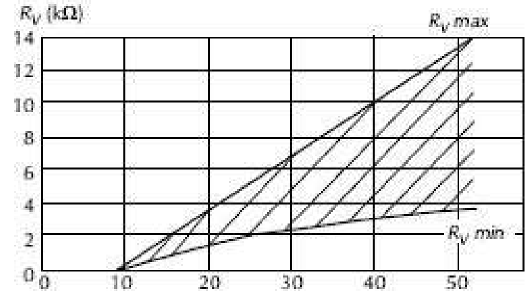


Figure 20-63. Dropping resistor value chart for phantom powering AKG C-451 microphones. Courtesy AKG Acoustics.

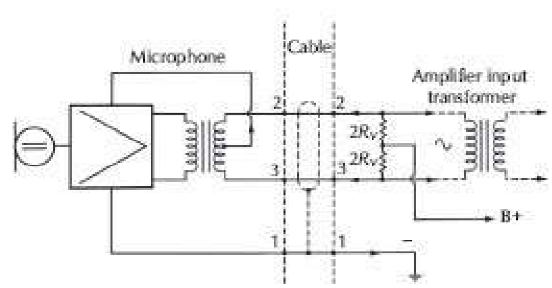


Figure 20-64. Artificial center tap connected method of powering capacitor microphones. Courtesy AKG Acoustics.

The basic circuitry is shown in Fig. 20-65. By means of a single transistor, two oscillatory circuits are excited and tuned to the exact same frequency of 3.7MHz. The output voltage from the circuits is rectified by a phase-bridge detector circuit, which operates over a large linear modulation range with very small RF voltages from the oscillator. The amplitude and polarity of the output voltage from the bridge depend on the phase angle between the two high-frequency voltages. The microphone capsule (head) acts as a variable capacitance in one of the oscillator circuits. When a sound wave impinges on the surface of the diaphragm of the microphone head, the vibrations of the diaphragm are detected by the phase curve of the oscillator circuit, and an audio frequency voltage is developed at the output of the bridge circuit. The microphone-head diaphragm is metal to guarantee a large constant capacitance. An automatic frequency control (afc) with a large range of operation is provided by means of capacitance diodes to preclude any influence caused by aging or temperature changes on the frequency-determining elements, that might throw the circuitry out of balance.

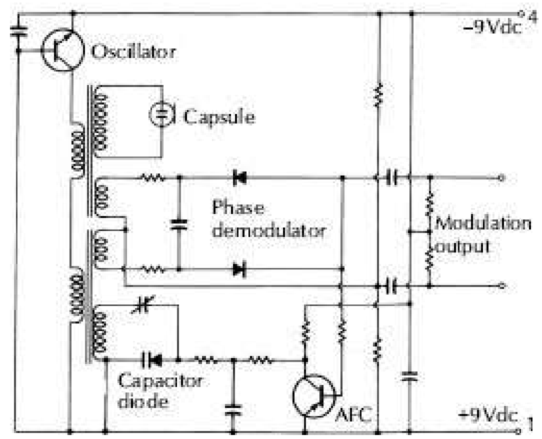


Figure 20-65. Basic circuit for a radio-frequency capacitor microphone.

Internal output resistance is about 200Ω. The signal, fed directly from the bridge circuit through two capacitors, delivers an output level of −51dB to −49dB (depending on the polar pattern used) into a 200Ω load for a sound pressure level of 10dyn/cm^2. The *SNR* and the distortion are independent of the load because of the bridge circuit; therefore, the microphone may be operated into load impedances ranging from 30 to 200Ω.

20.10.4 Symmetrical Push-Pull Transducer Microphone

Investigations on the linearity of condenser microphones customarily used in the recording studios was carried out by

Sennheiser using the *difference frequency method* incorporating a twin tone signal, Fig. 20-66. This is a very reliable test method as the harmonic distortions of both loudspeakers that generate the test sounds separately do not disturb the test result. Thus, difference frequency signals arising at the microphone output are arising from nonlinearities of the microphone itself.

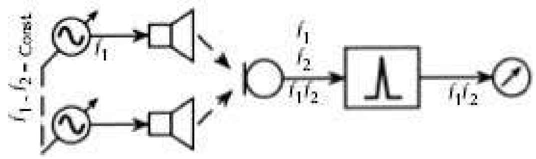


Figure 20-66. Difference frequency test. Courtesy Sennheiser Electronic Corporation.

Fig. 20-67 shows the distortion characteristics of eight unidirectional studio condenser microphones which were stimulated by two sounds of 104dB SPL (3 Pa). The frequency difference was fixed to 70Hz while the twin tone signal was swept through the upper audio range. The curves show that unwanted difference frequency signals of considerable levels were generated by all examined microphones. Although the curves are shaped rather individually, there is a general tendency for increased distortion levels (up to 1% and more) at high frequencies.

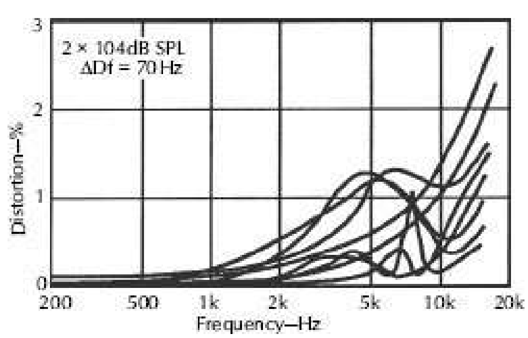


Figure 20-67. Frequency distortion of eight unidirectional

microphones. Courtesy Sennheiser Electronic Corporation.

The measurement results can be extended to higher signal levels simply by linear extrapolation. This means, for instance, that 10 times higher sound pressures will yield 10 times higher distortions, as long as clipping of the microphone circuit is prevented. Thus, two sounds of 124dB SPL will cause more than 10% distortion in the microphones. Sound pressure levels of this order are beyond the threshold of pain of human hearing but may arise at close-up micing. Despite the fact that the audibility of distortions depends significantly on the tonal structure of the sound signals, distortion figures of this order will affect the fidelity of the sound pick-up.

The Cause of Nonlinearity. Fig. 20-68 shows a simplified sketch of a capacitive transducer. The diaphragm and backplate form a capacitor, the capacity of which depends on the width of the air gap. From the acoustical point of view the air gap acts as a complex impedance. This impedance is not constant but depends on the actual position of the diaphragm. Its value is increased if the diaphragm is moved toward the backplate and it is decreased at the opposite movement, so the air gap impedance is varied by the motion of the diaphragm. This implies a parasitic rectifying effect superimposed to the flow of volume velocity through the transducer, resulting in nonlinearity-created distortion.

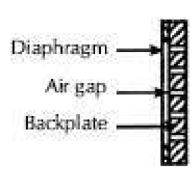


Figure 20-68. Conventional capacitor microphone transducer.

Solving the Linearity Problem. A push-pull design of the transducer helps to improve the linearity of condenser microphones, Fig. 20-69. An additional plate equal to the backplate is positioned symmetrically in front of the diaphragm, so two air gaps are formed with equal acoustical impedances as long as the diaphragm is in its rest position. If the diaphragm is deflected by the sound signal, then both air gap impedances are deviated opposite to each other. The impedance of one side increases while the other impedance decreases. The variation effects compensate each other regardless of the direction of the diaphragm motion, and the total air gap impedance is kept constant, reducing the distortion of a capacitive transducer.

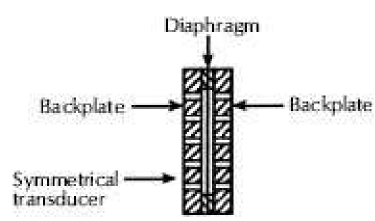


Figure 20-69. Symmetrical capacitor microphone transducer.

Fig. 20-70 shows the distortion characteristics of the Sennheiser MKH series push-pull element transformerless RF condenser microphones. The improvement on linearity due to the push-pull design can be seen by comparing Fig. 20-70 to Fig. 20-67.

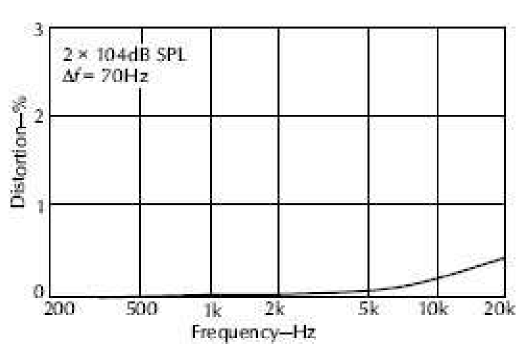


Figure 20-70. Distortion characteristics of the symmetrical

capacitor microphone transducer.

20.10.5 Noise Sources

The inherent noise of condenser microphones is caused partly by the random incidence of the air particles at the diaphragm due to their thermal movement. The laws of statistics imply that sound pressure signals at the diaphragm can be evaluated by a precision that improves linearly with the diameter of the diaphragm. Thus, larger diaphragms yield better noise performance than smaller ones.

Another contribution of noise is the frictional effects in the resistive damping elements of the transducer. The noise generation from acoustical resistors is based on the same principles as the noise caused by electrical resistors so high acoustical damping implies more noise than low damping.

Noise is also added by the electrical circuit of the microphone. This noise contribution depends on the sensitivity of the transducer. High transducer sensitivity reduces the influence of the circuit noise. The inherent noise of the circuit itself depends on the operation principle and on the technical quality of the electrical devices.

Noise Reduction. Large-diameter diaphragms improve noise performance. Unfortunately, a large diameter increases the directivity at high frequencies. A 1in (25mm) transducer diameter is usually a good choice.

A further method to improve the noise characteristics is the reduction of the resistive damping of the transducer. In most directional condenser microphones, a high amount of resistive damping is used in order to realize a flat frequency response of the transducer itself. With this design the electrical circuit of the microphone is rather simple. However, it creates reduced sensitivity and increased noise.

Keeping the resistive damping of the transducer moderate will be a more appropriate method to improve noise performance, however it leads to the transducer frequency response that is not flat so equalization has to be applied by electrical means to produce a flat frequency response of the complete microphone. This design technique requires a more sophisticated electrical circuit but produces good noise performance.

The electrical output of a transducer acts as a pure capacitance. Its impedance decreases as the frequency increases so the transducer impedance is low in an RF circuit but high in an AF circuit. Moreover, in an RF circuit the electrical impedance of the transducer does not depend on the actual audio frequency but is rather constant due to the fixed frequency of the RF oscillator. Contrary to this, in an AF design, the transducer impedance depends on the actual audio frequency, yielding very high values especially at low frequencies. Resistors of extremely high values are needed at the circuit input to prevent loading of the transducer output. These resistors are responsible for additional noise contribution.

The RF circuit features a very low output impedance which is comparable to that of dynamic-type microphones. The output signal can be applied directly to bipolar transistors, yielding low noise performance by impedance matching.

The Sennheiser MKH 20, Fig. 20-71, is a pressure microphone with omnidirectional characteristics. The MKH 30 is a pure pressure-gradient microphone with a highly symmetrical bidirectional pattern due to the symmetry of the push-pull transducer. The MKH 40, Fig. 20-72, operates as a combined pressure and pressure-gradient microphone yielding a unidirectional cardioid pattern.

- The microphones are phantom powered by 48Vdc and 2mA. The outputs are transformerless floated, Fig. 20-73.
- The *SPLmax* is 134dB at nominal sensitivity and 142dB at reduced sensitivity.
- The equivalent *SPL* of the microphones range from 10–12dBA corresponding to CCIR-weighted figures of 20–22dB.
- The directional microphones incorporate a switchable bass roll-off to cancel the proximity effect at close-up micing. The compensation is adjusted to about 5cm (2in) distance.

Figure 20-71. Omnidirectional pressure capacitor microphone. Note the lack of rear entry holes in the case. Courtesy Sennheiser Electronic Corporation.

A special feature of the omnidirectional microphone is a switchable diffuse field correction that corrects for both direct and diffuse sound field conditions. The normal switch position is recommended for a neutral pickup when close-up micing and the diffuse field position is used if larger recording distances are used where reverberation becomes significant.

Figure 20-72. Unidirectional pressure/pressure-gradient capacitor microphone. Courtesy Sennheiser Electronic Corporation.

The distinction between both recording situations arises because omnidirectional microphones tend to attenuate lateral and reverse impinging sound signals at high frequencies. Diffuse sound signals with random incidence cause a lack of treble response, which can be compensated by treble emphasis at the microphone. Unfortunately, frontally impinging sounds are emphasized also, but this effect is negligible if the reverberant sound is dominant.

20.11 Electret Microphones

An *electret microphone* is a capacitor microphone in which the head capacitor is permanently charged, eliminating the need for a high-voltage bias supply.

The Shure SM81 cardioid condenser microphone[3] Fig. 20-74, uses an electret material as a means of establishing a bias voltage on the transducer. The backplate carries the electret material based upon the physical properties of halocarbon materials such as Teflon™ and Aclar, which are excellent electrets, and materials such as polypropylene and polyester terephthalate (Mylar™), which are more suitable for diaphragms.

The operation of the Shure SM-81 microphone is explained in Section 20.5.1.

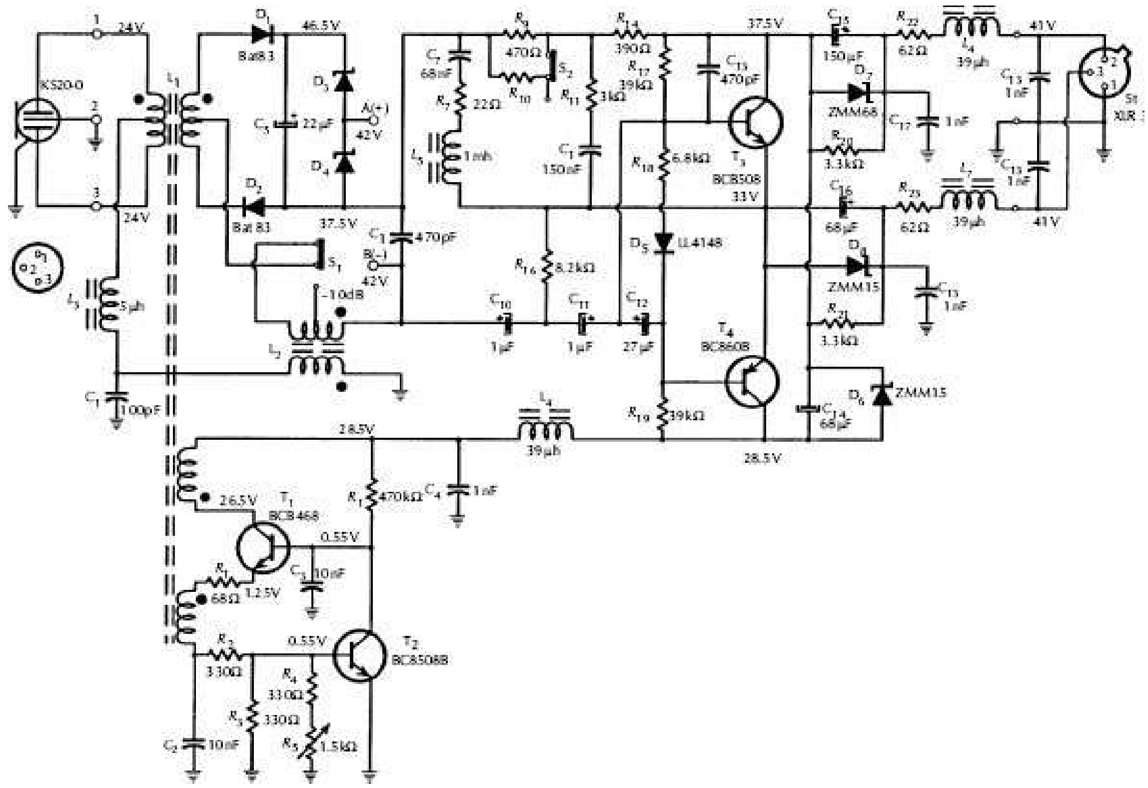


Figure 20-73. Schematic of a Sennheiser MKH 20 P 48 U 3 capacitor microphone. Courtesy Sennheiser Electronic Corporation.

Figure 20-74. Shure SM81 electret capacitor microphone. Courtesy Shure Incorporated.

20.12 Pressure Zone Microphones (PZM)

The *pressure zone microphone*, referred to as a *PZMicrophone* or *PZM*, is a miniature condenser microphone mounted face-down next to a sound-reflecting plate or boundary. The microphone diaphragm is placed in the pressure zone just above the boundary where direct and reflected sounds combine effectively in-phase over the audible range.

In many recording and reinforcement applications, the sound engineer is forced to place microphones near hard reflective surfaces such as when recording an instrument surrounded by reflective baffles, reinforcing drama or opera with the microphones near the stage floor, or recording a piano with the microphone close to the open lid.

In these situations, sound travels from the source to the microphone via two paths: directly from the source to the microphone, and reflected off the surface to the microphone. The delayed sound reflections combine with the direct sound at the microphone, resulting in phase cancellations of various frequencies, see Fig. 20-4. This creates a series of peaks and dips in the frequency response called the *comb-filter effect*, affecting the recorded tone quality and giving an unnatural sound.

The PZM was developed to avoid the tonal coloration caused by microphone placement near a surface. The microphone diaphragm is arranged parallel with and very close to the reflecting surface and facing it, so that the direct and reflected waves combine at the diaphragm in-phase over the audible range, Fig. 20-75.

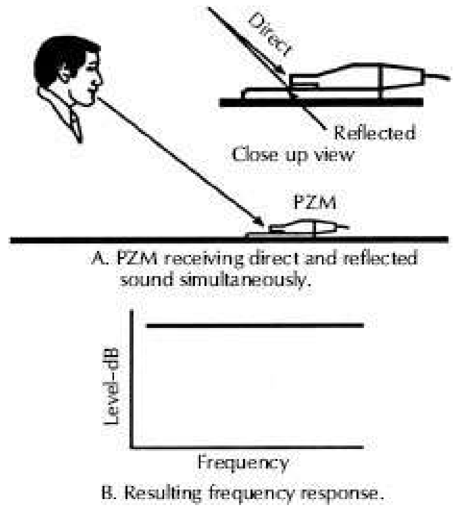


Figure 20-75. Effects of receiving direct and reflected sound simultaneously.

This arrangement can provide several benefits:

- Wide, smooth frequency response (natural reproduction) because of the lack of phase interference between direct and reflected sound.
- A 6dB increase in sensitivity because of the coherent addition of direct and reflected sound.
- High SNR created by the PZM's high sensitivity and low internal noise.
- A 3dB reduction in pickup of the reverberant field compared to a conventional omnidirectional microphone.
- Lack of off-axis coloration as a result of the sound entry's small size and radial symmetry.
- Good-sounding pickup of off-mic instruments due to the lack of off-axis coloration.
- Identical frequency response for random-incidence sound (ambience) and direct sound due to the lack of off-axis coloration.
- Consistent tone quality regardless of sound-source movement or

microphone-to-source distance.

- Excellent reach (clear pickup of quiet distant sounds).
- Hemispherical polar pattern, equal sensitivity to sounds coming from any direction above the surface plane.
- Inconspicuous low-profile mounting.

20.12.1 Sensitivity Effects

If a PZM capsule is placed very near a single large boundary, within 0.020in (0.50mm), such as a large plate, floor, or wall, incoming sound reflects off the surface. The reflected sound wave adds to the incoming sound wave in the Pressure Zone next to the boundary. This coherent addition of sound waves doubles the sound pressure at the microphone, effectively increasing the microphone sensitivity or output by 6dB over a standard microphone.

If the PZM capsule is placed at the junction of two boundaries at right angles to each other, such as the floor and a wall, the wall increases sensitivity 6dB, and the floor increases sensitivity another 6dB, increasing sensitivity 12dB.

With the PZM element at the junction of three boundaries at right angles, such as in the corner of the floor and two walls, microphone sensitivity will be 18dB higher than what it was in open space.

Note that the acoustic sensitivity of the microphone rises as boundaries are added, but the electronic noise of the microphone stays constant, so the effective *SNR* of the microphone improves 6dB every time a boundary is added at right angles to previous boundaries.

20.12.2 Direct-to-Reverberant Ratio Effects

Direct sound sensitivity increases 6dB per boundary added, while reverberant or random-incidence sound increases only 3dB per boundary added. Consequently, the direct-to-reverberant ratio increases 3dB ($6dB_{dir}$ – $3dB_{rev}$) whenever a boundary is added at right angles to previous boundaries.

20.12.3 Frequency-Response Effects

The low-frequency response of the PZM or PCC depends on the size of the surface it is mounted on. The larger the surface, the more the low-frequency response is extended. The low-frequency response shelves down to a level 6dB below the mid-frequency level at the frequency where the wavelength is about 6 times the boundary dimension. For example, the frequency response of a PZM on a 2ft × 2ft (0.6m × 0.6m) panel shelves down 6dB below 94Hz. On a 5in × 5in (12cm × 12cm) plate, the response shelves down 6dB below 376Hz.

For best bass and flattest frequency response, the PZM or PCC must be placed on a large hard boundary such as a floor, wall, table, or baffle at least 4ft × 4ft (1.2m × 1.2m).

To reduce bass response, the PZM or PCC can be mounted on a small plate well away from other reflecting surfaces. This plate can be made of thin plywood, Masonite, clear plastic, or any other hard, smooth material. When used on a carpeted floor the PZM or PCC should be placed on a hard-surfaced panel at least 1ft × 1ft (0.3m × 0.3m) for flattest high-frequency response.

To determine the frequency f_{-6dB} where the response shelves down 6dB, use

$$f_{-6\text{ dB}} = \frac{188^*}{D} \qquad (20\text{-}19)$$

*57.3 for SI units
where,
D is the boundary dimension in ft or m.

For example, if the boundary is 2ft (0.6m) square, the 6dB down point is

$$
\begin{aligned}
f_{-6dB} &= \frac{188}{D} \\
&= \frac{188}{2} \\
&= 94\,\text{Hz}
\end{aligned}
$$

Below 94Hz, the response is a constant 6dB below the upper mid-frequency level. Note that there is a response shelf, not a continuous roll-off.

When the PZM is on a rectangular boundary, two shelves appear. The long side of the boundary is D_{max} and the short side D_{min}. The response is down 3dB at

$$f_{-3dB} = \frac{188^*}{D_{max}} \qquad (20\text{-}20)$$

*57.3 for SI units

and is down another 3 dB at

$$f_{-3dB} = \frac{188^*}{D_{min}} \qquad (20\text{-}21)$$

*57.3 for SI units

The low-frequency shelf varies with the angle of the sound source around the boundary. At 90° incidence (sound wave motion parallel to the boundary), there is no low-frequency shelf.

The depth of the shelf also varies with the distance of the sound source to the panel. The shelf starts to disappear when the source is closer than a panel dimension away. If the source is very close to the PZM mounted on a panel, there is no low-frequency shelf; the frequency response is flat.

If the PZM is at the junction of two or more boundaries at right angles to each other, the response shelves down 6dB per boundary at the above frequency. For example, a two-boundary unit made of 2ft (0.6m) square panels shelves down 12dB below 94Hz.

There are other frequency-response effects in addition to the low-frequency shelf. For sound sources on-axis to the boundary, the response rises about 10dB above the shelf at the frequency where the wavelength equals the boundary dimension.

For a square panel,

$$F_{peak} = \frac{0.88c}{D} \tag{20-22}$$

where,
c is the speed of sound (1130ft/s or 344m/s)
D is the boundary dimension in ft or m.

For a circular panel

$$F_{peak} = \frac{c}{D} \tag{20-23}$$

As an example, a 2ft (0.6m) square panel has a 10dB rise above the shelf at

$$
\begin{aligned}
F_{peak} &= \frac{0.88c}{D} \\
&= \frac{0.88 \times 1130}{2} \\
&= 497\,\text{Hz}
\end{aligned}
$$

Note that this response peak is only for the direct sound of an on-axis source. The effect is much less if the sound field at the panel is partly reverberant, or if the sound waves strike the panel at an angle. The peak is also reduced if the microphone capsule is placed off-center on the boundary.

20.12.4 Phase Coherent Cardioid (PCC)

The *phase coherent cardioid microphone* (PCC) is a surface-mounted supercardioid microphone with many of the same benefits as the PZM. Unlike the PZM, however, the PCC uses a subminiature supercardioid microphone capsule.

Technically, the PCC is not a pressure zone microphone. The diaphragm of a PZM is parallel to the boundary; the diaphragm of the PCC is perpendicular to the boundary. Unlike a PZM, the PCC aims along the plane on which it is mounted, i.e., the main pickup axis is parallel with the plane.

20.13 Ribbon Microphones

A ribbon microphone, Fig. 20-76, consists of a baffle, often rectangular in shape, with a rectangular slit in the middle of it which houses a very thin, flexible ribbon, Fig. 20-77. Magnets produce a powerful magnetic field in the slit, Fig. 20-78.

Figure 20-76. Royer Labs Model R-122 Active Ribbon™ Velocity Microphone. Courtesy Royer Labs.

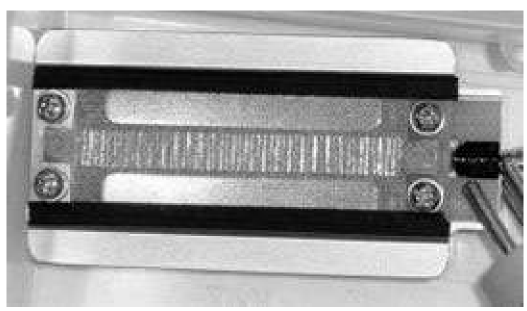


Figure 20-77. Royer Labs ribbon microphone structure. Courtesy Royer Labs.

When a sound wave approaches the microphone from either in front of or behind it, the result will be a difference in pressure between the front and the back of the microphone, and the ribbon will move accordingly. The motion of the ribbon will be the same as the sound wave actuating it and the voltage induced into the ribbon will be proportional to the velocity of the ribbon's motion.

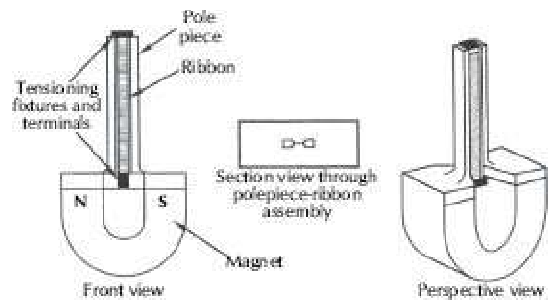


Figure 20-78. Ribbon microphone assembly.

Due to the fact that the ribbon displacements is dependent on a difference in air pressure between the front and the back of the microphone, the sound-pickup pattern will be strongly bi-directional, a sound-source located in the same plane as the ribbon will produce equal-but-opposite forces on the ribbon, and, hence, no ribbon motion and no output.* This produces the figure 8 pattern, Fig. 20-79A. The frequency response is shown in Fig. 20-79B.

Since the voltage induced into the ribbon is proportional to the velocity of its motion and its displacement is proportional to the velocity of the sound wave, the frequency response can be made very flat if care is taken with the design of the microphone.

Ribbon microphones generally have excellent bass response and also exhibit "proximity effect." A typical bi-directional ribbon microphone will have a flat frequency response at a distance of about six feet from the microphone, but at shorter distances the bass response becomes boosted; the effect becomes increasingly pronounced as the distance between the microphone and the sound source is reduced, Fig. 20-80.

This bass boosting characteristic can become quite intense and, if desired, can be corrected by equalization. However, for a multiple microphone setup, the pronounced bass boosting (due to proximity effect) can be turned to an advantage. If an instrument, such as a trumpet, is extremely close mic'ed and the bass is cut to restore flat response, unwanted low frequency sounds are cut back by as much as 20dB compared to an unequalized microphone with a flat response. This discrimination is independent of the microphone's polar response.

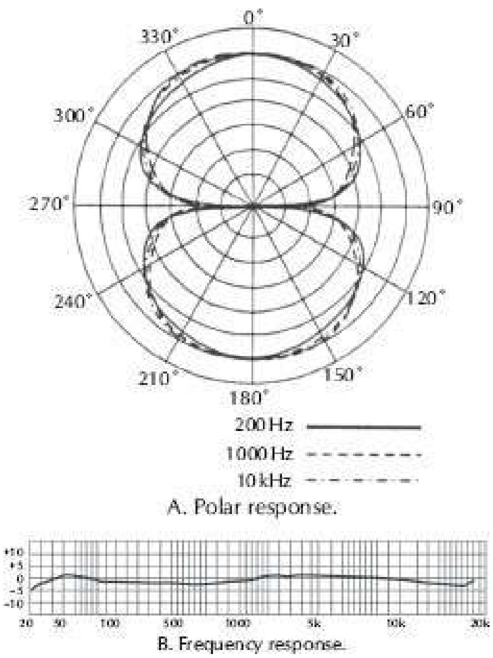


Figure 20-79. Polar and frequency response of a Royer Labs R-122 microphone. Courtesy Royer Labs.

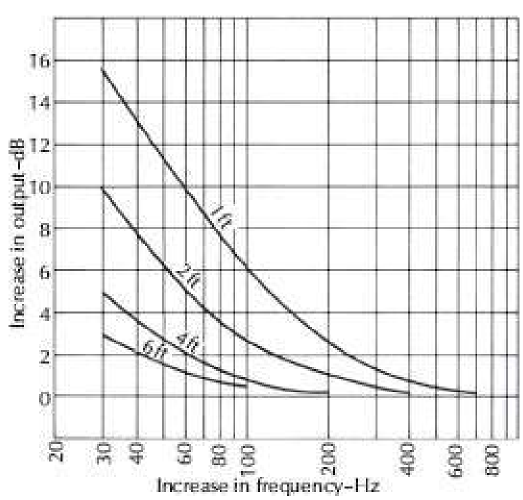


Figure 20-80. Typical relationship of microphone distance to frequency response for ribbon velocity bidirectional microphone.

Another area where proximity effect can be turned to an advantage is to make things sound more “real than real”. For example, many voices and certain musical instruments produce fundamental frequencies within the bass range (below 150Hz or so) but the fundamentals are weak. If a microphone that has no proximity effect and a rising high frequency response is used on an upright piano, or on a person with a thin, weak voice, the recorded sound is likely to sound even thinner than it was in real life. In contrast, using a microphone with strong proximity effect on such sound sources can deliver a “better than real” sound, since the boosted bass response will compensate for the weak fundamentals in the sound source. Since the fundamentals are present, but weakened, boosting them by several dB will sound “natural”, even though the sound has been “sweetened”.

In a typical ribbon microphone design, the open circuit voltage across the ribbon will be on the order of −80dB referred to a volt per pascal (V/Pa); far too low to be usable, however, due to the extremely low impedance of the ribbon, around 0.1Ω, an output of on the order of −50dB referred to V/Pa can be easily managed using a suitable step-up transformer; by combining a step-up transformer and a suitable “head-amp” circuit, an output of −36dB referred to a V/Pa and an equivalent noise of 15dB or better is attainable.

20.14 Lavalier Microphones

Lavalier microphones may be dynamic, condenser (capacitor), pressure-zone, electret, or high-impedance ceramic.

Clip-on lavalier microphones, Figs. 20-81 and 20-82, do not require frequency response correction because there is no coupling to the chest cavity and the small diameter of the diaphragm does not create pressure build-up at high frequencies, creating directionality. Most lavalier microphones are omnidirectional, however, more manufacturers are producing directional lavalier microphones. The Sennheiser MKE 104 clip-on lavalier microphone, Fig. 20-82, has a cardioid pickup pattern, Fig. 20-83. This reduces feedback, background noise, and comb filtering caused by the canceling between the direct sound waves and sound waves that hit the microphone on a reflective path from the floor, lectern, and so forth.

Figure 20-81. Shure MX183 omnidirectional condenser lavalier microphone. Courtesy Shure Incorporated.

Figure 20-82. Sennheiser MKE 104 lavalier clip-on directional microphone. Courtesy Sennheiser Electronic Corporation.

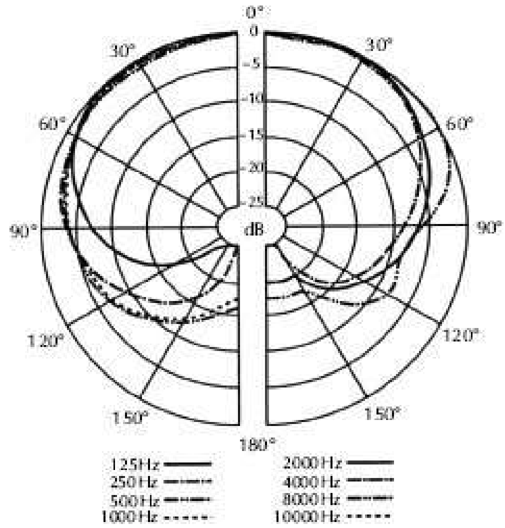


Figure 20-83. Polar response of the microphone in Fig. 20-75. Courtesy Sennheiser Electronic Corporation.

One of the smallest microphones is the Countryman B6, Fig. 20-84. The B6 microphone has a diameter of 0.1 in (2.54mm) and has replaceable protective caps. Because of its small size, it can be hidden even when it's in plain sight. By choosing a color cap to match the environment, the microphone can be pushed through a button hole or placed in the hair.

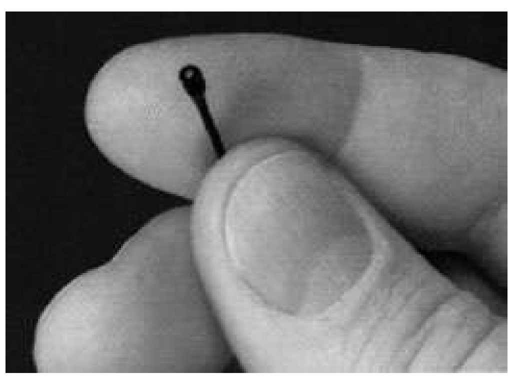


Figure 20-84. Countryman B6 miniature lavalier microphone. Courtesy of Countryman Associates, Inc.

Lavalier microphones are normally used to give the talker freedom of movement. This causes problems associated with motion—for instance, noise being transmitted through the microphone cable. To reduce this noise, soft, flexible microphone cable with good fill to reduce wire movement should be used, see Chapter 18 *Transmission Techniques: Wire and Cable*. The cable, or power supply for electret/condenser microphones, should be clipped to the user's belt or pants to reduce cable noise to only that created between the clip and the microphone, about 2ft (0.6m). Clipping to the waist also has the advantage of acting as a strain relief when the cord is accidentally pulled or stepped on.

A second important characteristic of the microphone cable is size. The cable should be as small as possible to make it unobtrusive and light enough so it will not pull on the microphone and clothing.

Because the microphone is normally 10in (25cm) from the mouth of the talker and out of the signal path, the microphone output is less than a microphone on a stand in front of the talker. Unless the torso is between the microphone and loudspeaker, the lavalier microphone is often a prime candidate for feedback. For this reason, the microphone response should be as smooth as possible.

As in any microphone situation, the farther the microphone is away from the source, the more freedom of movement between microphone and source without adverse effects. If the microphone is worn close to the neck for increased gain, the output level will be greatly affected by the raising and lowering and turning of the talker's head. It is important that the microphone be worn chest high and free from clothing, etc. that might cover the capsule, reducing high-frequency response.

20.15 Head-Worn Microphones

20.15.1 Shure SM10A and Beta 53

The Shure Model SM10A, Fig. 20-85, and Shure Beta 53, Fig. 20-86, are low-impedance, dynamic microphones designed for sports and news announcing, for interviewing and intercommunications systems, and for special-event remote broadcasting. The Shure SM10A is a unidirectional microphone while the Beta 53 is an omnidirectional microphone.

Figure 20-85. Shure SM10A dynamic unidirectional head-worn microphone. Courtesy Shure Incorporated.

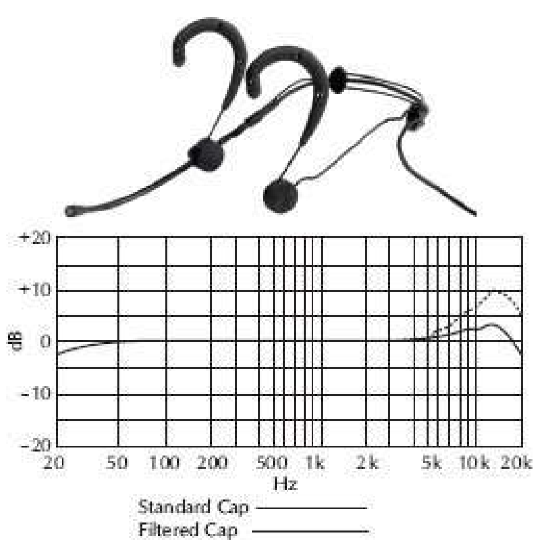


Figure 20-86. Shure Beta 53 omnidirectional microphone and its frequency response. Courtesy Shure Incorporated.

Head-worn microphones offer convenient, hands-free operation without user fatigue. As close-talking units, they may be used under noisy conditions without losing or masking voice signals. They are small, lightweight, rugged, and reliable units that normally mount to a cushioned headband. A pivot permits the microphone boom to be moved 20° in any direction and the distance between the microphone and pivot to be changed 9cm (3½in).

20.15.2 Countryman Isomax E6 Directional EarSet

Countryman's Isomax E6 Directional EarSet microphone is extremely small. The microphone clips around the ear rather than around the head. The units come in different colors to blend in with the background. The ultra-miniature condenser element is held close to the mouth by a thin boom and comfortable ear clip. The entire assembly weighs less than 0.07oz (2g) and almost disappears against the skin, so performers can forget it's there and audiences barely see it, Fig. 20-87.

Figure 20-87. Countryman Isomax E6 Directional EarSet microphone. Courtesy Countryman Associates, Inc.

The microphone requires changeable end caps that create a cardioid pickup pattern for ease of placement, or a hypercardioid pattern when more isolation is needed. The C (cardioid) and H (hypercardioid) end caps modify the EarSet's directionality, Fig. 20-88.

The EarSet series should always have a protective cap in place to keep sweat, makeup and other foreign material out of the microphone.

The hypercardioid cap provides the best isolation from all directions, with a null toward the floor where wedge monitors are often placed. The hypercardioid is slightly more sensitive to air movement and handling noise and should always be used with a windscreen.

The cardioid cap is slightly less directional, with a null roughly toward the performer's back. It's useful for trade show presenters or others who have a monitor loudspeaker over their shoulders or behind them.

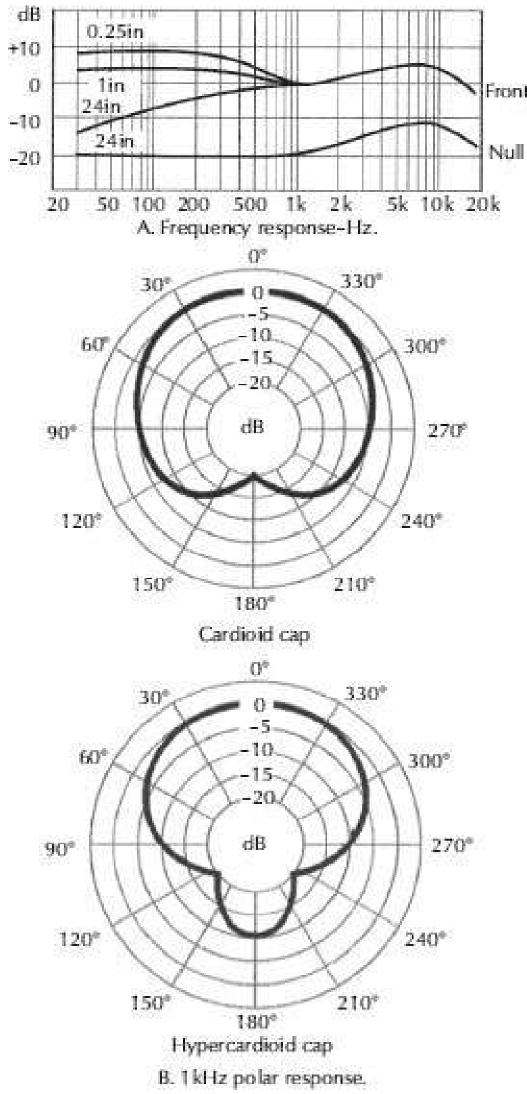


Figure 20-88. Frequency response and polar response of the Countryman E6 EarSet microphone in Fig. 20-80. Courtesy Countryman Associates, Inc.

20.15.3 Audio-Technica BP894

The Audio-Technica BP894, Fig. 20-89, is a headworn condenser microphone with a cardioid polar pattern, Fig. 20-90.

Figure 20-89. Audio-Technica BP894 headworn microphone. Courtesy Audio-Technica U. S., Inc.

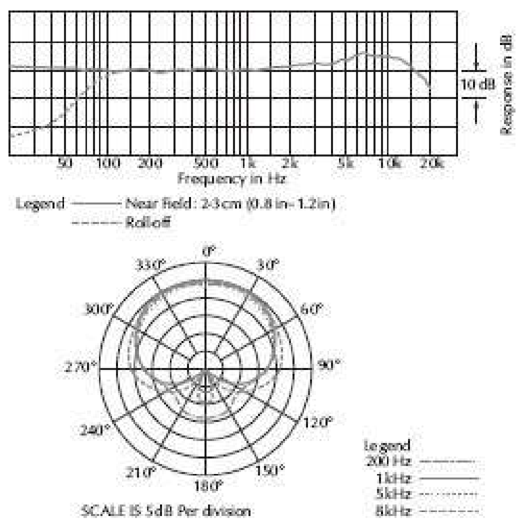


Figure 20-90. Audio-Technica BP894 headworn microphone frequency and polar response. Courtesy Audio-Technica U. S., Inc.

An integral 80Hz high-pass UniSteep® filter provides switching from a flat frequency response to a low-end roll-off, reducing the microphone's sensitivity to popping in close vocal use. It also reduces the pickup of low-frequency ambient noise (such as traffic, air-handling systems, etc.), room reverberation and mechanically coupled vibrations.

The microphone includes a 1.4m (55in) permanently attached miniature cable and requires 11–52V phantom power for operation.

The microphone is also available in a variety of wireless models, including a variety of terminations for use with many manufacturers' wireless systems.

The BP894 features a 180° rotating capsule housing, allowing the user to wear the microphone on either the left or right ear. A talk-side indicator shows the user which way to turn the capsule housing for their desired position.

A cable clip is provided for strain relief, allowing the microphone to remain securely in place without the weight of the cable pulling on the headset.

The included AT8464 Dual-Ear Microphone Mount, Fig. 20-89, allows the single ear-worn BP894 MicroSet® to be worn as a dual-ear-worn unit. The microphone capsule fits on either side of the Dual-Ear Microphone Mount, allowing the microphone to be worn to either the left or right of your mouth. The headband easily adjusts to fit both children and adults.

20.16 Base Station Microphones

Base station power microphones are designed specifically for citizens band transceivers, amateur radio, and two-way radio applications. For clearer transmission and improved reliability, transistorized microphones can be used to replace ceramic or dynamic, high- or low-impedance microphones supplied as original equipment.

The Shure Model 450 Series II, Fig. 20-91, is a high output dynamic microphone designed for paging and dispatching applications. The microphone has an omnidirectional pickup pattern and a frequency response tailored for optimum speech intelligibility, Fig. 20-92. It includes an output impedance selection switch for high, 30,000Ω, and low, 225Ω, and a locking press-to-talk switch, Fig. 20-93.

Figure 20-91. Shure 450 Series II base station microphone. Courtesy Shure Incorporated.

20.17 Differential Noise-Canceling Microphones

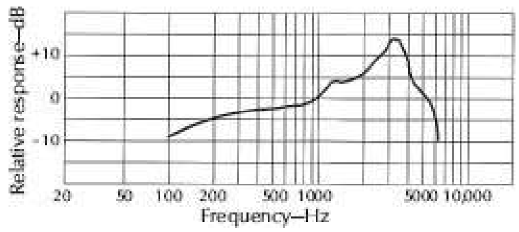


Figure 20-92. Frequency response of the Shure 450 Series II microphone shown in Fig. 20-82. Courtesy Shure Incorporated.

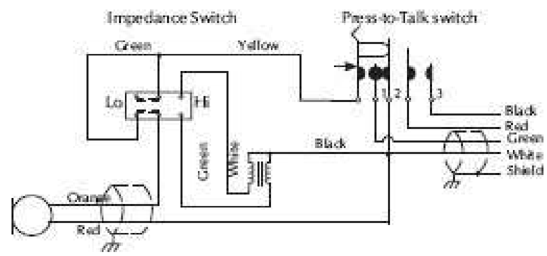


Figure 20-93. Block diagram of the Shure 450 Series II microphone. Courtesy Shure Incorporated.

Figure 20-94. Shure 577B dynamic noise-canceling microphone. Courtesy Shure Incorporated.

Differential noise-canceling microphones, Fig. 20-94, are essentially designed for use in automobiles, aircraft, boats, tanks, public-address systems, industrial plants, or any service where the ambient noise level is 80dB or greater and the microphone is handheld. All sounds originating more than ¼in (6.4mm) from the front of the microphone are canceled. The noise-canceling is achieved through the use of a balanced port opening, which directs the unwanted sound to the rear of the dynamic unit diaphragm out of phase with the sound arriving at the front of the microphone. The noise canceling is most effective for frequencies above 2000Hz. Only speech originating within ¼in (6.4mm) of the aperture is fully reproduced. The average discrimination between speech and noise is 20dB with a frequency response of 200Hz to 5000Hz.

20.18 Controlled-Reluctance Microphones

The *controlled-reluctance microphone* operates on the principle that an electrical current is induced in a coil, located in a changing magnetic field. A magnetic armature is attached to a diaphragm suspended inside a coil. The diaphragm, when disturbed by a sound wave, moves the armature and induces a corresponding varying voltage in the coil. High output with fairly good frequency response is typical of this type of microphone.

Handheld Entertainer Microphones

The *handheld entertainer microphone* is most often used by a performer on stage and, therefore, requires a special frequency response that will increase articulation and presence. The microphones are often subjected to rough handling, extreme shock, and vibration. For live performances, the proximity effect can be useful to produce a low bass sound.

Probably the most famous entertainer's microphone is the Shure SM58, Fig. 20-95. The microphone has a highly effective spherical wind screen that also reduces breath pop noise. The cardioid pickup pattern helps reduce feedback. The frequency response, Fig. 20-96, is tailored for vocals with brightened midrange and bass roll-off. Table 20-2 gives the suggested microphone placement for best tone quality.

Figure 20-95. Shure SM58 vocal microphone. Courtesy Shure Incorporated.

To overcome rough handling and handling noise, special construction techniques are used to reduce wind, pop noise, and mechanical noise and to ensure that the microphone will withstand sudden collisions with the floor. The Sennheiser MD431, Fig. 20-97, is an example of a high-quality, rugged, and low-mechanical-noise microphone. To eliminate feedback, the MD431 incorporates a supercardioid directional characteristic, reducing side pickup to 12% or less than half that of conventional cardioids.

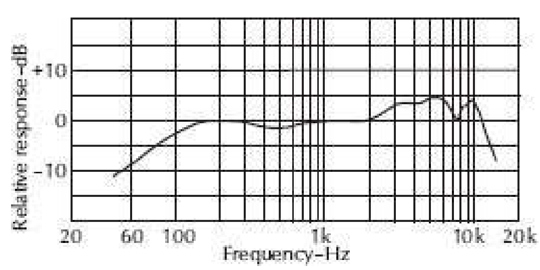


Figure 20-96. Frequency response of the Shure SM58 vocal microphone. Courtesy Shure Incorporated.

Table 20-2. Suggested Placement for the SM58 Microphone

Application	Suggested Microphone Placement	Tone Quality
Lead and backup vocals	Lips less than 150mm (6 in) away or touching the windscreen, on axis to microphone	Robust sound, emphasized bass, maximum isolation from other sources
Speech	from mouth, just above nose height	Natural sound, reduced bass
	200mm (8in) to 0.6m (2ft) away from mouth, slightly off to one side	Natural sound, reduced bass and minimal “s” sounds
	1m (3 ft) to 2m (6 ft) away	Thinner; distant sound; ambience

Another problem, particularly with powerful sound reinforcement systems, is mechanical (handling) noise. Aside from disturbing the audience, it can actually damage equipment. As can be seen in the cutaway, the MD 431 is actually a microphone within a microphone. The dynamic transducer element is mounted within an inner capsule, isolated from the outer housing by means of a shock absorber. This protects it from handling noise as well as other mechanical vibrations normally encountered in live performances.

To screen out noise still further, an internal electrical high pass filter network is incorporated to insure that low-frequency disturbances will not affect the audio signal. A built-in mesh filter in front of the diaphragm reduces the popping and excessive sibilance often produced by close micing.

The microphone case is a heavy-duty cast outer housing with a stainless steel front grille and reed-type on–off switch. A hum bucking coil is mounted behind the transducer to cancel out any stray magnetic fields.

20.19 Pressure-Gradient Condenser Microphones

One of the most popular studio microphones is the Neumann U-87, Fig. 20-98, multidirectional condenser microphone, and its cousin, the Neumann U-89, Fig. 20-99. This microphone is used for close micing where high SPLs are commonly encountered. The response below 30Hz is rolled off to prevent low-frequency blocking and can be switched to 200Hz to allow compensation for the bass rise common to all directional microphones at close range.

The figure-eight characteristic is produced by means of two closely spaced or assembled cardioid characteristic capsules, whose principal axes are pointed in opposite directions and are electrically connected in antiphase.

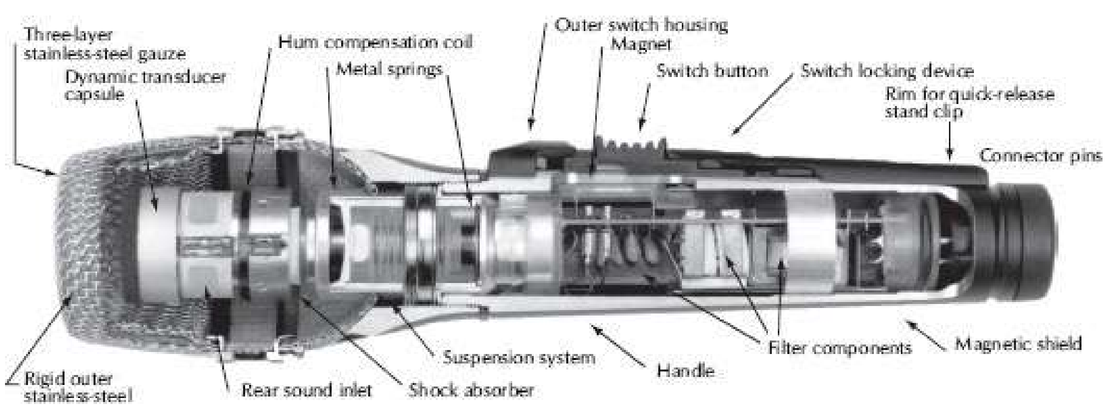


Figure 20-97. Cut-away view of a Sennheiser MD 431 handheld entertainment microphone. Courtesy Sennheiser Electronic Corporation.

These microphones are usually made with backplates equipped with holes, slots, and chambers forming delay elements whose perforations act as part friction resistances and part energy storage (acoustic inductances and capacitances), giving the backplate the character of an acoustic low-pass network. In the cutoff range of this low-pass network, above the transition frequency f_t, the membrane is impinged upon only from the front, and the microphone capsule changes to a pressure or interference transducer.

Figure 20-98. Neumann U-87 microphone. Courtesy Neumann USA.

The output voltage $e(t)$ of a condenser microphone using dc polarization is proportional to the applied dc voltage E_o and, for small displacement amplitudes of the diaphragm, to the relative variation in capacity $c(t)/C_o$ caused by the sound pressure

$$e(t) = E_o \frac{c(t)}{C_o} \qquad (20\text{-}24)$$

where,
E_o is the applied dc voltage,
$c(t)$ is the variable component of capsule capacity,
C_o is the capsule capacity in the absence of sound pressure,
t is the time.

Figure 20-99. Neumann U-89 microphone. Courtesy Neumann USA.

The dependence of output voltage $e(t)$ on E_o is utilized in some types of microphones to control the directional characteristic. Two capsules with cardioid characteristics as shown in Fig. 20-100 are placed back to back. They can also be assembled as a unit with a common backplate. The audio signals provided by the two diaphragms are connected in parallel through a capacitor C. The intensity and phase relationship of the outputs from the two capsule halves can be affected by varying the dc voltage applied to one of them (the left cartridge in Fig. 20-100). This can be accomplished through a switch, or a potentiometer. The directional characteristic of the microphone may be changed by remote control via long extension cables.

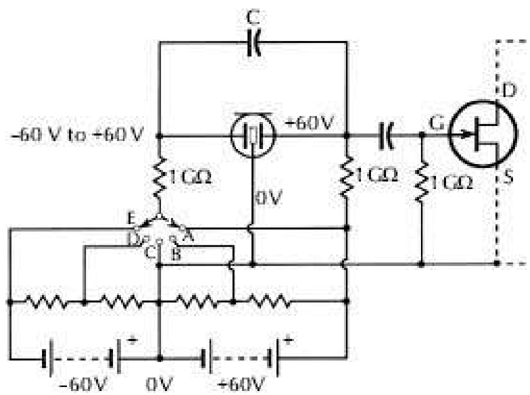


Figure 20-100. Circuit of the Neumann U-89 condenser microphone with electrically switchable direction characteristic. Courtesy Neumann USA.

If the switch is in its center position C, then the left capsule-half does not contribute any voltage, and the microphone has the cardioid characteristic of the right capsule-half. In switch position A, the two ac voltages are in parallel, resulting in an omnidirectional pattern. In position E the two halves are connected in antiphase, and the result is a figure-8 directional response pattern.

The letters A to E given for the switch positions in Fig. 20-100 produce the patterns given the same letters in Fig. 20-101.

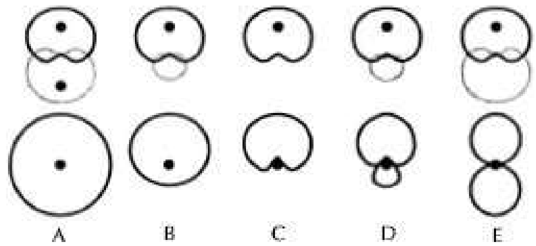


Figure 20-101. By using a microphone as shown in Fig. 20-99 and superimposing two cardioid patterns (top row), directional response patterns (bottom row) can be obtained. Courtesy Neumann USA.

20.20 Steering Array Microphone System

The TOA AM-1 Real-Time microphone, Fig. 20-102, is an array microphone system comprised of (8) cardioid electret microphone elements. The array's pick-up beam is a half-cardioid pattern in the vertical axis since it is a boundary-type array, Fig. 20-103. In the horizontal axis the beam is a constant 50° pick-up angle from 500Hz to 20kHz. This high directivity enables the AM-1 to provide gain-before-feedback that is superior to other fixed-position microphones.

Figure 20-102. Toa AM-1 Real-Time microphone. Courtesy TOA

Electronics Inc.

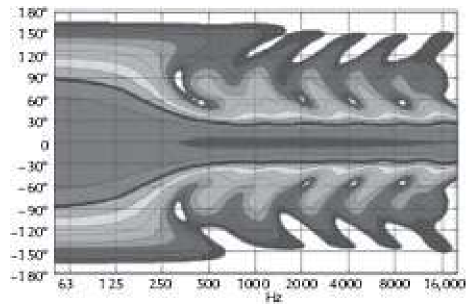


Figure 20-103. Toa AM-1 Real-Time microphone vertical coverage. Courtesy TOA Electronics Inc.

The array's beam is steered in real time with an on-board DSP using triangulation comparing the different arrival times of a sound source's impulse response at particular elements in the array. The array has to perform these calculations. The array's DSP also enables both angular and distance limits of pick-up to avoid steering the beam towards reinforcement loudspeakers or undesirable background noise sources. It can be software controlled via a hardwired or wireless network employing either a PC or an Apple iPad.

Steerable microphone arrays are not new. Various applications have been developed over the years. However, the spacing of the AM-1's elements is unique in that the constant-directivity pick-up beam for each octave from 415Hz to 18.3kHz is horizontally co-axial. This is the first commercially available product to offer a real-time steerable microphone array that has both constant-directivity and a constant horizontal axis for the pick-up beam.

20.21 Interference Tube Microphone

The *interference tube microphone*[8] as described by Olson in 1938 is often called a *shotgun microphone* because of its physical shape and directional characteristics.

Important characteristics of any microphone are its sensitivity and directional qualities. Assuming a constant sound pressure source, increasing the microphone to the source distance requires an increase in the gain of the amplifying system to produce the same output level. This is accompanied by a decrease in *SNR* and an increase in environmental noises including reverberation and background noise to where the indirect sound may equal the direct sound. The wanted signal then deteriorates to where it is unusable. Distance limitations can be overcome by increasing the sensitivity of the microphone, and the effect of reverberation and noise pickup can be lessened by increasing the directivity of the pattern. The interference tube microphone has these two desirable qualities.

The DPA 4017 is a supercardioid shotgun microphone. It is 8.3in (210mm) long and weighs 2.5oz (71g), making it useful on booms, Fig. 20-104. The polar pattern and frequency response is shown in Fig. 20-105.

The difference between interference tube microphones and standard microphones lies in the method of pickup.

Figure 20-104. A supercardioid interference tube microphone. Courtesy DPA Microphones, Inc.

An interference tube is mounted over the diaphragm and is schematically drawn in Fig. 20-106.

The microphone consists of four parts as shown in the schematic:

1. Interference tube with one frontal and several lateral inlets covered with fabric or other damping material.
2. Capsule with the diaphragm and counter electrode(s).
3. Rear inlet.
4. Electronic circuit.

The directional characteristics are based on two different principles:

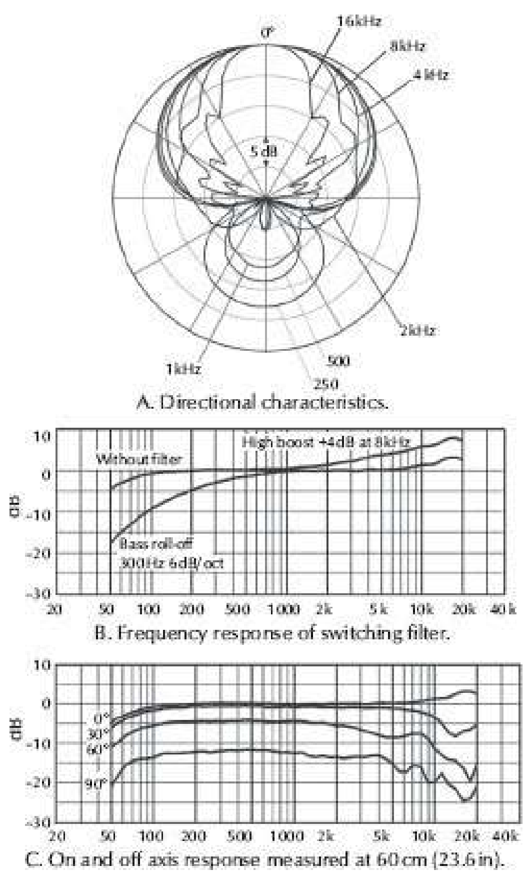


Figure 20-105. Directional characteristics and frequency response of a DPA 4017 microphone. Courtesy DPA Microphones, Inc.

1. In the low-frequency range, tube microphones behave as first-order directional receivers. The tube in front of the capsule can be considered as an acoustic element with a compliance due to the enclosed air volume and a resistance determined by the lateral holes or slits of the tube. The rear inlet is designed as an acoustic low-pass filter to achieve the phase shift for the desired polar pattern (normally cardioid or supercardioid).
2. In the high-frequency range, the acoustical properties of the interference tube determine the polar patterns. The transition frequency between the two different directional characteristics depends on the length of the tube and is given by

$$f_o = \frac{c}{2L} \qquad (20\text{-}25)$$

where,
f_o is the transition frequency,
c is the velocity of sound in air in ft/s or m/s,
L is the length of the interference tube in ft or m.

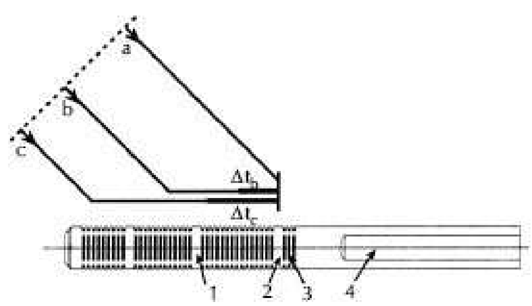


Figure 20-106. Schematic of an interference tube microphone.

Referring to Fig. 20-106, if the tube is exposed to a planar sound wave, every lateral inlet is the starting point of a new wave traveling inside the tube toward the capsule as well as towards the frontal inlet. Apart from the case of frontal sound incidence, every particular wave covers a different distance to the capsule and, therefore, arrives at a different time. Fig. 20-106 shows the delay times of waves *b* and *c* compared to wave a. Note that they increase with the angle of sound incidence.

The resulting pressure at the capsule can be calculated by the sum of all particular waves generated over the tube's length, all with equal amplitudes but different phase shifts. The frequency and phase response curves can be described by

$$\frac{P\theta}{P(\theta = 0°)} = \frac{\sin\left[\frac{\pi L}{\lambda} \times (1 - \cos\theta)\right]}{\frac{\pi L}{\lambda} \times (1 - \cos\theta)} \qquad (20\text{-}26)$$

where,
$P(\theta)$ is the microphone output at a given angle of sound incidence,
$P(\theta = 0°)$ is the microphone output along principal axis,
λ is the wavelength,
L is the length of the tube,
θ is the angle of sound incidence.

The calculated curves and polar patterns are plotted in Figs. 20-107 and 20-108 for a tube length of 9.8in (25cm) without regard to the low-frequency directivity caused by the rear inlet. The shape of the response curves looks similar to that of a comb filter with equidistant minima and maxima decreasing with 6dB/octave. The phase response is frequency independent only for frontal sound incidence. For other incidence angles, the phase depends linearly on frequency, so that the resulting pressure at the capsule shows an increasing delay time with an increasing incidence angle.

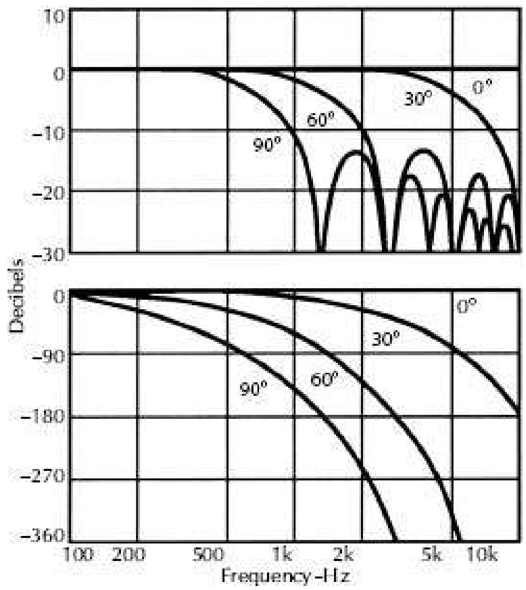


Figure 20-107. Calculated frequency and phase response curves of an interference tube microphone (250mm) without rear inlet for different angles of sound incidence. Courtesy Sennheiser Electronic Corporation.

In practice, interference tube microphones show deviations from this simplified theoretical model. Fig. 20-109 is the polar pattern of the Sennheiser MKH 60P48. The built-in tube delivers a high-frequency roll-off for lateral sound incidence with a sufficient attenuation especially for the first side lobes. The shape of the lateral inlets as well as the covering material influences the frequency and phase response curves. The transition frequency can be lowered with an acoustic mass in the frontal inlet of the tube to increase the delay times for low frequencies.

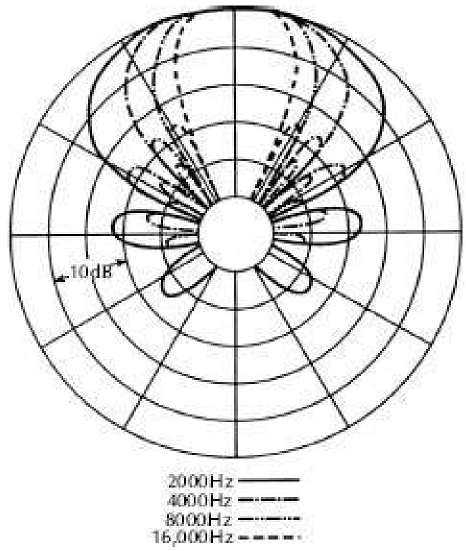


Figure 20-108. Calculated polar patterns of an interference tube microphone (250mm) without rear inlet. Courtesy Sennheiser Electronic Corporation.

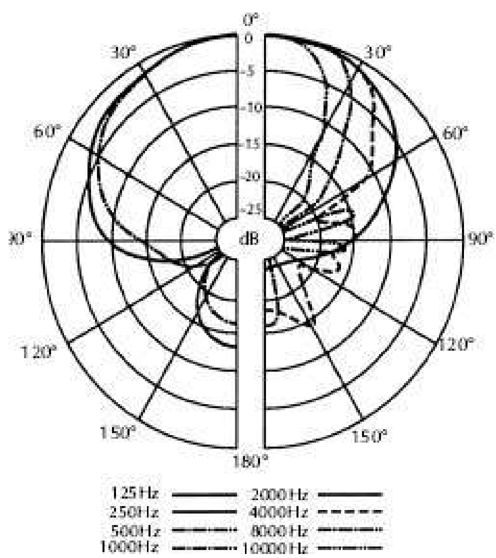


Figure 20-109. Characteristic of a supercardioid microphone (MKH 60P48). Courtesy Sennheiser Electronic Corporation.

Another interference tube microphone is the Shure VP89L with an overall length of 19.23in (488mm), Fig. 20-110.[9] In this microphone, a tapered acoustic resistance is placed over the elongated interference tube slit, varying the effective length of the tube with frequency so that L/M (the ratio of tube length to wavelength) remains nearly constant over the desirable frequency range. This allows the polar response to be more consistent as frequency increases, Fig. 20-111, because the resistance in conjunction with the compliance of the air inside the tube forms an acoustical low-pass filter. High frequencies are attenuated at the end of the tube because it is the high-resistance end, allowing the high frequencies to enter the tube only near the diaphragm. This makes the tube look shorter at high frequencies, Eq. 20-26.

Figure 20-110. Shure VP89L condenser shotgun microphone. Courtesy Shure Incorporated.

While a cardioid microphone may be capable of picking up satisfactorily at 3ft (1m), a cardioid in-line may reach 6–9ft (1.8–2.7m), and a super in-line may reach as far as 40ft (12m) and be used for picking up a group of persons in a crowd from the roof of a nearby building, following a horse around a race track, picking up a band in a parade, and picking up other hard-to-get sounds from a distance.

There are precautions that should be followed when using interference tube microphones. Because they obtain directivity by cancellation, frequency response and phase are not as smooth as omnidirectional microphones. Also, since low frequencies become omnidirectional, the frequency response drops rapidly below 200Hz, which helps control directivity.

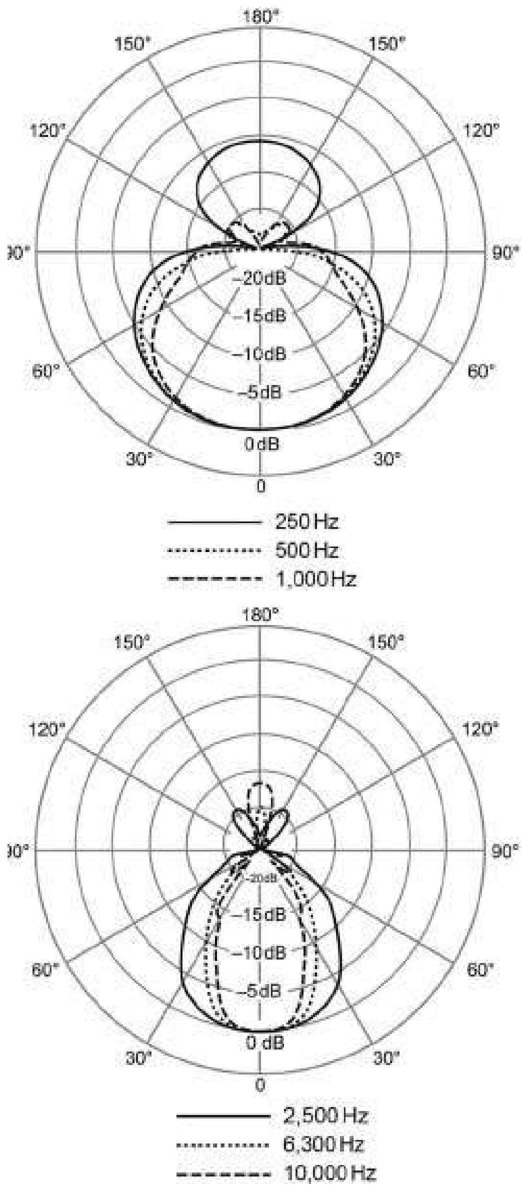


Figure 20-111. Polar response of the VP89 shotgun microphone. Courtesy Shure Incorporated.

It should not be assumed that no sound will be picked up outside the pickup cone. As the microphone is rotated from an on-axis position to a 180° off-axis position, there will be a progressive drop in level. Sounds originating at angles of 90° to 180° off-axis will cancel by 20dB or more; however, the amount of cancellation depends on the level and distance of the microphone from the sound source. As an example, if an on-axis sound originated at a distance of 20ft (6m), a 90° to 180° off-axis sound occurring at the same distance and intensity will be reduced by 20dB or more, providing none of the off-axis sound is reflected into the front of the microphone by walls, ceiling, and so on. On the other hand, should the off-axis sound originate at a distance of 2ft (0.6m) and at the same sound pressure level as the sound at 20ft (6m) on axis, it will be reproduced at the same level. The reason for this behavior is that the microphone is still canceling the unwanted sound as much as 20dB, but due to the difference in the distances of the two sounds, the off-axis sound is 20dB louder than the on-axis sound at the microphone, so they are reproduced at the same level. For a pickup in an area where random noise and reverberation are problems, the microphone should be located with the back end to the source of unwanted sound and as far from the disturbances as possible.

If the microphone is being used inside a truck or enclosed area, and pointing out a rear door, poor pickup may be experienced because all sounds, both wanted and unwanted, arrive at the microphone on-axis. Since the only entrance is through the truck door, no cancellation occurs because the truck walls inhibit the sound from entering the sides of the microphone. In this instance, the microphone will be operating as an omnidirectional microphone. Due to the reflected sound from the walls, the same condition will prevail in a room where the microphone is pointed through a window or when operating in a long hallway. For good pickup, the microphone should be operated in the open and not in closely confined quarters.

Difficulties may be encountered using interference tube microphones on stage and picking out a talker in the audience, particularly where the voice is 75–100ft (23–30m) away and fed back through a reinforcement system for the audience to hear. Under these circumstances, only about 30–50ft (9–15m) is possible without acoustic feedback; even then, the system must be balanced very carefully.

20.22 Parabolic Microphones

Parabolic microphones use a parabolic reflector with a microphone to obtain a highly directional pickup response. The microphone diaphragm is mounted at the focal point of the reflector, Fig. 20-112. Any sound arriving from an angle other than straight on will be scattered and therefore will not focus on the pickup. The microphone is focused by moving the diaphragm in or out from the reflector for maximum pickup. This type concentrator is often used to pick up a horse race or a group of people in a crowd.

The greatest gain in sound pressure is obtained when the reflector is large compared to the wavelength of the incident sound. With the microphone in focus, the gain is the greatest at the mid-frequency range. The loss of high frequencies may be improved somewhat by defocusing the microphone a slight amount, which also tends to broaden the sharp directional characteristics at the higher frequencies. A bowl 3ft (0.91m) in diameter is practically nondirectional below 200Hz but is very sharp at 8000Hz, Fig. 20-

113. For a diameter of 3ft, the gain over the microphone without the bowl is about 10dB and, for a 6 ft (1.8m) diameter bowl, approximately 16dB.

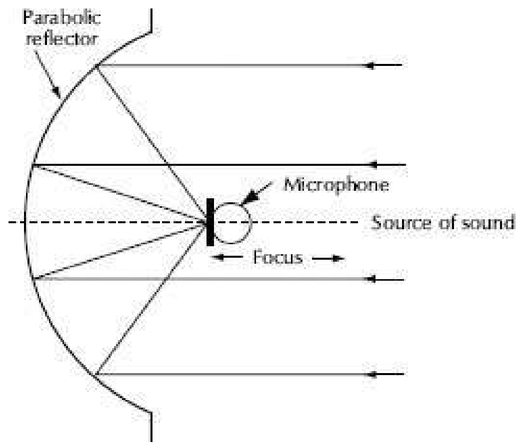


Figure 20-112. A parabolic bowl concentrator for directional microphone pickup.

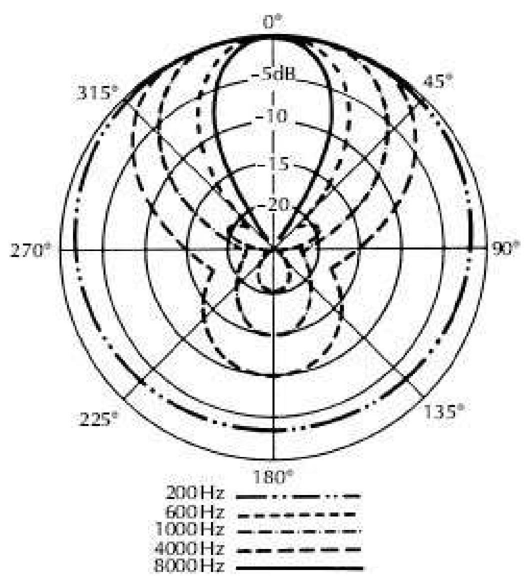


Figure 20-113. Polar pattern for a parabolic concentrator.

20.23 Zoom Microphones

A *zoom microphone*,[10] or *variable-directivity microphone*, is one that operates like and in conjunction with a zoom lens. This type of microphone is useful with television and motion-picture operations.

The optical perception of distance to the object is simply determined by the shot angle of the picture. On the other hand, a sound image is perceived by:

- Loudness.
- Reverberation (ratio of direct sound to reflected sound).
- Acquired response to sound.
- Level and arriving time difference between the two ears.

If the sound is recorded in monophonic, the following factors can be combined to reproduce a natural sound image with respect to the perceived distance:

- *Loudness*: Perceived loudness can be controlled by varying microphone sensitivity.
- *Reverberation*: The representation of the distance is made by changing the microphone directivity or the ratio between direct and reverberant sound. In a normal environment, we hear a combination of direct sound and its reflections. The nearer a listening point is to the source, the larger the ratio of direct to reverberant sound. The farther the listening point is from the source, the smaller the ratio; therefore use of a high-directivity microphone to keep direct sound greater than reflected sound permits the microphone to get apparently closer to the source by reducing reverberant sound pickup. For outdoor environments, use of directional microphones allows the ambient noise level to be changed for natural representation of distances.
- *Acquired human response to sound*: Normally we can tell

approximately how far a familiar object as a car or a person is by the sound generated by the objects because we acquire the response to sound through our daily experiences.

By changing the sensitivity and directivity of a microphone simultaneously, an acoustical zoom effect is realized, and more reality becomes possible in sound recording. Fig. 20-114 is the basic block diagram of a zoom microphone system. The system consists of three unidirectional microphone capsules (1 through 3) arranged on the same axis. The three capsules have the same characteristics, and capsule 3 faces the opposite direction. The directivity can be varied from omnidirectional to second-order gradient unidirectional by varying the mixing ratio of the output of each capsule and changing the equalization characteristic accordingly. An omnidirectional pattern is obtained by simply combining the outputs of capsule 2 and 3. In the process of directivity change from omnidirectional to unidirectional, the output of capsule 3 is gradually faded out, while the output of capsule 1 is kept off. Furthermore, the equalization characteristic is kept flat, because the on-axis frequency response does not change during this process. In the process of changing from unidirectional to second-order gradient unidirectional, the output of capsule 3 is kept off. The second-order gradient unidirectional pattern is obtained by subtracting the output of capsule 1 from the output of capsule 2. To obtain the second-order gradient unidirectional pattern with minimum error, the output level of capsule 1 needs to be trimmed. Since the on-axis response varies according to the mixing ratio, the equalization characteristics also have to be adjusted along with the level adjustment of the output of capsule 1. The on-axis sensitivity increase of second-order gradient setup over the unidirectional setup allows the gain of the amplification to be unchanged.

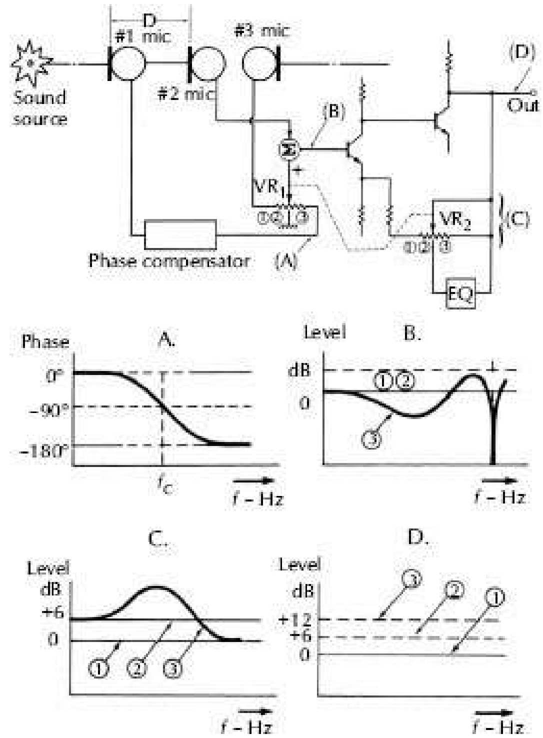


Figure 20-114. Configuration of the zoom microphone.

In order to obtain a good matching of picture and sound, a mechanism that synchronizes the optical zooming and acoustical zooming becomes inevitable. Electrical synchronization would also be possible by using voltage-controlled amplifiers (VCA) or voltage-controlled resistors (VCR).

20.24 PolarFlex™ Microphone System

The PolarFlex system by Schoeps models any microphone. The system features two output channels with two microphones per channel, Fig. 20-115. The standard system consists of an omnidirectional and a figure 8 microphone for each channel and an analog/digital processor.

Essential sonic differences between condenser microphones of the same nominal directional pattern are due to frequency response and the fact that the polar pattern is not always uniformly maintained throughout the entire frequency range particularly at the lowest and highest frequencies. Though ostensibly a defect, this fact can also be used to advantage, e.g., adaptation to the acoustic of the recording room. While the frequency response at a given pickup angle can be controlled by equalizers, there was no way to alter the polar pattern correspondingly. The only way to control this situation was through the choice of microphones having different variations of the polar pattern versus frequency. With the DSP-4P processor, nearly ideal directional characteristics can be selected, and nearly any frequency-dependent directional characteristic that may be desired, e.g., a cardioid becomes omnidirectional below the midrange, so that it has better response at the very lowest frequencies. Also modeling a large-diaphragm microphone is possible.

In excessively reverberant spaces one could record a drier sound (cardioid or supercardioid setting) or, in spaces that are dry, accept more room reflections (wide cardioid or omni setting) in the corresponding frequency range.

In such cases it is not the frequency response but rather the ratio of direct to reflected sound, that will be altered. That cannot be done with an equalizer nor can a reverb unit reduce the degree of reflected sound after the fact.

In the arrangement of Fig. 20-115, an omnidirectional microphone with a mild high-frequency emphasis in the direct sound field is used. Because of its angle of orientation, the capsule has ideal directional response in the horizontal plane; the high-frequency emphasis compensates for the high-frequency losses due to lateral sound incidence.

A figure 8 microphone is set directly above the omni. The direction to which it is aimed will determine the orientation of the resulting adjustable virtual microphone. The hemispherical device attached to the top of the figure 8 flattens the response of the omnidirectional microphone at the highest frequencies.

Figure 20-115. A Schoeps PolarFlex microphone with an omnidirectional and a figure 8 microphone. Courtesy Schoeps GmbH.

By using the DSP-4P processor, Fig. 20-116, the following settings can be made independently of one another in three adjustable frequency ranges. With the three knobs in the upper row, the directional patterns in each of the three frequency bands can be set.

The settings are indicated by a circle of LEDs around each of the knobs. At the lower left of each knob is the omnidirectional setting; at the lower right is the figure 8 setting. Eleven intermediate pattern settings are available. The knobs in the lower row are set between those in the upper row. They are used for setting the boundaries between the frequency ranges; 100Hz–1kHz and 1kHz–10kHz, respectively, in 1/3-octave steps.

The three buttons at the lower right are used to store and recall presets. If the unprocessed microphone signals have been recorded, these adjustments can be made during postprocessing.

Figure 20-116. Schoeps DSP-4P microphone processor. Courtesy Schoeps GmbH.

The processor operates at 24-bit resolution with either a 44.1kHz or 48kHz sampling rate. When a digital device is connected to the input, the PolarFlex™ processor adapts to its clock signal.

20.25 Piano Microphone

The Earthworks PM40 PianoMic™ System was designed to pickup piano sound for live and recording applications without microphone stands or booms, Fig. 20-117. The sound quality is the same with the piano lid either up or down. The system consists of two random incidence, omnidirectional, 40kHz High Definition

Microphones™ with a frequency response of 9Hz to 40kHz, fast impulse response and extremely short diaphragm setting times, Fig. 20-118. The 40kHz High Definition Microphones have been designed for random incidence pickup. When microphones are placed inside a piano, they are within a sound field that has multiple sound sources: i.e. every string, the sound board, multiple reflections of the sound from the sound board, the sides and the lid of the piano. All these sound sources and reflections produce sound waves that arrive at the microphones from all directions. This is called a diffused sound field, and the High Definition Random Incidence Microphones are designed to be used within such diffused sound field. There is no proximity effect, regardless of their distance from the piano strings or soundboard.

A large amount of gain before feedback is achieved because the microphones are placed very close to the sound source and are within the sound field of the piano. There is almost no leakage of sounds from outside the piano because the microphones are placed inside the piano shell and are somewhat isolated from sounds outside the piano plus the microphones are placed 3 to 6 inches above the piano strings, making the sound level of the piano picked up by the microphones much louder than sounds coming from outside the piano.

Figure 20-117. PM40 PianoMic™ System. Courtesy Earthworks, Inc.

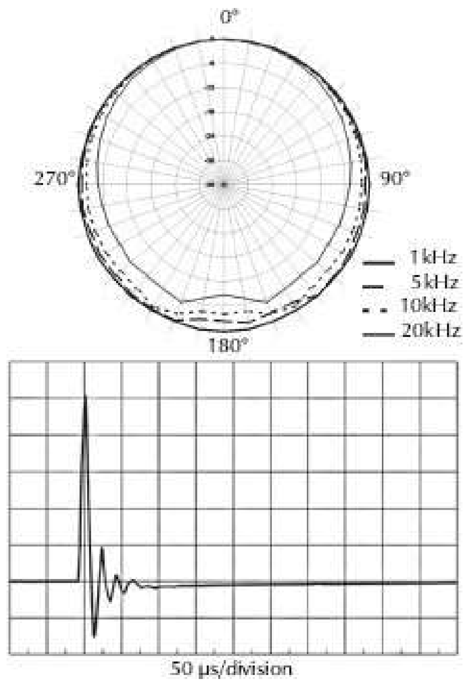


Figure 20-118. PM40 PianoMic™ System polar pattern and impulse response. Courtesy Earthworks, Inc.

The system uses an adjustable telescoping tube that is supported by the sides of the piano shell. The telescoping tube can be adjusted from 46 to 64 inches to easily accommodate various grand pianos. The support arms that sit on the side of the piano case are less than ⅛ in thick, and are smooth with a protective coating that will not harm the finish of the piano. The piano lid can be closed on top of them without stressing the piano lid hinge.

The telescoping tube can be moved so the two microphones can be placed close to the hammers or further away from the hammers. The telescoping tube can be placed on the piano with the microphones facing the keyboard or with the microphones facing away from the keyboard. The center section of the tube can be moved as much as eight inches to the left or the right to favor either the low or high strings of the piano. The flex arms allow the microphone heads to move about four inches to the left or right as well as up or down.

20.26 Stereo Microphones

Stereo microphones are used for coincident, XY, M/S, ORTF, NOS, Blumlien, SASS, binaural in-the-head, and binaural in-the-ear (ITE) recording. They may be individual microphones mounted in a certain way, or may be a composite microphone. These systems have the microphones close together (in proximity of a point source or ear-to-ear distance) and produce the stereophonic effect by intensity stereo, time-based stereo, or a combination of both, Fig. 20-119.

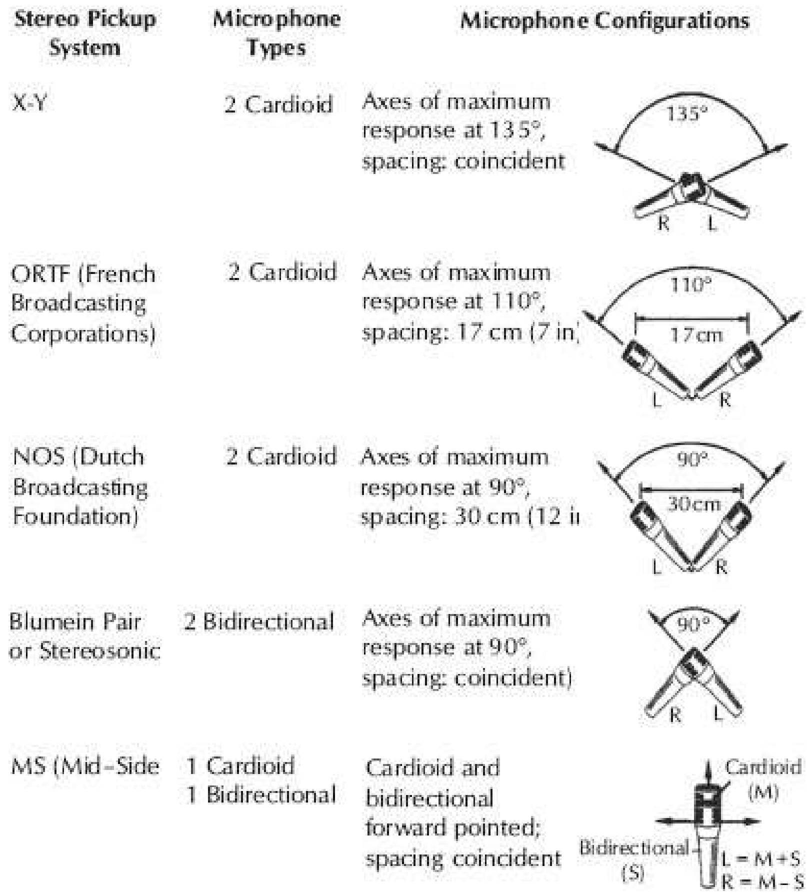


Figure 20-119. Various types of stereo recording options. Courtesy Shure Incorporated.

20.26.1 Coincident Microphones

A highly versatile stereo pickup is the *coincident microphone technique*.[11,12,13] Coincident means that sound reaches both microphones at the same time, implying that they are at the same point in space. In practice, the two microphones cannot occupy the same point, but they are placed as closely together as possible. There are special-purpose stereo microphones available that combine the two microphones in one case. Since they are essentially at the same point, there is no time differences between arrival of any sound from any direction; thus no cancellation occurs. It might first appear that there could be no stereophonic result from this configuration. The two microphones are usually unidirectional and oriented at 90° to one another. The combination is then aimed at the sound source, each microphone 45° to a line through the source. Stereo results from intensity differences—the left microphone (which is to the right of the pair) will receive sounds from the left-hand part of the stage with greater volume than it will receive from the right-hand side of the stage.

The stereo result, although often not as spectacular as that obtained from spaced microphones, is fully mono compatible, and it most accurately reproduces the sound of the acoustic environment. It is quite foolproof and quick to set up.

Variations of the coincident technique include changing the angle between the microphone (some stereo microphones are adjustable); using bidirectional microphones, which results in more reverberant sound; using combinations of unidirectional and bidirectional microphones; and using matrix systems, which electrically provide sum and difference signals from the left and right channels (these can be manipulated later for the desired effect).

The basic coincident technique was developed in the 1930s (along with the first stereo recordings) by English engineer Alan Blumlein.[14] Blumlein used two figure 8 pattern ribbon microphones mounted so that their pattern lobes were at right angles (90°) to each other, as shown in Fig. 20-120. The stereo effect is produced primarily by the difference in amplitude generated in the two microphones by the sound source. A sound on the right generates a larger signal in microphone B than in microphone A. A sound directly in front produces an equal signal in both microphones, and a sound on the left produces a larger signal in microphone A than in microphone B. The same process takes place with spaced omnidirectional microphones, but because of the spacing, there is also a signal delay between two signals (comb filter effect). It can also produce a loss in gain and unpleasant sound if the two channels are combined into a single monosignal. Since the coincident microphone has both its transducers mounted on the same vertical axis, the arrival time is identical in both channels, reducing this problem to a large degree.

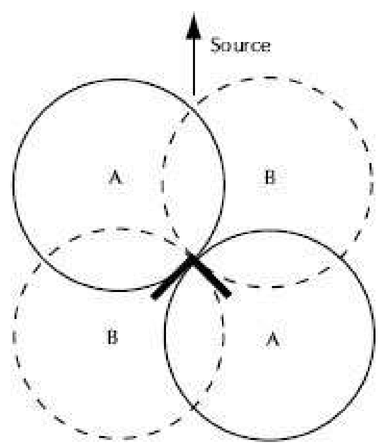


Figure 20-120. Coincident microphone technique using two bidirectional microphones.

Modern coincident microphones often use cardioid or hypercardioid patterns. These patterns work as well as the figure 8 pattern microphones in producing a stereo image, but they pick up less of the ambient hall sound.

Probably the strongest virtue of the coincident microphone technique is its simplicity under actual working conditions. Just place the microphone in a central location that gives a good balance between the musicians and the acoustics of the hall. It is this simplicity that makes coincident microphones a favorite of broadcast engineers recording (or transmitting) live symphonic concerts.

20.26.2 XY Stereo Technique

The XY technique uses two identical directional microphones that, in relation to the recording axis, are arranged at equal and opposed offset angles. The leftward pointing X microphone supplies the L signal directly, and the rightward pointing Y microphone supplies the R signal, Fig. 20-121. The stereophonic properties depend on the directional characteristics of the microphones and the offset angle.

One property specific to a microphone system is the recording angle, which defines the angle between the center axis (symmetry axis of the system) and the direction where the level differences between the L and R define the angular range of sound incidence where regular stereophonic reproduction is obtained. In most cases there is another opening for backward sound reception besides the recording angle for frontal sound pick-up.

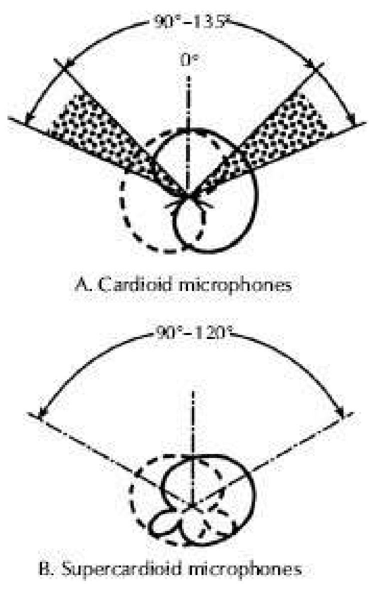


Figure 20-121. XY stereo technique patterns.

Another important aspect concerns the relationship between the sound incidence angle and the stereophonic reproduction angle. As both XY and M/S recording techniques supply pure intensity cues, a relationship can be applied that relates the reproduction angle to the level difference of the L and R signals for the standard listening configuration based on an equilateral triangle, Fig. 20-122. This relationship is shown in Fig. 20-123 and is valid at frequencies between 330 and 7800Hz within ±3°. The level difference is plotted on the horizontal axis and the reproduction angle can be read on the vertical scale. A 0° reproduction angle means localization at the center of the stereo base, and 30° means localization at one of the loudspeakers.

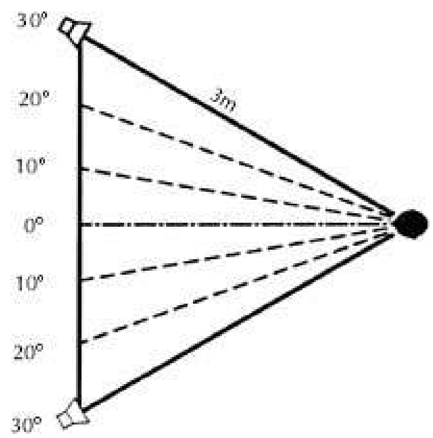


Figure 20-122. Standard listening configuration.

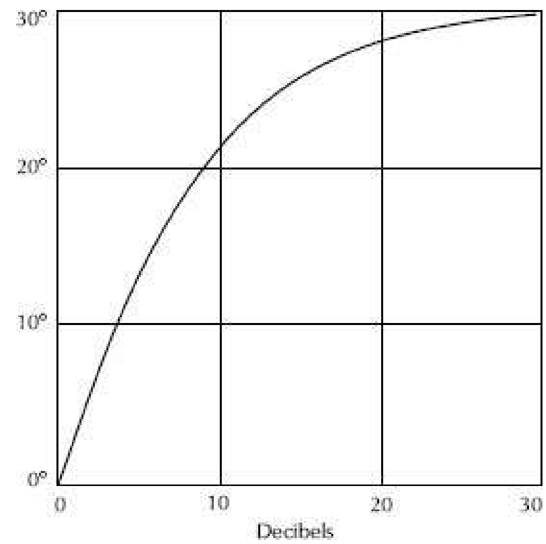


Figure 20-123. Stereophonic localization.

20.26.3 The ORTF Technique

A variation on the basic XY coincident technique is the *ORTF technique*. The initials ORTF stand for *Office de Radiodiffusion Télévision Français*, the French government radio network that developed this technique. The ORTF method uses two cardioid microphones spaced 17cm (7in) apart and facing outward at an angle of 110° between them, Fig. 20-124. Due to the spacing between the transducers, the ORTF method does not have the time-coherence properties of M/S or XY micing.

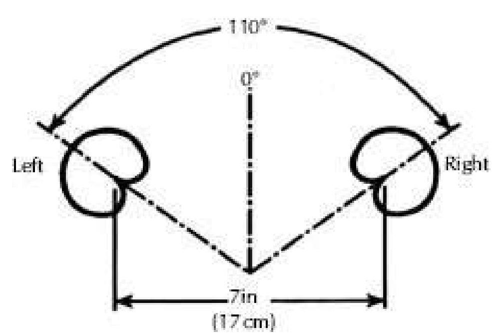


Figure 20-124. ORTF microphone technique.

20.26.4 The M/S Stereo Technique

The M/S technique employs a mid (M) cartridge that directly picks up the mono sum signal, and a side (S) cartridge that directly picks up the stereo difference signal (analogous to the broadcast stereo subcarrier modulation signal). Although two individual microphones may be used, single-unit M/S microphones are more convenient and generally have closer cartridge placement.

Fig. 20-125 indicates the pickup patterns for a typical M/S microphone configuration. The mid cartridge is oriented with its front (the point of greatest sensitivity) aimed at the center of the incoming sound stage. A cardioid (unidirectional) pattern as shown is often chosen for the mid cartridge, although other patterns may also be used. For symmetrical stereo pickup, the side cartridge must have a side-to-side facing bidirectional pattern (by convention, the lobe with the same polarity as the front mid signal aims 90° to the left, and the opposite polarity lobe to the right).

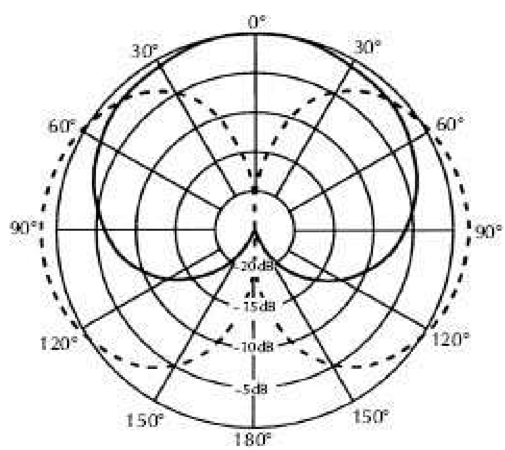


Figure 20-125. MS Microphone Pickup Patterns.

In a stereo FM or television receiver, the mono sum baseband signal and the stereo difference subcarrier signal are demodulated and then decoded, using a sum-and-difference matrix, into left and right stereo audio signals. Similarly, the mid (mono) signal and the side (stereo difference) signal of the MS microphone may be decoded into useful left and right stereo signals.

The mid cartridge signal's relation to the mono sum signal, and the side cartridge signal's relation to the stereo difference signal, can be expressed simply by

$$M = \frac{1}{2(L+R)} \tag{20-27}$$

$$S = \frac{1}{2(L-R)} \tag{20-28}$$

Solving for the left and right signals,

$$L = M+S \tag{20-29}$$

$$R = M-S \tag{20-30}$$

Therefore, the left and right stereo signals result from the sum and difference, respectively, of the mid and side signals. These stereo signals can be obtained by processing the mid and side signals through a sum and difference matrix, implemented with transformers, Fig. 20-126, or active circuitry. This matrix may be external to the M/S microphone or built-in.

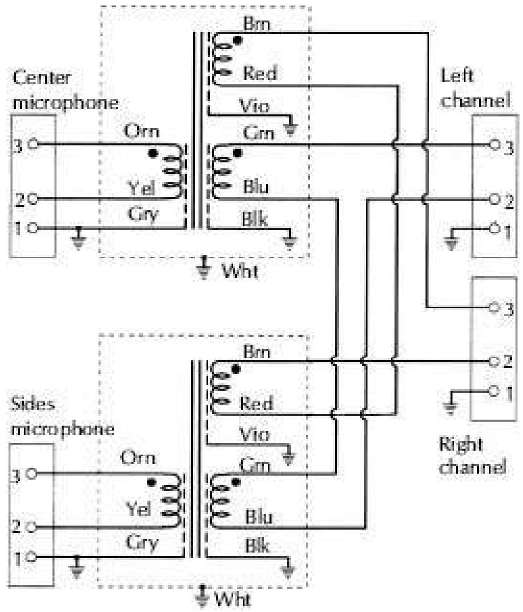


Figure 20-126. Transformer sum and difference matrix for M/S microphones.

In theory, any microphone pattern may be used for the mid signal pickup. Some studio M/S microphones provide a selectable mid pattern. In practice, however, the cardioid mid pattern is most often preferred in M/S microphone broadcast applications.

The Shure VP88 stereo microphone, Fig. 20-127, employs a switchable pattern. Fig. 20-128 shows the polar response of the mid capsule and the side capsule.

Figure 20-127. Shure VP88 stereo condenser microphone. Courtesy Shure Incorporated.

The left and right stereo signals exhibit their own equivalent pickup patterns corresponding to left-forward-facing and right-forward-facing microphones. Fig. 20-129 shows the relative levels of the mid and side microphones and the stereo pickup pattern of the Shure VP88 microphone in the L position with the bidirectional side pattern maximum sensitivity 6dB lower than the maximum mid sensitivity. The small rear lobes of each pattern are 180° out of polarity with the main front lobes. For sound sources arriving at 0° the left and right output signals are equal, and a center image is reproduced between the loudspeakers. As the sound source is moved off-axis, an amplitude difference between left and right is created, and the loudspeaker image is moved smoothly off-center in the direction of the higher amplitude signal.

When the mid (mono) pattern is fixed as cardioid, the stereo pickup pattern can be varied by changing the side level relative to the mid level. Fig. 20-129 shows an M/S pattern in the L, M, and H position with the side level 1.9dB lower than the mid level in the M position and 1.6dB higher than the mid level in the H position. The three resultant stereo patterns exhibit pickup angles of 180°, 127°, and 90°, respectively. The incoming sound angles, which will create left, left-center, center, right-center, and right images, are also shown. Note the changes in the direction of the stereo patterns and the size of their rear lobes.

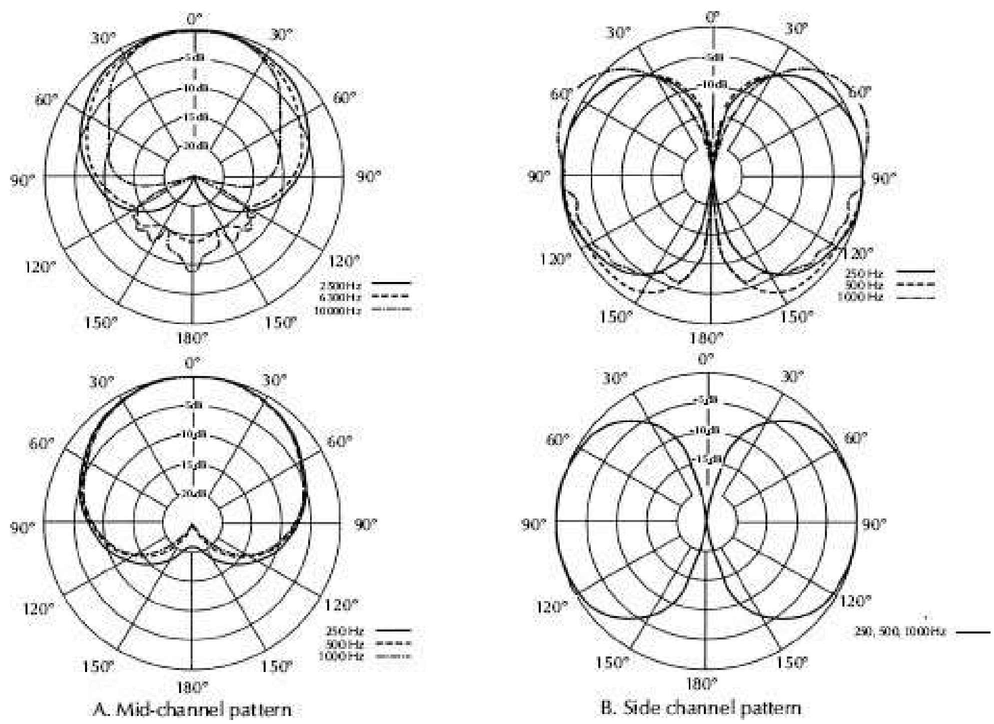


Figure 20-128. Polar response of the Shure VP88 M/S microphone. Courtesy Shure Incorporated.

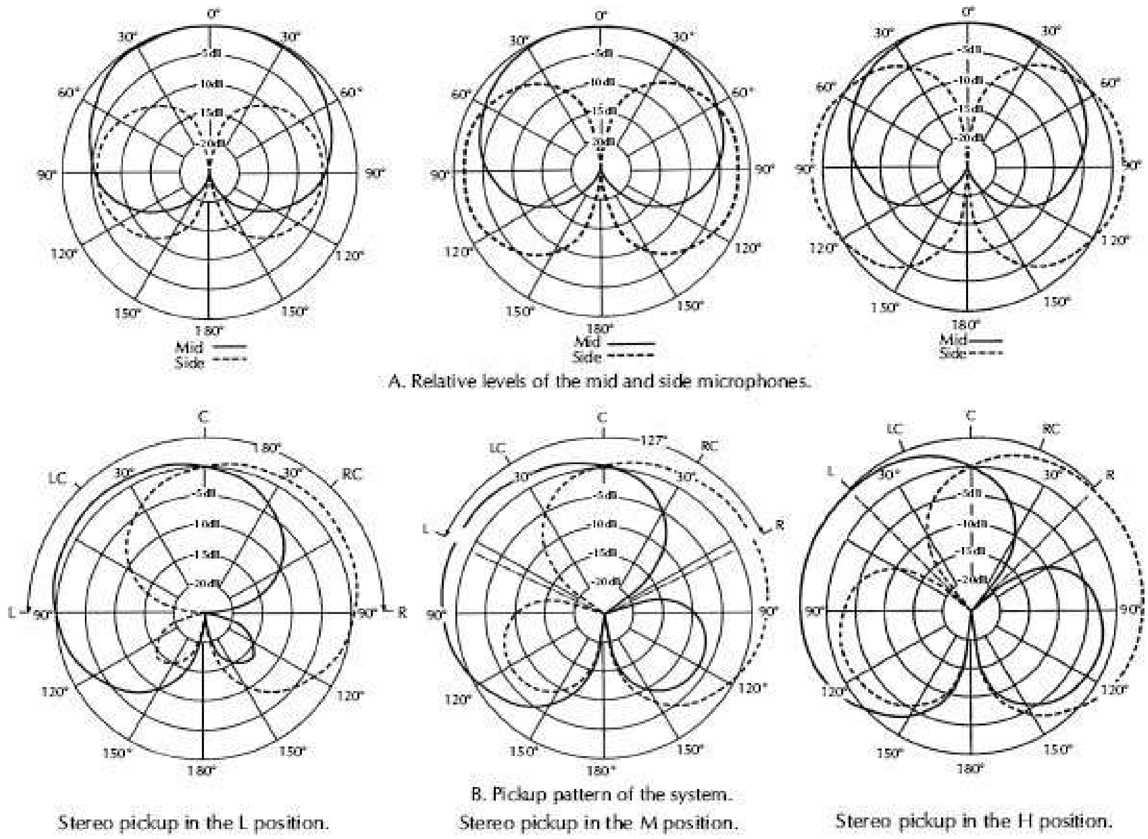


Figure 20-129. Stereo pickup pattern of the Shure VP88 M/S in the L, M, and H position. Courtesy Shure Incorporated.

Taking the directional properties of real microphones into consideration, it becomes clear that the M/S technique provides a higher recording fidelity than the XY technique. There are at least three reasons for this:

1. The microphones in an XY system are operated mainly at off-axis conditions, especially at larger offset angles. The influence of directivity imperfections is more serious than with MS systems, where the M microphone is aimed at the performance center. This is illustrated by Fig. 20-130.
2. The maximum sound incidence angle for the microphone is only half that of the X and Y microphones, although the covered performance area is the same for all microphones. This area is

symmetrically picked up by the M microphone, but unsymmetrically by the X and Y microphones. The M/S system can supply the more accurate monophonic (M) signal in comparison with the XY system.

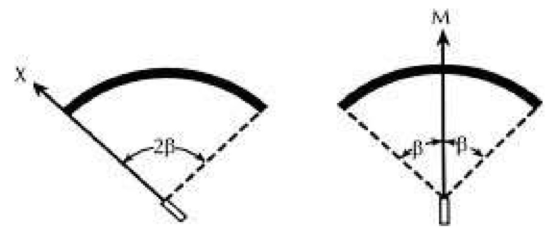


Figure 20-130. Exposure of X and M microphones to 0° pickup.

3. The M/S system picks up the S signal with a bidirectional microphone. The directivity performance of this type of microphone can be designed with a high degree of perfection, so errors in the S signal can be kept particularly small for all directions of sound incidence. The M/S system can supply a highly accurate side (S) signal.
4. In the M/S technique, the mono directivity does not depend on the amount of S signal applied to create the stereophonic effect. If recordings are made in the M/S format, a predictable mono signal is always captured. On the other hand, the stereophonic image can be simply influenced by modifying the S level without changing the mono signal. This can even be done during postproduction.

20.26.5 The Stereo Boom Microphone

Microphones for stereophonic television broadcasting have their own special problems. The optimal distance from the TV screen is considered to be five times the screen diagonal, at which distance the line structure of the TV image can no longer be resolved by the human eye. The resulting minimum observer distance is therefore 11ft (3.3m) and the two loudspeakers should be at least 12.5ft (3.8m) apart. This is certainly not realistic for television viewing.

The sound engineer must take into account that the reproduction will be through loudspeakers right next to the TV screen as well as through hi-fi equipment. If, for instance, the full base width is used during the sound recording of a TV scene, an actor appearing at the right edge of the picture will be heard through the right loudspeaker of the hi-fi system and sound as if he is far to the right of the TV set. This will result in an unacceptable perception of location.

The viewer must be able to hear realistically the talker on the TV screen in the very place where the viewer sees the talker. To achieve this goal, German television proposed to combine a unidirectional microphone for the recording of the actors with a figure-eight microphone for the recording of the full stereophonic basic width.

This recording technique utilizes a figure-eight microphone suspended from the boom in such a way that it maintains its direction when the boom is rotated while a second microphone with unidirectional pattern is mounted on top and follows the movement of an actor or reporter, Fig. 20-131. To make sure that the directional characteristic of the moving microphone does not have too strong an influence and that a slight angular error will not lead to immediately perceptible directional changes, the lobe should be somewhat wider as with the customary shotgun microphones in use today.

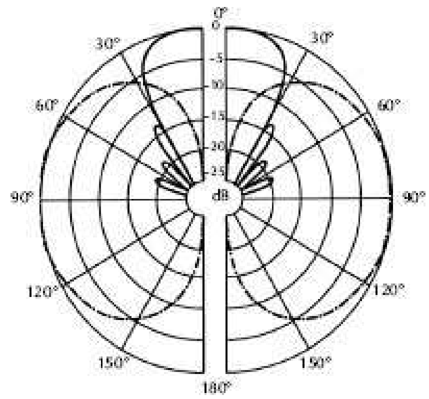


Figure 20-131. Stereo boom microphone using Sennheiser MKH 30 and MKH 70 microphones. Courtesy Sennheiser Electronic Corporation.

It is now possible to produce a finished stereo soundtrack on location by positioning the microphones in the manner of an M/S combination. The level of the figure 8 microphone can be lowered and used only for the recording of voices outside of the picture, ambience and music. This microphone should always remain in a fixed position and its direction should not be changed. The S-signal generated in this way must be attenuated to such a degree that the M-signal microphone will always remain dominant. This microphone is the one through which the actors pictured on the screen will be heard.

20.26.6 SASS Microphones

The Crown® SASS-P MK II or Stereo Ambient Sampling System™, Fig. 20-132, is a patented, mono-compatible, near-coincident array stereo condenser microphone using PZM technology.

Figure 20-132. Crown® SASS-P MK II stereo microphone. Courtesy of Crown Audio, Inc.

The SASS uses two PZMs mounted on boundaries (with a foam barrier between them) to make each microphone directional.

Controlled polar patterns and human-head-sized spacing between capsules create a focused, natural stereo image with no hole-in-the-middle for loudspeaker reproduction, summing comfortably to mono if required.

The broad acceptance angle (125°) of each capsule picks up ambient sidewall and ceiling reflections from the room, providing natural reproduction of acoustics in good halls and ambient environments. This pattern is consistent for almost ±90° vertical.

A foam barrier/baffle between the capsules shapes the pickup angle of each capsule toward the front, limiting overlap of the two sides at higher frequencies. Although the microphone capsules are spaced a few centimeters apart, there is little phase cancellation when both channels are combined to mono because of the shadowing effect of the baffle. While there are phase differences between channels, the extreme amplitude differences between the channels caused by the baffle, reduce phase cancellations in mono up to 20kHz.

The SASS has relatively small boundaries. However, it has a flat response down to low frequencies because there is no 6dB shelf as in standard PZM microphones (see section 20.12). The flat response is attained because the capsules are omnidirectional below 500Hz, and their outputs at low frequencies are equal in level, which, when summed in stereo listening, causes a 3dB rise in perceived level. This effectively counteracts one half of the low frequency shelf normally experienced with small boundaries.

In addition, when the microphone is used in a reverberant sound field, the effective low-frequency level is boosted another 3dB because the pattern is omnidirectional at low frequencies and unidirectional at high frequencies. All of the low-frequency shelf is compensated, so the effective frequency response is uniform from 20Hz–20kHz. Fig. 20-133 is the polar response of the left channel (the right channel is the reverse of the left channel).

20.27 Surround Sound Microphone System

20.27.1 Schoeps 5.1 Surround System

The Schoeps 5.1 surround system consists of the KFM 360 sphere microphone, two figure-eight microphones with suspension and the 24 bit DSP-4 KFM 360 processor, Fig. 20-134.

The central unit in this system is the sphere microphone KFM 360. It uses two pressure transducers and can, even without the other elements of the system, be used for stereophonic recording. Its recording angle is about 120°, permitting closer micing than a standard stereo microphone. The necessary high-frequency boost is built into the processor unit.

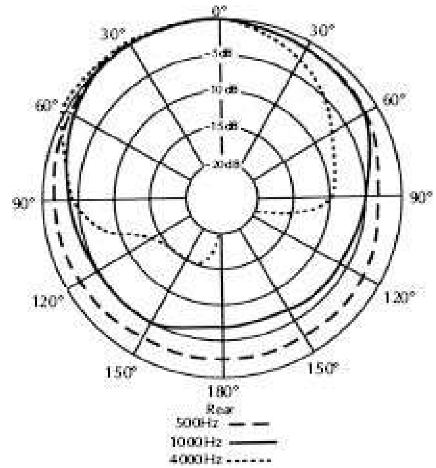


Figure 20-133. SASS-P MK II polar response, of the left channel. 0° sound incidence is perpendicular to the boundary. The right channel is a mirror image of the left channel. Courtesy Crown Audio, Inc.

Figure 20-134. Schoeps KFM 360 sphere microphone. Courtesy Schoeps GmbH.

Surround capability is achieved through the use of two figure-eight microphones, which can be attached beneath the pressure transducers by an adjustable, detachable clamp system with bayonet-style connectors. The two microphones should be aimed forward.

The DSP-4 KFM 360 processor derives the four corner channels from the microphone signals. A center channel signal is obtained from the two front signals, using a special type of matrix. An additional channel carries only the low frequencies, up to 70Hz. To avoid perceiving the presence of the rear loudspeakers, it is possible to lower the level of their channels, to delay them and/or to set an upper limit on their frequency response, Fig. 20-135.

The front stereo image width is adjustable and the directional patterns of the front-facing and rear-facing pairs of virtual microphones can be chosen independently of one another.

The processor unit offers both analog and digital inputs for the microphone signals. In addition to providing gain, it offers a high-frequency emphasis for the built-in pressure transducers as well as a low-frequency boost for the figure-eights.

As with M/S recording, matrixing can be performed during post-production in the digital domain.

The system operates as follows: the front and rear channels result from the sum (front) and difference (rear) of the omnidirectional and figure-eight microphones on each side, respectively, Fig. 20-136. The four resulting virtual microphones that this process creates will seem to be aimed forward and backward, as the figure-eights are. At higher frequencies they will seem to be aimed more toward the sides, i.e., apart. Their directional pattern can be varied, anywhere from omnidirectional to cardioid to figure-eight. The pattern of the two rear-facing virtual microphones can be different from that of the two forward-facing ones. Altering the directional patterns alters the sound as well, in ways that are not possible with ordinary equalizers. This permits a flexible means of adapting to a recording situation—to the acoustic conditions in the recording space—and this can even be done during postproduction, if the unprocessed microphones signals are recorded.

This four-channel approach yields a form or surround reproduction without a center channel—something that is not what everybody requires.

20.27.2 Holophone® H2-PRO Surround Sound System

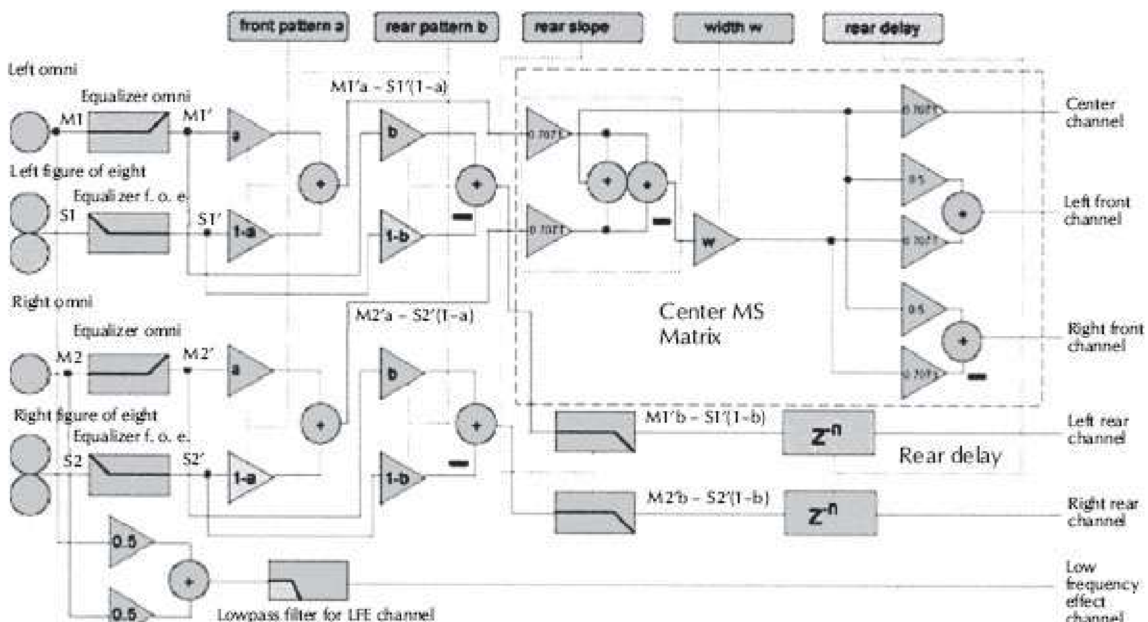


Figure 20-135. Schoeps DSP-4 KFM 360 processor. Courtesy Schoeps GmbH.

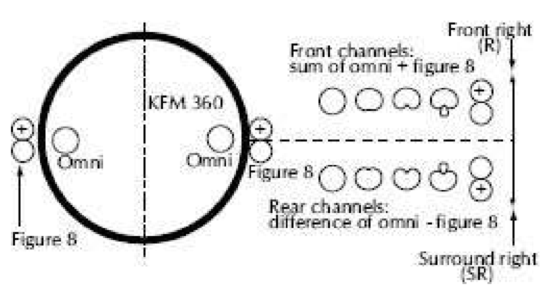


Figure 20-136. Derivation of the right (R) and right surround signals (SR) of the Schoeps 5.1 Surround System. Courtesy Schoeps GmbH.

Figure 20-137. Holophone H2-PRO 7.1 surround sound system. Courtesy Holophone®.

The elliptical shape of the Holophone® H2-PRO emulates the characteristics of a human head, Fig. 20-137. Sound waves bend around the H2-PRO as they do around a human head providing an accurate spatiality, audio imaging, and natural directionality. Capturing the directionality of these soundwaves translates into a very realistic surround sound experience. The total surface area of the eight individual elements combines with the spherical embodiment of the H2-PRO to capture the acoustic textures required for surround reproduction, Fig. 20-138. The embodiment acts as an acoustic lens capturing lows and clean highs.

A complete soundfield can be accurately replicated without the use of additional microphones—a simple point-and-shoot operation. The Holophone H2-PRO is capable of recording up to 7.1

channels of discrete surround sound. It terminates in eight XLR microphone cable ends (Left, Right, Center, Low Frequency, Left Surround, Right Surround, Top, and Center Rear). These co-relate to the standard 5.1 channels and add a top channel for formats such as IMAX and a center rear channel for extended surround formats such as Dolby EX, DTS, ES, and Circle Surround. Because each microphone has its own output, the engineer may choose to use as many or as few channels as the surround project requires as channel assignments are discrete all the way from the recording and mixing process to final delivery. It is well suited for television broadcasters (standard TV, DTV, and HDTV), radio broadcasters, music producers and engineers, film location recording crews, and independent project studios.

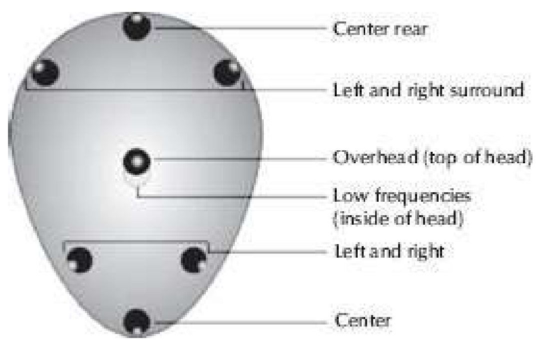


Figure 20-138. Location of the microphones on the H2-PRO head. Courtesy Holophone®.

Holophone H4 SuperMINI Surround Sound System

The H4 SuperMINI head, Fig. 20-139 contains six microphone elements that translate to the standard surround sound loudspeaker configuration; L, R, C, LFE, LS, RS. The LFE collects low-frequency signals for the subwoofer. The six discrete channels are fed into a Dolby® Pro-Logic II encoder which outputs the audio as a stereo signal from a stereo mini-plug to dual XLRs, dual RCAs, or dual mini-plugs. The left and right stereo signals can then be connected to the stereo audio inputs of a video camera or stereo recorder. The encoded signal is recorded onto the media in the camera or recorder and the captured audio can be played back in full 5.1-channel surround over any Dolby® Pro Logic II equipped home theater system. The material can be edited and the audio can be decoded via a Dolby® Pro Logic II Decoder and then brought into an NLE including Final Cut or iMovie, etc. The stereo recording can also be broadcast directly through the standard infrastructure. Once it is received by a home theater system containing a Dolby® Pro-Logic II or any compatible decoder, the six channels are completely unfolded to their original state. Where no home theater receiver is detected, the signal will simply be heard in stereo. The SuperMINI has additional capabilities that include an input for an external, center-channel-placed shotgun or lavalier microphone to enhance sonic opportunity options and features an audio zoom button that increases the forward bias of the pickup pattern. It also includes virtual surround monitoring via headphones for real-time on-camera 3D audio monitoring of the surround field.

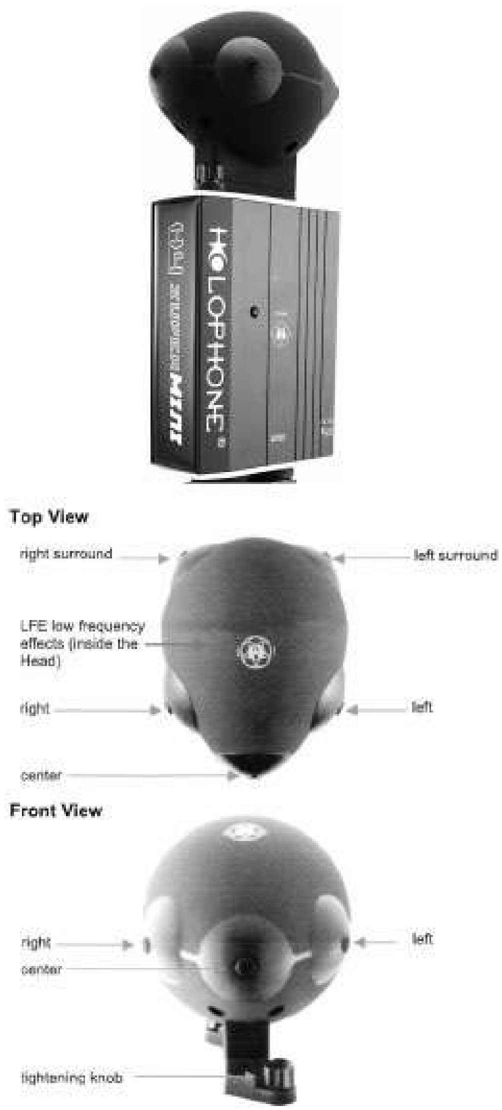


Figure 20-139. Microphone pattern for Holophone Super-Mini. Courtesy Holophone®.

20.28 Microphones for Binaural Recording

20.28.1 Artificial Head Systems

Human hearing is capable of selecting single sounds from a mixture of sounds while suppressing the unwanted components (the cocktail party effect). This is done in the listener's brain by exploiting the ear signals as two spatially separated sound receivers in a process frequently referred to as *binaural signal processing*. A simple test will verify this statement: when listening to a recording of several simultaneous sound events recorded by a single microphone, the individual sources cannot be differentiated.

Two spaced microphones or more elegant multi-element spatially sensitive microphones, such as a stereo coincident microphone, have been used to capture the spatial characteristics of sounds, but they have frequently been deficient when compared to what a person perceives in the same environment. This lack of realism is attributed to absence of the spectrum modification inherent in sound propagation around a person's head and torso and in the external ear, i.e., the transfer function of the human and the fact that the signals are kept separate until very late in the human analysis chain.

The acoustic transfer function of the human external ear is uniquely related to human body geometry. It is composed of four parts that can be modeled mathematically, as shown in Fig. 20-140, or recreated by an artificial head system.[15,16,17]

Reflections and diffraction of sound at the upper body, the shoulder, the head, and the outer ear (pinna), as well as resonances caused by the concha and ear canal, are mainly responsible for the transfer characteristic. The cavum concha is the antechamber to the ear. The spectral shape of the external ear transfer function varies from person to person due to the uniqueness of people and the dimensions of these anatomical features. Therefore, both artificial heads and their mathematical models are based on statistical composites of responses and dimensions of a number of persons.

All of these contributions to the external ear transfer function are direction sensitive. This means that sound from each direction has its own individual frequency response. In addition, the separation of the ears with the head in between affects the relative arrival time of sounds at the ears. As a result the complete outer-ear transfer function is very complicated, Fig. 20-141, and can only be partially applied as a correction to the response of a single or even a pair of microphones. In the figure, the base of each arrow indicates reference SPL. The solid curves represent the free-field (direct sound) external ear transfer function, while the dashed curves represent the difference, at each direction, relative to frontal free-field sound incidence.

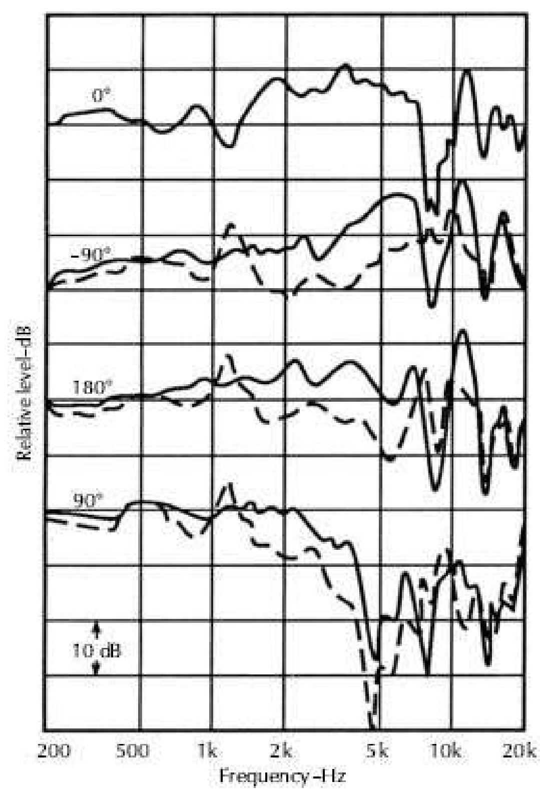


Figure 20-140. Transfer function of the left ear, measured 4 mm inside the entrance of the ear canal, for four angles of incidence (straight ahead, to the left, straight behind, and to the right).

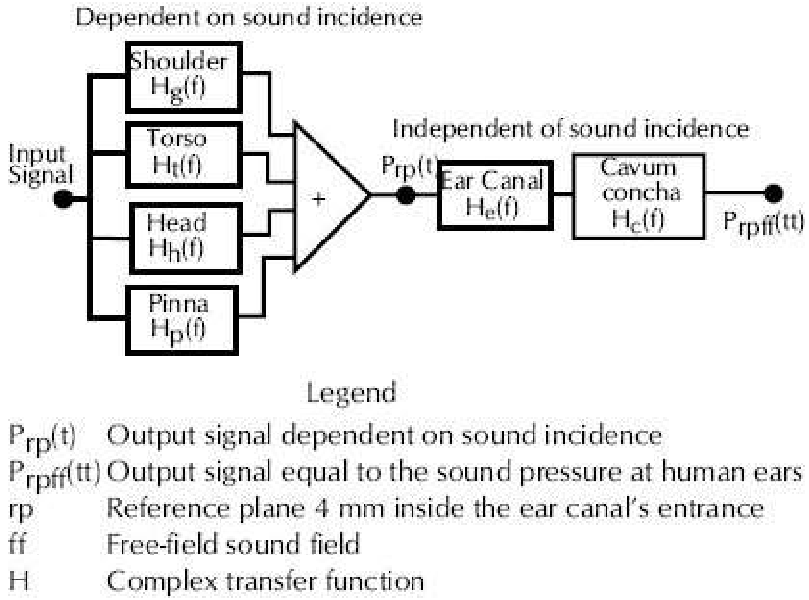


Figure 20-141. The human external-ear transfer function.

Artificial heads have been used for recording for some time. However, the latest heads and associated signal processing electronics have brought the state of the art close to "in the ear" (ITE) recording, which places microphones in human ears.

The KU 100 measurement and recording system by Georg Neumann GmbH in Germany is an example of a high-quality artificial head system, Fig. 20-142. Originally developed by Dr. Klaus Genuit and his associates at the Technical University of Aachen, the artificial head, together with carefully designed signal processing equipment, provides binaural recording systems that allow very accurate production of spatial imaging of complex sound fields.

Figure 20-142. Georg Neumann KU 100 dummy head. Courtesy Georg Neumann GmbH.

The head is a realistic replica of a human head and depends on a philosophy of sound recording and reproduction—namely, that the sound to be recreated for a listener should not undergo two transfer functions, one in the ears of the artificial head and one in the ears of the listener.

Fig. 20-143 is a block diagram of a head microphone and recording system. A high-quality microphone is mounted at the ear canal entrance position on each side of the head. Signals from each microphone pass through diffuse-field equalizers in the processor and are then available for further use in recording or reproduction. The diffuse-field equalizer is specifically tuned for the head to be the inverse of the frontal diffuse-field transfer function of the head. This signal is then recorded and can be used for loudspeaker playback and for measurement. The headphone diffuse-field equalizers in the Reproduce Unit yield a linear diffuse-field transfer function of the headphone, so the sound pressures presented at the entrance of the listener's ear canals will duplicate those at the entrance of the head's ear canals.

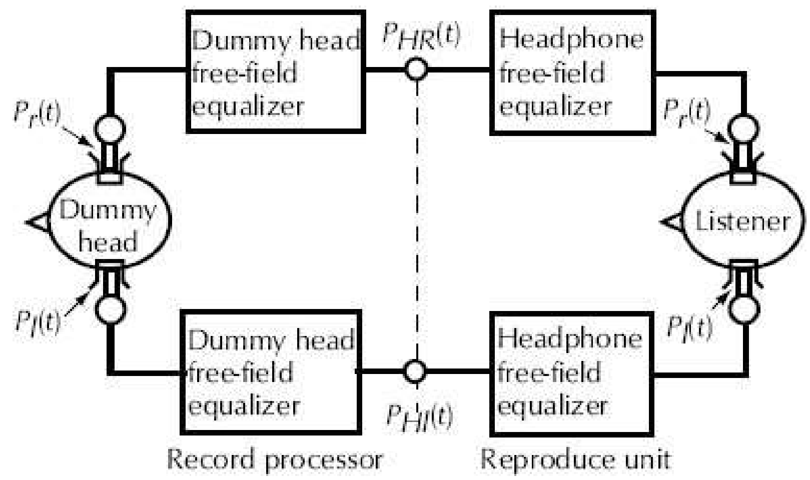


Figure 20-143. Block diagram of a dummy head binaural stereo microphone system.

Diffuse-field equalization is suitable for situations where the source is at a distance from the head. For recordings close to a sound source or in a confined space, such as the passenger compartment of an automobile, another equalization called *independent of direction* (ID) is preferred. The equalization is internal in the head in Fig. 20-143.

Signals $P_{Hr}(t)$ and $P_{Hl}(t)$ from the heads can be recorded and used directly for loudspeaker playback, analysis, or playback through headphones. As a recording tool this method can surpass many other recording techniques intended for loudspeaker reproduction. The full benefits of spatial imaging can be heard and enjoyed with earphone playback as well as with high quality loudspeakers.

The heads are constructed of rugged fiberglass. The microphones can be calibrated by removal of the detachable outer ears and applying a pistonphone. Preamplifiers on the microphones provide polarization and have balanced transformerless line drivers. A record processor and modular unit construction provides dc power to the dummy head and act as the interface between the head and the recording medium or analysis equipment. The combination of low noise electronics and good overload range permits full use of the 135dB dynamic range of the head microphones and 145dB with the 10dB attenuator switched in.

For headphone playback, a reproduce unit provides an equalized signal for the headphones that produces earcanal entrance sound signals that correspond to those at the corresponding location on the artificial head.

An important parameter to consider in any head microphone recording system is the dynamic range available at this head signal output. For example, the canal resonance can produce a sound pressure that may exceed the maximum allowed on some ear canal-mounted microphones.

20.28.2 In-the-Ear Recording Microphones

In-the-Ear (ITE™)[18] recording and Pinna Acoustic Response (PAR™) playback represent a new-old approach to the recording of two channels of spatial images with full fidelity and their playback over two channels, Fig. 20-144. It is important that the loudspeakers are in signal synchronization and that they be placed at an angle so that the listener position is free of early reflections.

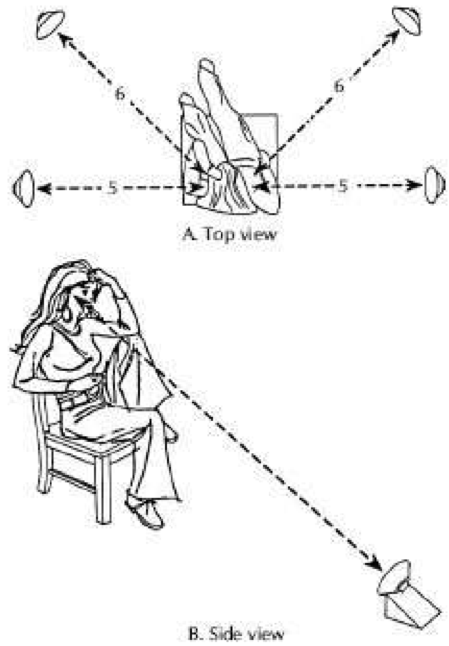


Figure 20-144. The loudspeaker arrangement used for PAR playback of ITE recordings. Courtesy SynAudCon.

Low noise, wide frequency response, and dynamic range probe microphones employing soft silicone probes are placed in the pressure zone of the eardrum of live listeners. This microphone system allows recording with or without equalization to compensate for the ear canal resonances while leaving the high-frequency comb filter spatial clues unaltered. The playback system consists of synchronized loudspeaker systems spaced approximately equal distances from the listener in the pattern shown in Fig. 20-144. Both left loudspeakers are in parallel, and both right loudspeakers are in parallel. However, the front and back loudspeakers are on individual volume controls. This is to allow balancing side-to-side and to adjust the front-to-back relative levels for each individual listener. The two front loudspeakers are used to provide hearing signals forward of the listener.

Fig. 20-145A shows an ETC made in a listening room $(L_D - L_R =$

0.24). Fig. 20-145B is the identical position measured with the ITE technique ($L_D - L_R = 5.54$). Note particularly the difference in $L_D - L_R$ for the two techniques. ITE recording and PAR playback allow a given listener to hear a given speech intelligibility environment as perceived by another person's head and outer ear configuration right up to the eardrum.

Recordings made using ITE microphones in two different people's ears of the same performance in the same seating area sound different. Playback over loudspeakers where the system is properly balanced for perfect geometry for one person may require as much as 10dB different front to back balance for another person to hear perfect geometry during playback.

Since ITE recordings are totally compatible with normal stereophonic reproduction systems and can provide superior fidelity in a many cases, the practical use of ITE microphony would appear to be unlimited.

20.29 USB Microphones

The computer has become an important part of sound systems. Many consoles are digital and microphones are connected directly to them. Microphones are also connected to computers through the USB input.

The audio-technica AT2020 USB cardioid condenser microphone, Fig. 20-146 is designed for computer-based recording. It includes a USB (Universal Serial Bus) digital output that is Windows and Mac compatible. The sample rate is 44.1kHz with a bit depth of 16 bits. The microphone is powered directly from the 5Vdc USB output.

The MXL.006 USB is a cardioid condenser microphone with a USB output that connects directly to a computer without the need for external mic preamps through USB 1.1 and 2.0, Fig. 20-147.

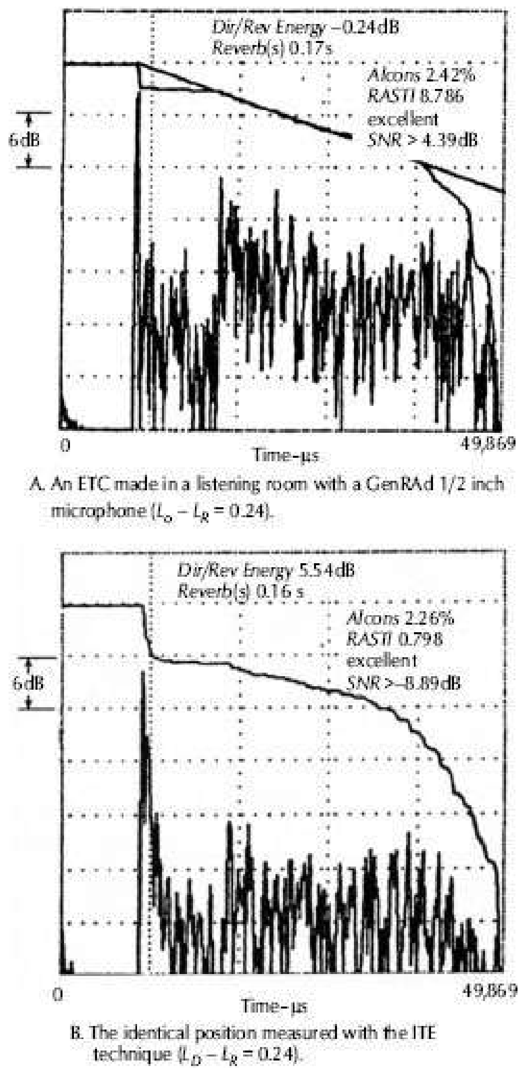


Figure 20-145. An ETC comparison of a measurement microphone and the ITE technique at the same position in the room. Courtesy SynAudCon.

The analog section of the MXL.006 microphone features 20Hz–20kHz frequency response, a 22mm, 6 micron sputtered gold diaphragm, pressure-gradient condenser capsule, and a three-position, switchable attenuation pad. The digital section features a 16-bit Delta Sigma A/D converter with a sampling rate of 44.1kHz and 48kHz.

The MXL.006 includes a red LED behind the protective grill to inform the user that the microphone is active and correctly oriented.

20.30 Wireless Communication Systems

Figure 20-146. Audio-Technica AT2020 USB microphone. Courtesy Audio-Technica US, Inc.

Figure 20-147. MXL.006 USB microphone. Courtesy Marshall Electronics, Inc.

Wireless communication systems are wireless microphones (*radio microphones*), Fig. 20-148, and a related concept, wireless intercoms. Often the same end user buys both the microphones and intercoms for use in television and radio broadcast production, film production, and related entertainment-oriented applications.

Wireless microphone systems can be used with many of the preceding microphones discussed. Some wireless microphone systems include a fixed microphone cartridge while others allow the use of cartridges by various manufacturers.

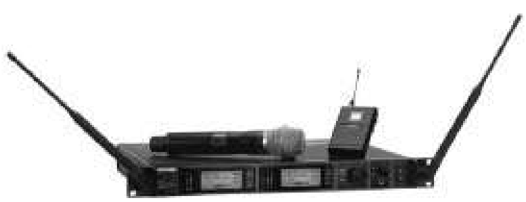


Figure 20-148. Shure UHF-R Wireless Microphone System. Courtesy Shure Incorporated.

20.30.1 Wireless Analog Microphone Systems

A block diagram of aa analog wireless microphone system is shown in Fig. 20-149. The sending end of a wireless microphone system has a dynamic, condenser, electret, or pressure zone microphone connected to a preamplifier, compressor, and a small transmitter/modulator and antenna.

The receiving end of the system is an antenna, receiver/discriminator, expander, and preamplifier, which is connected to external audio equipment.

20.30.2 Criteria for Selecting a Wireless Microphone

There are a number of criteria that must be considered in obtaining a wireless microphone system suitable for professional use.[19,20] Ideally, such a system must work perfectly and reliably in a variety of tough environments with good intelligibility and must be usable near strong RF fields, lighting dimmers, and sources of electromagnetic interference. This relates directly to the type of modulation (standard frequency modulation or narrow-band frequency modulation), the operating frequency, high frequency (HF), very high frequency (VHF), ultrahigh frequency (UHF), the receiver selectivity, etc. The system must be very reliable and capable of operating at least five hours on one set of disposable batteries (or on one recharge if Ni-Cads are used).

20.30.2.1 Frequency Bands of Operation

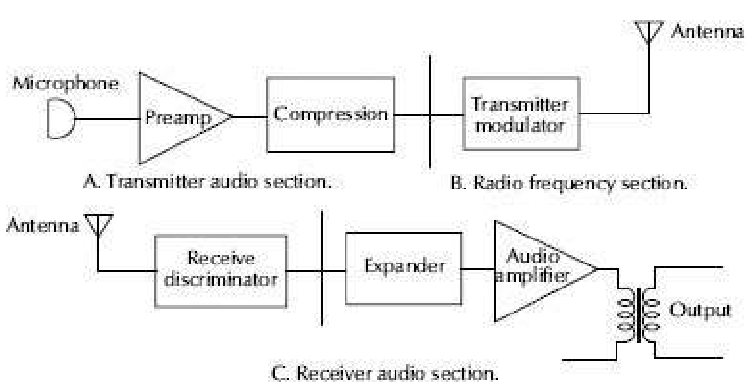


Figure 20-149. Wireless microphone transmitter section with built-in preamplifier, compressor, and transmitter, and the receiver with a built-in discriminator expander.

Based on the FCC's reallocation of frequencies and the uncertainty of current and future allocations, some wireless manufacturers are offering systems that avoid the VHF and UHF bands completely. The ISM (industrial, science, and medicine) bands provide a unique alternative to the TV bands. By international agreement, all devices are low powered so there will never be any grossly high-powered RF interference potential. The 2.4GHz band provides a viable alternative to traditional UHF bands, and as long as line of sight between transmitters and receivers is monitored users can easily get a 100 meter range. Another benefit of 2.4GHz is that it can simplify wireless inventory for traveling shows. The same wireless frequencies are accepted worldwide, so there is no need to adhere to the country-specific frequency rules that severely complicate the situation for international tours. The same applies within the United States—the same frequencies work in all areas.

Currently wireless microphones are licensed on several frequencies, the most common being:

• VHF low band (AM and FM)	25 to 50MHz
	72 to 76MHz
• FM broadcast (FM)	88 to 108MHz
• VHF high band (FM)	150 to 216MHz
• UHF (FM)	470 to 746MHz
	902 to 952MHz
• UHF (FM) banned	698–806MHz

The VHF bands are seldom used today and can only be found on old equipment. The low band is in the noisiest radio spectrum and, because the wavelength is about 20ft (6m) it requires a long antenna (5ft or 1.5m). The VHF low band is susceptible to skip, which can be defined as external signals from a long distance away bouncing off the ionosphere back to earth, creating interference.

The VHF high band is more favorable than the low band. The 1/4-wavelength antenna is only about 17in (43cm) long and requires little space. The VHF band has some penetration through buildings that can be advantageous and disadvantageous. It is advantageous in being able to communicate between rooms and around surfaces. It is disadvantageous in that transmission is not controlled (security), and outside noise sources can reach the receiver.

Most often the frequencies between 174MHz and 216MHz are used in the VHF band, corresponding to television channels 7 to 13. The VHF high band is free of citizens band and business radio interference, and any commercial broadcast stations that might cause interference are scheduled so you know where they are and can avoid them. Inherent immunity to noise is built in because it uses FM modulation. Better VHF high-band receivers have adequate selectivity to reject nearby commercial television or FM broadcast signals. If operating the microphone or intercom on an unused television channel—for instance Channel 7—protection might be required against a local television station on Channel 8. Another problem could be caused by an FM radio station. If a multi–thousand watt FM station is broadcasting near a 50mW wireless microphone, even a well-suppressed second harmonic can have an RF field strength comparable to the microphone or intercom signal because the second harmonic of FM 88MHz is

176MHz, which is in the middle of television Channel 7. The second harmonic of FM 107MHz is 214MHz, which is in the middle of Channel 13. Thus, if a VHF wireless system is to be utilized fully, especially with several microphones or intercoms on adjacent frequencies, the wireless receiver must have a very selective front end.

One television channel occupies a 6MHz wide segment of the VHF band. Channel 7, for example, covers from 174–180MHz. A wireless intercom occupies about 0.2MHz (200kHz). By FCC Part 74 allocation, up to 24 discrete VHF high-band microphones and/or intercoms can be operated in the space of a single television channel. In order to use multiple systems on adjacent frequencies, the wireless microphone/intercom receivers must be very selective and have an excellent capture ratio. On a practical basis, this means using narrow-deviation FM (approximately 12kHz modulation). Wide-deviation systems (75kHz modulation or more) can easily cause interference on adjacent microphone/intercom frequencies; such systems also require wide bandwidth receivers that are more apt to be plagued by interference from adjacent frequencies. The trade-off between narrowband FM and wideband FM favor wideband for better overall frequency response, lower distortion, and inherently better *SNR* versus maximum possible channels within an unused TV channel for equal freedom from interference (max. 6). Poorly designed FM receivers, are also subject to *desensing*. Desensing means the wireless microphone/intercom receiver is muted because another microphone, intercom, television station, or FM station (second harmonic) is transmitting in close proximity; this limits the effective range of the microphone or intercom.

The UHF band equipment is the band of choice and is the only one used by manufacturers today. The wavelength is less than 3ft (1m) so the antennas are only 9in (23cm). The range is not as good as VHF, because it can sneak through small openings and can reflect off surfaces more readily.

All of the professional systems now are in the following UHF bands:

- A band 710–722MHz.
- B band of 722–734MHz.
- 728.125–740.500MHz band.

The FCC has assigned most of the DTV channels between channel 2 and 51, and only four channels between 64 and 69, which is where most of the professional wireless microphones operate.

20.30.2.2 Adjustment of the System's Operating Frequency

Many of the professional wireless microphones are capable of being tuned to many frequencies. In the past the systems were fixed frequency, often because that was the only way they could be made stable. With PLL-synthesized channels (Phase Lock Loop), it is not uncommon for systems to be switch tunable to 100 different frequencies in the UHF band and have a frequency stability of 0.005%. This is especially important with DTV in the scene.

20.30.2.3 Capture Ratio and Muting

Capture ratio and muting specifications of the receiver are important. The capture ratio is the ability of the receiver to discriminate between two transmitters transmitting on the same frequency. When the signal is frequency modulated (FM), the stronger signal controls what the receiver receives. The capture ratio is the difference in the signal strength between the capturing transmitter and the captured transmitter that is blanketed. The lower the number, the better the receiver is at capturing the signal. For instance, a receiver with a capture ratio of 2dB will capture a signal that is only 2dB stronger than the second signal.

Most systems have a muting circuit that squelches the system if no RF signal is present. To open the circuit, the transmitter sends a special signal on its carrier that breaks the squelch and passes the audio signal.

20.30.2.4 RF Power Output and Receiver Sensitivity

The maximum legal RF power output of a VHF high-band microphone or intercom transmitter is 50mW; most deliver from 25–50mW. Up to 120mW is permissible in the business band (for wireless intercoms) under FCC part 90.217, but even this represents less than 4dB more than 50mW. The FCC does not permit the use of high-gain transmitter antennas, and even if they did, such antennas are large and directional so they would not be practical for someone who is moving around. Incidentally, high-gain receiving antennas are also a bad idea because the transmitter is constantly moving around with the performer so much of the received radio signal is actually caught on the bounce from walls, props, and so on.

Even if an offstage antenna is aimed at the performer, it probably would be aiming at the wrong target. Diversity receiving antenna systems, where two or more antennas pick up and combine signals to feed the receiver, will reduce dropouts or fades for fixed receiver installations.

The received signal level can't be boosted, given the restrictions on antenna and transmitted power, so usable range relies heavily on receiver sensitivity and selectivity (i.e., capture ratio and *SNR*) as well as on the audio dynamic range. In the pre-1980 time frame, most wireless microphones and intercoms used a simple compressor to avoid transmitter overmodulation. Today, systems include compandor circuitry for 15–30dB better audio *SNR* without changing the RF *SNR*. This is achieved by building a full-range compressor into the microphone or intercom transmitter, and then providing complementary expansion of the audio signal at the receiver—much like the encoder of a tape noise-reduction system. The compression keeps loud sounds from overmodulating the transmitter and keeps quiet sounds above the hiss and static. The expander restores the loud sounds after reception and further reduces any low-level hiss or static. Companding the audio signal can provide from 80–85dB of dynamic range compared to the 50–60dB of a straight noncompanded transmit/receive system using the same deviation.

20.30.2.5 Frequency Response, Dynamic Range, and Distortion

No wireless microphone will provide flat response from 20Hz–20kHz, nor is it really needed. Wireless or not, by the time the audience hears the broadcast, film, or concert, the frequency response has probably been reduced to a bandwidth from 40Hz–15kHz. Probably the best criteria for judging a handheld wireless microphone system is to compare it to the microphone capsule's naked response. If the transmit/receive bandwidth basically includes the capsule's bandwidth, it is enough. Generally speaking, a good wireless microphone should sound the same as a hard-wired microphone that uses the same capsule. Wireless intercom systems, because they are primarily for speech communication, are less critical with regard to audio bandwidth; 300Hz–3kHz is telephone quality, and 50Hz–8kHz is excellent for an intercom.

Dynamic range is probably the most critical aspect of performance for natural sound. A good compandor system will provide 80–85 dB of dynamic range, assuming the microphone is adjusted to 100% modulation on the loudest sounds. Leaving a margin of safety by turning down the microphone modulation level sacrifices *SNR*. Even with extra headroom and a working *SNR* of 75dB, the microphone will still have about twice the dynamic range of a typical optical film sound track or television show.

The system should provide at least 40–50dB SNR with a 10μV signal and 70–80dB SNR with an 80μV signal. This shows up when no audio signal is being transmitted.

When an electret condenser microphone is used, a major limitation in dynamic range can be the capsule itself, not the wireless system. Typically, an electret powered by a 1.5V battery is limited to about 105dB SPL. Powered by a 9V battery, the same microphone may be usable to 120dB SPL. The wireless microphone system should be able to provide a high enough bias voltage to ensure adequate dynamic range from the microphone capsule. Although the condenser may be hotter in output level than a dynamic microphone, its background noise level is disproportionately higher, so the overall *SNR* specification may be lower.

Wireless intercom systems do not need the same dynamic range as a microphone. They do not have to convey a natural musical performance. However, natural dynamics are less fatiguing than highly compressed audio, especially given a long work shift. So aside from greater range, there are other benefits to seeking good *SNR* in the intercom: 40dB or 50dB would be usable, and 60dB or 70dB is excellent. An exception is in a very high-noise industrial environment, where a compressed loud intercom is necessary to overcome background noise. A good intercom headset should double as a hearing protector and exclude much of the ambient noise.

Distortion is higher in a wireless system than in a hard-wired system—a radio link will never be as clean as a straight piece of wire. Still, total harmonic distortion (THD) specifications of less than 1% overall distortion are available in today's better wireless microphones. In these microphones, one of the largest contributors to harmonic distortion is the compandor, so distortion is traded for *SNR*. The wireless intercom can tolerate more THD, but lower distortion will prevent fatigue and improve communication.

20.30.3 Receiving Antenna Systems

RF signal dropout or multipath cancellation is caused by the RF signal reflecting off a surface and reaching a single receiver antenna 180° out-of-phase with the direct signal, Fig. 20-150. The signal can be reflected off surfaces such as armored concrete walls, metal grids, vehicles, buildings, trees, and even people.

Although you can often eliminate the problem by experimenting with receiver antenna location, a more foolproof approach is to use a space diversity system where two or more antennas pick up the transmitted signal, as shown in Fig. 20-151. It is highly unlikely that the obstruction or multipath interference will affect two or more receiver antennas simultaneously.

There are three diversity schemes: *switching diversity, true diversity*, and *antenna combination*.

- **Switching Diversity**. In the switching diversity system, the RF signals from two antennas are compared, and only the stronger one is selected and fed to one receiver.
- **True Diversity**. This receiving technique uses two receivers and two antennas set up at different positions, Fig. 20-152. Both receivers operate on the same frequency. The AF signal is taken from the output of the receiver that at any given moment has the stronger signal at its antenna. The probability of no signal at both antennas at the same time is extremely small. The advantages of diversity compared to conventional RF transmission are shown in Fig. 20-153. Only the receiving chain with the better input signal delivers audio output. Not only does this system provide redundancy of the receiving end, but it also combines signal strength, polarity and space diversity.

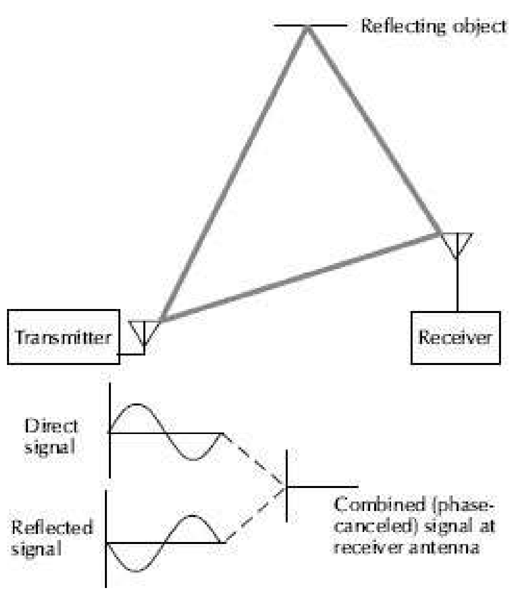


Figure 20-150. Phase cancellation of radio frequency signals due to reflections.

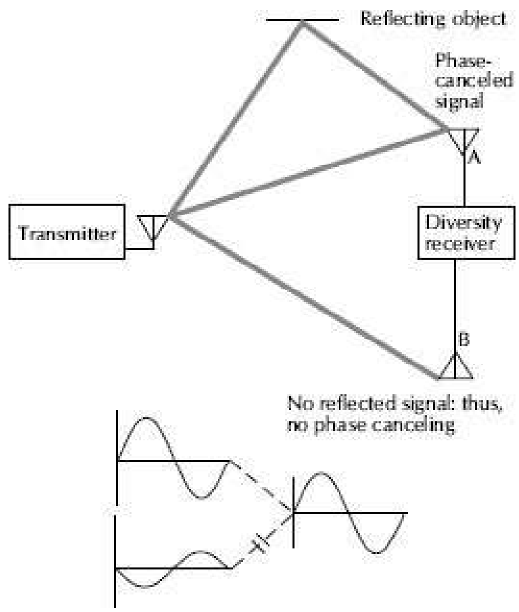


Figure 20-151. Diversity antenna system used to reduce multipath radio frequency phase cancellation.

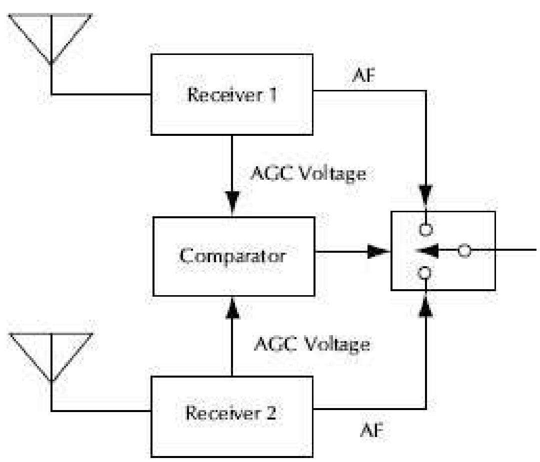


Figure 20-152. Functional diagram of a true diversity receiver.

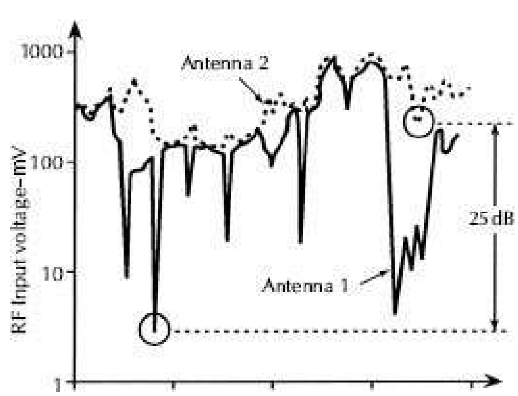


Figure 20-153. Effect of switch-over diversity operation. Solid line RF level at antenna 1 and the dotted line RF level at antenna 2. Courtesy Sennheiser Electronic Corporation.

- **Antenna Combination Diversity**. The antenna combination diversity system is a compromise of the other methods. This system uses two or more antennas, each connected to a wideband RF amplifier to boost the received signal. The signals from both receiving antennas are then actively combined and fed to one standard receiver per microphone. In this way, the receiver always gets the benefit of the signals present at all antennas. There is no switching noise, no change in background noise, and only requires one receiver for each channel. A drawback is the possibility of complete signal cancellation when phase and amplitude relationships due to multipath provide the proper unfavorable conditions.

Antenna Placement

It is often common to use a near antenna and a far antenna. The near antenna, which is the one nearest the transmitter, produces the majority of the signal most of the time; in fact, it may even be amplified with an in-line amplifier. The far-field antenna may be one or more antennas usually offset in elevation and position; therefore, the possibility of dropout is greatly reduced. Because the antennas are common to all receivers, many wireless microphones can be used at the same time on the same antenna system. This means that there are fewer antennas and a greater possibility of proper antenna placement.

The following will generally prevent dead spots:

- Do not set up antennas in niches or doorways.
- Keep the antennas away from metal objects including armored concrete walls. Minimum distance: 3ft (1m).
- Position the antennas as close as possible to the point where the action takes place.
- Keep antenna cables short to keep RF losses at a minimum. It is better to use longer AF leads instead. Note: If long runs of antenna cable are used, be sure they are of the low-loss type.
- Make a walkaround test, i.e., operate the transmitter at all positions where it will be used later. Mark all points where field strength is weak. Try to improve reception from these points by changing the antenna position. Repeat this procedure until the optimum result is achieved.

Interference is mainly caused by spurious signals arriving at the receiver input on the working frequency. These spurious signals may have various causes:

- Two transmitters operating on the same frequency (not permissible).
- Intermodulation products of a multichannel system whose frequencies have not been selected carefully enough.
- Excessive spurious radiation from other radio installations, e.g., taxi, police, CB-radio, etc.
- Insufficient interference suppression on electric machinery, vehicle ignition noise, etc.
- Spurious radiation from electronic equipment, e.g., light control equipment, digital displays, synthesizers, digital delays, computers, etc.

20.30.4 Companding

Two of the biggest problems with using wireless microphones are *SNR* and dynamic range. To overcome these problems, the signal is compressed at the transmitter and expanded at the receiver. Figs. 20-149 and 20-154 graphically illustrate how and what this can accomplish with respect to improving the *SNR* and reducing the susceptibility to low-level incidental FM modulation, such as buzz zones. The transmitter has an audio gain of 40dB and the receiver has an audio reduction of 40dB, therefore any noise originating in the transmission of the signal is also reduced 40dB in the receiver. Note that the −80dB signal has not been altered while the −20dB signal has been altered significantly.

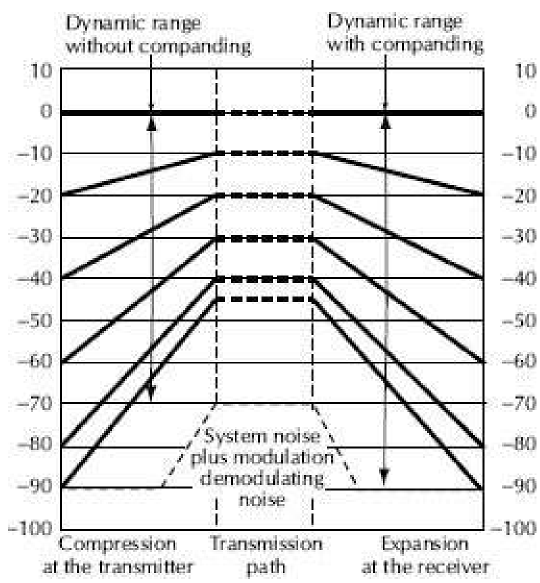


Figure 20-154. Compression and expansion of the audio signal.

As the typical input level changes by a factor of 80dB, the audio output to the modulator undergoes a contoured compression, so a change in input audio level is translated into a pseudo-logarithmic output. This increases the average modulation level, which reduces all forms of interference encountered in the transmission medium.

By employing standard narrowband techniques at the receiver, the recovered audio is virtually free of adjacent channel and spurious response interference. In addition, up to ten times the number of systems can be operated simultaneously without cross-channel interference. The ability of the receiver to reject all forms of interference is imperative when utilizing expansion and compression techniques because the receiver must complimentarily expand the audio component to restore the original signal integrity.

20.30.5 Digital Wireless Microphone Systems

Lectrosonics L Series Wireless Microphone Systems

The Lectrosonics L Series microphone systems have a tuning range of three standard Lectrosonics blocks, or 67.5 to 76.8MHz, depending on the specific frequency band, and employ Lectrosonics' patented Digital Hybrid Wireless® technology for compandor-free audio along with compatibility modes for interoperability with any previous Digital Hybrid units along with older analog systems. They have 25 or 100kHz tuning steps (yielding up to 3,072 selectable frequencies). The new L Series product group consists of the LMb and LT beltpack transmitters, and the LR miniature receiver, Figs. 20-155 and 20-156.

Figure 20-155. Lectrosonics L series microphones. Courtesy Lectrosonics, Inc.

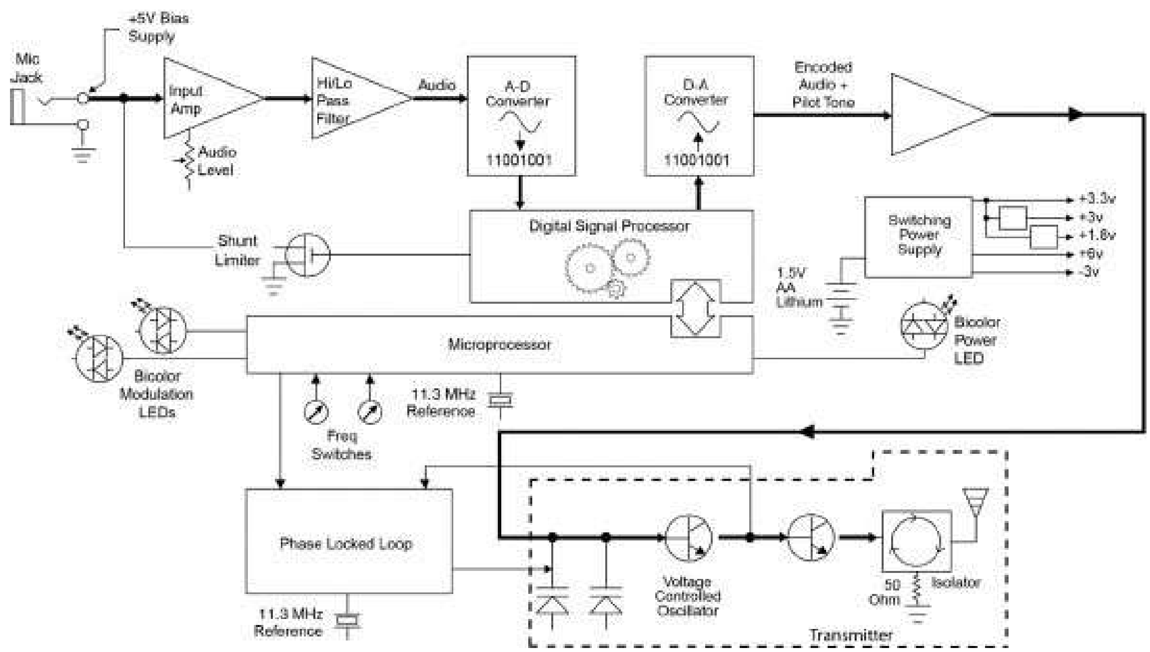


Figure 20-156. Block diagram of the Lectrosonics DSW wireless system transmitter. Courtesy Lectrosonics, Inc.

The LMb is an economical beltpack transmitter designed for applications including theater, house of worship, TV, ENG, video and film production. It features an all-metal housing, wire belt clip, 50mW of RF power, and an integrated multi-function switch with menu-selectable modes for power, mute or talkback. It incorporates a graphic LCD and membrane switch panel plus IR synch. Power is provided via two AA batteries.

The LT transmitter features an integrated multi-function switch with menu-selectable modes for power, mute or talkback, and a graphic LCD and membrane switch panel plus IR synch port for ease of setup and operation. It has a user-selectable RF transmission power of 50 or 100mW, and a detachable antenna allowing for remote mount antennas. The LT responds to remote commands from either the dedicated Lectrosonics RM remote unit or smartphone apps with this capability. It has two inputs, a Lectrosonics microphone input, and a user-selectable 1MΩ input.

The LR receiver is designed to be ultra-portable for use with DSLR, 4/3, small 4K, and other compact HD cameras. The wide tuning bandwidth with tracking front-end filters provides flexibility while avoiding the vulnerabilities of wider pass bands. Compatibility modes enable the LR to operate with older analog transmitters. Dual antenna diversity improves range and resistance to dropouts. The RF spectrum analyzer and Lectrosonics' SmartTune capability simplifies finding clean frequencies on site. A large, backlit LCD provides instant, clear information for setup and status monitoring while operational. Detachable antennas allow for remote antenna mounting. The unit is powered with two AA batteries.

The L Series units are available in four different frequency ranges for worldwide use:

- A1 (470.100–537.575MHz).
- B1 (537.600–614.375MHz).

- C1 (614.400–691.175MHz).
- D1 (691.200–767.975MHz).

Overall system audio specifications are:

- Digital conversion: 24 bit/88.2kHz sampling rate.
- Frequency response: 40 to 20kHz ±1dB.
- THD+N = 0.3% in Digital Hybrid mode.
- > 95dB SNR ratio.
- Audio output levels are adjustable from −50 to +5dBu in 1dB steps.

20.30.6 Digital Secure Wireless Microphone Systems

Wireless microphone systems are often used in environments where privacy is of paramount concern, such as closed film & TV sets, corporate boardrooms, shareholder meetings, and government facilities. In such cases, a method of protecting the transmitted information is very important. To assure protection a digital wireless system using digital modulation and encryption is often used.

Lectrosonics DSW system

The Lectrosonics Patented Digital Hybrid Wireless® technology (U.S. Patent No. 7,225,135) DSW (Digital Secure Wireless) system, Fig. 20-157, uses a proprietary algorithm to encode digital audio information into analog format which can be transmitted over an analog FM wireless link. It utilizes AES (Advanced Encryption Standard) encryption technology employing a 256-bit key, which was approved as a government standard encryption algorithm in 2001 [FIPS 197]. The CTR (Counter) Encryption mode in the DSW maintains low latency—2.5ms overall—and keeps channel noise at a minimum. At the receiver, the encoded signal is captured and decoded back to the original digital audio. This combination offers the high audio quality of a pure digital system and the operating range of high quality FM wireless systems. The digital audio chain eliminates a compandor and its artifacts, and provides audio frequency response flat to 20kHz. The RF link takes advantage of the spectral efficiency characteristics of a highly optimized FM radio system. Key management is user selectable in one of two modes: "persistent", where the key is kept from session to session, and "one-time", where the key must be generated for each session which is the most secure. The system utilizes a chip for entropy generation in order to ensure the key is generated in a truly random manner—thus complying with another government standard [FIPS 140-2]. The key transfer requires a simple cabled download, avoiding security issues with easily detected infrared key transfer systems.

Figure 20-157. Lectrosonics DSW wireless system. Courtesy Lectrosonics, Inc.

The DSW system consists of three pieces, the DR digital wireless receiver frame which can hold six DRM receiver modules, individual DRM digital receiver modules, and the DB digital wireless beltpack transmitter.

Analog and digital (AES/EBU) XLR outputs are selectable on the device menu. The frame supports wideband reception (470.100 to 691.175MHz), 50Ω BNC antenna inputs, and outputs for cascading up to 3 additional frames (24 channels total) on one set of antennas without an external multi-coupler. The clock input and output enables the DR to be the master clock in a digital audio system or to be the slave to an external master clock. A 1/4 inch headphone output can be fed from a mixture of channel signals, or from isolated channels.

The DB digital wireless beltpack transmitter has wideband tuning (470–698MHz), a linear RF output stage for reduced intermodulation distortion and high channel counts with minimum RF spectrum available. The transmitter provides 50mW transmission RF. The TA5M mic/line input accepts all lavaliere and headworn microphones wired for Lectrosonics transmitters.

Specifications are:

- Audio frequency response: 20Hz to 20kHz ±1dB.
- Digital conversion: 24 bits, 48kHz sampling.
- System latency: 2.5ms.
- Distortion: 0.05% THD+N, 1kHz @–10dBFS.
- Dynamic range: 108dB A-wtd.

20.30.7 Waterproof Wireless Microphone Systems

Wireless microphones that are worn are very useful for coaching all forms of athletics including swimming and aquatic aerobics. If the instructor always stays on the pool deck, a weatherproof system might be adequate. If the instructor is in the water, a completely submersible and waterproof system will be required.

Hydrophonics assembles a completely waterproof and submersible wireless microphone system. Assembled with Telex components, the system includes a headset microphone with a special waterproof connector and a Telex VB12 waterproof beltpack transmitter. The transmitter can operate on a 9V alkaline battery or a 9V NiMH rechargeable battery. The rechargeable battery is recommended as it does not require removing the battery from the transmitter for recharging and therefore reduces the chance of water leaking into the transmitter housing. The receiver is a Telex VR12 for out-of-pool operation, and can be connected to any sound system the same way as any other wireless microphone.

An interesting thing about this system is you can dive into the water while wearing the system and come up and immediately talk as the water drains out of the windscreen rapidly.

The DPA Type 8011 hydrophone, Fig. 20-158, is a 48V phantom powered waterproof microphone specially designed to handle the high sound pressure levels and the high static ambient pressure in water and other fluids. The hydrophone uses a piezoelectric sensing element, which is frequency compensated to match the special acoustic conditions under water. A 10m high-quality audio cable is vulcanized to the body of the hydrophone and fitted with a standard three-pin XLR connector. The output is electronically balanced and offers more than 100dB dynamic range. The 8011 hydrophone is a good choice for professional sound recordings in water or under other extreme conditions where conventional microphones would be adversely affected.

Figure 20-158. DPA 8011 hydrophone. Courtesy DPA Microphones A/S.

20.31 Multichannel Wireless Microphone and Monitoring Systems

By Joe Ciaudelli

Wireless microphones are ubiquitous in our society. They are critical tools for content creation (e.g. movies, TV, radio), news gathering, live stage performance, sports and political events. They are also routinely used in schools, houses of worship, government facilities convention centers and corporate offices. The definition of wireless mics includes in-ear monitoring systems as well as intercom and cue systems.

Demand for wireless audio systems has been fueled by the increasing volume of content being created, higher sophistication of productions, a trend towards greater mobility on stage, the desire to control volume and equalization of individual performers, the speed required to disseminate news, and the safety that lack of cables can provide. Consequently, applications in which the number of wireless microphones, referred to as channels, being used simultaneously has increased dramatically. Note, the term channels in this context should not to be confused with TV channels in which wireless audio systems often operate within. Productions and events with large multichannel systems, greater than thirty channels, are now common. Systems of this magnitude pose a difficult engineering challenge. Careful planning, installation, operation, and maintenance are required.

Wireless systems use electromagnetic waves, instead of a cable or fiber, to transport audio information. Electromagnetic waves are produced by the vibration of charged particles. For reasons of practicality, audio systems generally use the radio frequency (RF) portion of the electromagnetic spectrum.

A wireless system requires a transmitter and complementary receiver, each having a tuned antenna, to process sound via radio frequency (RF) transmission. First, the transmitter processes the audio signal and superimposes it on a RF carrier through a process called modulation. The transmit antenna then acts as a launch pad for the modulated carrier, and broadcasts the signal over the air. The signal travels a certain space or distance and reaches the pick-up element, which is the receiving antenna. Finishing the operation, the receiver—which selects the desired carrier—strips off the signal through demodulation, processes it, and finally reconstitutes the original audio signal, Fig. 20-159. Each wireless channel operating in the same area at the same time needs to be on a unique carrier frequency.

20.31.1 Audio Signals

Acoustic sound sources are captured by a microphone element, which transforms the mechanical sound wave to an electrical signal that the transmitter can then process. Alternately other input signals, such as from electric instruments, can be fed into the transmitter.

Classically wireless microphone systems used these signals in their original analog form to modulate the carrier frequency. An analog signal is continuous and mimics the shape of a wave.

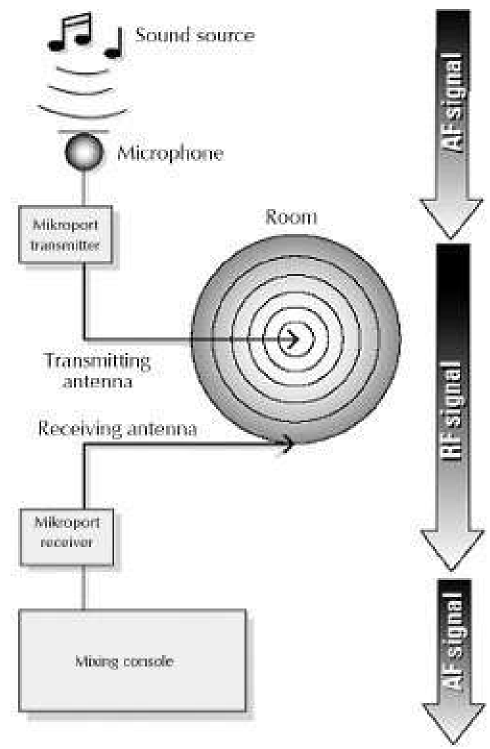


Figure 20-159. Signal path of a Mikroport system. Courtesy Sennheiser Electronic Corporation.

As time goes on, more digital transmission systems are employed. A digital signal is a stream of discrete numbers that represent instantaneous amplitudes of an analog signal, measured at equally spaced points in time. There are some advantages to transmitting a digitized version of an audio signal instead of its original analog form, noise being a primary one. Since analog signals can assume any value, any noise introduced is interpreted as being part of the original signal. Digital systems, on the other hand, can only understand two states numerically represented by zero and one respectively. Anything in between is disregarded. This property allows:

- Simple storage without degradation over time.
- Transmission without degradation.
- Digital Signal Processing (DSP).

Conversion of sound from the analog to the digital domain is done through the process of sampling. It must be understood that digital is an approximation. To reduce the degree of approximation, many samples of the analog signal must be measured within a short interval of time. The sampling rate must be at least twice the highest desired audio frequency. 20,000 cycles per second (20kHz) is the generally accepted maximum frequency for human hearing perception. Therefore, a sampling rate of 44.1kHz or higher is generally used for digital systems with full audio frequency response.

However, there are some trade-offs with digital systems, mainly latency, the delay between the input of a signal and its output. Each conversion between analog and digital domain takes finite time (typical for A/D and D/A: 0.1–1.5ms). Any processing, such as compression, adds more latency.

Clocking is also crucial. Digital audio is based on sampling rate and needs precise clock regeneration or master system clock for multiple devices, and/or accurate sample rate conversion.

Digital Audio Representation

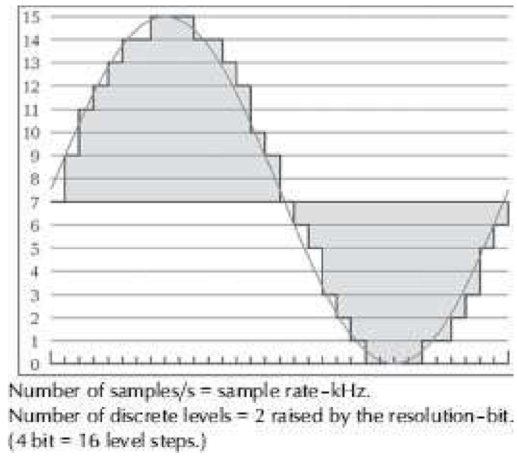


Figure 20-160. D/A conversion. Courtesy Sennheiser Electronic Corporation.

The A/D converter takes samples of the analog signal at discrete intervals. Based on its magnitude, each sample will be quantized to the closest available number value, Fig. 20-160.

20.31.2 Modulation

Modulation is basically adding a "code" to a carrier frequency by making a detectable change. Transmitting a signal requires three main processes:

- Generation of a pure carrier frequency.
- Modulation of that carrier with the information to be transmitted. Any reliably detectable change in carrier characteristic could translate information.
- Detection at the receiver of the signal modification in the transmitter and reconstruction of the information, also known as demodulation.

The characteristics of a carrier signal, Fig. 20-161, that can be modified over time to convey information are:

- Amplitude.
- Frequency.
- Phase.

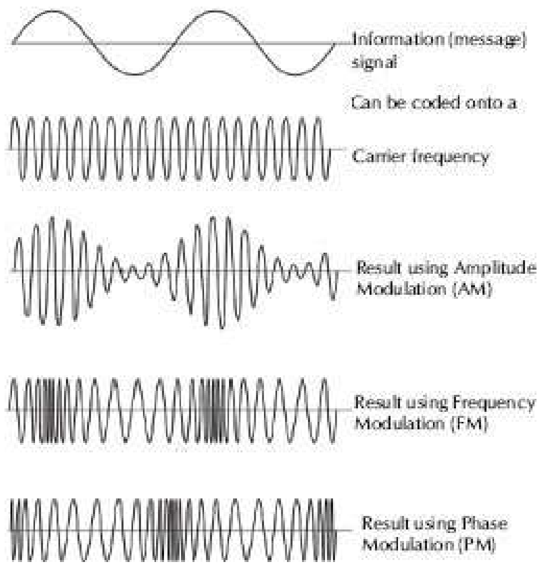


Figure 20-161. Classic analog modulation techniques. Courtesy Sennheiser Electronic Corporation.

20.31.2.1 Modulation – Analog

Amplitude Modulation (AM). In AM the frequency of the carrier is kept constant and its amplitude is changed in proportion to the instantaneous amplitude of the modulating message signal.

- Minimum required bandwidth (BW) is twice the maximum audio frequency (AF)

$$BW = 2 \times AF_{max} \qquad (20\text{-}31)$$

- Maximum signal to noise ratio (*SNR*) ≈ 30dBA (affected by fading).
- AF response limited by occupied *BW*.

- *SNR* dependent on signal strength and modulation depth.

Since the information is coded on the amplitude of the carrier, anything that degrades the carrier will adversely impact the desired message signal. For example, a lightning storm causes significant audible interference.

20.31.2.2 Frequency Modulation (FM)

In FM the amplitude of the modulated carrier is kept constant, while the frequency is varied in proportion to the instantaneous amplitude of the modulating information signal.

- Minimum required bandwidth equals twice the frequency deviation added to the maximum audio frequency:

$$BW_{min} = 2 \times (\Delta f + AF_{max}) \qquad (20\text{-}32)$$

- $SNR_{max} \approx 50$dBA (for typical $\Delta f = \pm 50$kHz) without AF processing, ≥100dB with dynamic processing.
- Audio Frequency (ΔF) response dependent on *BW*.
- *SNR* dependent on deviation Δf and received signal strength.

Since none of the message signal is coded in the amplitude of the carrier, FM is more robust and less susceptible to varying climatic conditions. This is the reason FM has classically been used for high fidelity personal and car radios. Likewise, an FM mic transmitter works like a miniature FM radio station.

FM Audio Processing. To improve the audio quality, several measures are necessary because of the inherent noise of the RF link. These techniques provide improvement of dynamic range and *SNR* of the audio signal:

Pre- and De-Emphasis. This method is a static measure which is used in most of the FM transmissions. By increasing the level of the higher audio frequencies on the transmitter side, the signal-to-noise ratio is improved because the desired signal is above the inherent noise floor of the RF link, Fig. 20-162.

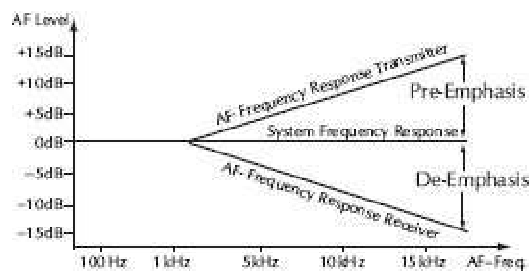


Figure 20-162. Pre-Emphasis and de-emphasis curves. Courtesy Sennheiser Electronic Corporation.

Companding. The compander is a synonym for “compressor” on the transmitter side, and is also a synonym for “expander” on the receiving end. The compressor raises low audio level above the RF noise floor. The expander does the mirror opposite and restores the audio signal. This measure increases the signal-to-noise ratio to CD quality level. All audio processing has consequences which are more evident with certain types of audio sources. A low level, short duration, high frequency sound, such as scissors closing, can activate the compander and a “breathing” sound can be heard which is the opening and closing of the compander circuit. Most often this is masked by typical audio input, Fig. 20-163.

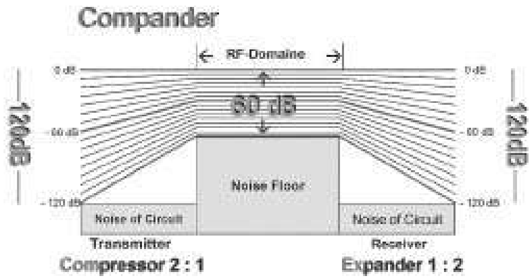


Figure 20-163. Companding curves. Courtesy Sennheiser Electronic Corporation.

Frequency Deviation. The modulation of the carrier frequency in an FM system greatly influences its audio quality. An increase in deviation yields better high frequency response and the dynamic range. The trade-off is that fewer channels can be used within a frequency range. Systems can be categorized as wideband and narrowband. A typical wideband system suitable for high fidelity music can have a peak deviation of 56kHz. If such a system, tuned to a 525.000MHz carrier, has a loud input signal, it could modulate the carrier between 524.944MHz and 525.056MHz. In contrast, a communications grade system, like a walkie-talkie may only have a deviation of 5kHz around its center carrier. So wideband allows higher quality audio but requires a larger portion of spectrum (occupied bandwidth).

Occupied Bandwidth. The total occupied bandwidth of spectrum is largely defined by the amount of information required to transmit. A U.S. TV channel is 6MHz wide (8MHz in Europe). The volume of information sent by a high definition TV signal can occupy the whole six MHz channel. Alternately, up to four standard definition stations could share a single channel. If a channel is not being used in an area for TV broadcast, it is ideal to place wireless mics in that locally vacant TV channel. Since they are not transmitting picture or color information, and the FCC limits the bandwidth of mics to 200kHz, about eight wideband FM mics could operate within a 6MHz TV channel. If only narrowband devices are required, more than twice could be packed into that channel.

Therefore a high fidelity signal (CD quality sound) such as a mic for music applications uses a larger portion of spectrum than a low fidelity device such as a walkie-talkie (limited frequency response and dynamic range). A simple analogy can be made with shopping bags (carriers) that are placed in a trunk of a car (the trunk could represent a TV channel: a six Megahertz block of spectrum). Big bags (wideband deviation) each hold more groceries (audio information), but fewer bags will fit in the trunk which is a fixed size. Many more smaller bags (narrowband deviation) will fit in the trunk but each bag will have less groceries (information).

20.31.2.3 Phase Modulation

Phase and frequency are closely related since they can be considered different ways to view or measure the same signal change. Frequency modulation (FM) is that form of angle modulation in which the cycles (completed alternations) per second are varied by the information signal. Phase modulation (PM) is that form of angle modulation in which the angle is varied by the information signal. FM has generally been used for analog transmission. Phase is more often used in digital systems.

20.31.2.4 Modulation—Digital Systems

The promise and potential advantages of converting the natively analog audio signal to digital data prior to transmission can be summarized to:

- Lower noise.
- A compander circuit can be avoided, providing highly transparent sound without breathing effects.
- More flexibility for audio processing.
- For applications where data compression/reduction is acceptable, higher spectral efficiency.
- Transmission security (encryption).
- Error correction capabilities.
- Additional features through software/firmware updates.

The main driving force remains BEST POSSIBLE AUDIO QUALITY! (SNR, frequency response, immunity to interference).

20.31.2.5 Digital Modulation

With analog modulation systems, the information signal is analog, a continuous wave with infinite degrees of variation within the signals dynamic range. With digital modulation, the information signal is binary with discreet values. Technically "Digital Modulation" is a misnomer and refers to an analog carrier modulated with digital data (digital representation of an audio signal). Any modulation always changes the analog properties of the carrier. Both systems use analog carriers.

In the most basic digital message, the presence of a signal represents a binary value of "1." Absence of the signal represents a "0."

Simple digital modulation schemes include Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), and Phase Shift Keying (PSK) Fig. 20-164.

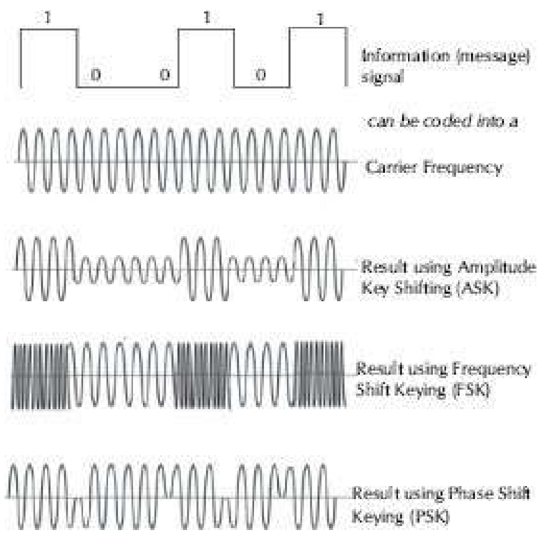


Figure 20-164. Simple binary digital modulation. Courtesy Sennheiser Electronic Corporation.

Many properties need to be considered for a digital wireless microphone (or any other radio communications device):

- Transmitter power (emitted from the antenna).
- Range.
- Tolerance for radio-frequency noise.
- Data rate (throughput in bits/second).
- Bit error rate (fraction of bits wrongly received).
- Occupied bandwidth (in Hz).

When a system operates near its limit, any of these properties can be improved, at least in principle, *but only at the expense of others*. Reduction of occupied bandwidth, for example, requires raising the transmitter power, reducing the operating range, reducing the noise tolerance, reducing the throughput data rate, and/or accepting more errors. These limitations are fundamental to the nature of information itself, much as the conservation of energy is fundamental to physics. In the same way that conservation of energy rules out a useful perpetual-motion machine, information theory rules out a useful wireless microphone that operates in a reduced bandwidth without sacrificing other qualities.

Many of the characteristics listed above are outside the microphone manufacturer's control. Transmitter power is limited by battery life (and government rules). Needed range is set by the layout of the performance venue. Needed noise tolerance depends on the local radio-frequency environment. Data rate depends mathematically on the audio quality required (subject to compression). The acceptable error rate likewise depends on quality requirements. Radio bandwidth, as we have seen, depends on all the others.

Other services that gained spectrum efficiency when converting to digital modulation (such as cell phones and broadcast TV) did not achieve those gains from digitization as such, but from implementing compression. The two are often confused because digitization is a prerequisite to efficient compression. But the processes are distinct. And even compression does not provide an end-run around information theory. Compression reduces the bit rate, but it also impairs audio quality or adds latency. So a balance of compromises must be made.

Regulations for wireless microphones in many frequency bands restrict the maximum occupied bandwidth to 200kHz. The minimum bandwidth required for transmission is equal to the bit rate (flow of data). Data Rate/Bandwidth = bits/Hz of BW or bits/Symbol.

Typical today is A/D conversion with 24 bit resolution at a sampling rate of 96kHz, resulting in bit rate of 2.304MBit/s. To facilitate reliable transmission additional data is required for framing and coding, necessary for control and synchronization, thereby expanding the data rate by a factor of around 1.5 and resulting in gross bit rate of 3.45MBit/s. This would require over 17 bits/s/Hz, a currently insurmountable feat given the limit of the 200kHz mask.

However, superior audio quality can be achieved with a frequency response up to 20kHz and a dynamic range (*SNR*) of >100dB with an A/D conversion of 18 bits of resolution and with a sampling rate of 44.1kHz, resulting in 793.4kbit/s. Factoring the required framing and coding, yields a data rate of 1.2Mbit/s.

Simple modulation schemes like ASK, FSK, PSK offer only a fraction of the needed data rate within the permissible bandwidth (typical ≤ 150–200kbit/s). One possible option could be further processing of the digital data to result in significant data reduction or compression. Any such data processing may affect the ultimate audio quality and inevitably will introduce additional processing delays (‘latency’).

A more sophisticated approach is a complex modulation technique that yields greater bits/s/Hz. In digital communication, the modulated parameter takes on only a discrete set of values, each of which represents a symbol. This symbol may consist of one or more bits, or binary ones and zeros. Since the demodulator must merely identify which amplitude, frequency and phase state is most closely represented in the received signal during each symbol period, the signal can be regenerated without any distortion.

For example, simple binary phase shift keying only transmits 1 bit of information representing either a “1” or a “0”. However, phase shifting of a carrier is not restricted to only two states of 0° and 180°. With Quadra-Phase Shift Keying (QPSK), four states are used, corresponding to 45°, 135°, 225°, and 315°. 2 bits are associated with each state, thus doubling the transmitted information, Fig. 20-165.

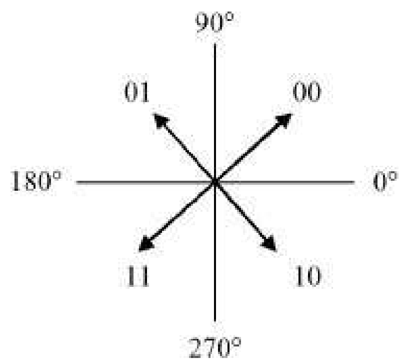


Figure 20-165. Quadra-Phase Shift Keying (QPSK). Courtesy Sennheiser Electronic Corporation.

QAM

It is also possible to combine phase shift keying and amplitude keying in a form of vector modulation known as Quadrature Amplitude Modulation, or QAM, Fig. 20-166.

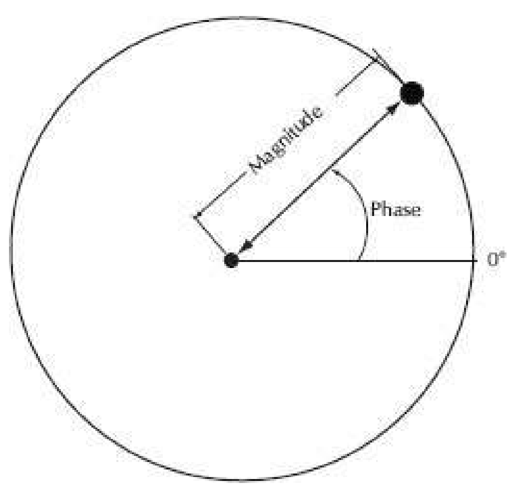


Figure 20-166. Digital (Vector) Modulation. Courtesy Sennheiser Electronic Corporation.

QAM is commonly generated as a signal in which two carriers shifted in phase by 90° are modulated. The resultant sum output includes both magnitude and phase variations. The advantage of QAM is that it is a modulation of higher order, and as a result it is able to carry more bits of information per symbol. By selecting a higher order format of QAM, the data rate of a link can be increased. Table 20-3 gives a summary of the bit rates of different forms of QAM and PSK.

A simple way to view specific conditions of magnitude and phase is made possible with constellation diagrams. Constellation diagrams show the different positions for the end points of vectors from corresponding polar diagrams. Shown here are representations for different forms of Quadrature Amplitude Modulation (QAM). As the order of the modulation increases, so does the number of points (CP) on the QAM constellation diagrams, Fig. 20-167.

Table 20-3. Bit Rates of QAM and PSK

Modulation	Bits Per Symbol	Symbol Rate
BPSK	1	1× bit rate
QPSK	2	1/2 bit rate
8PSK	3	1/3 bit rate
16QAM	4	1/4 bit rate
32QAM	5	1/5 bit rate
64QAM	6	1/6 bit rate

An alternate way to depict the constellation diagram for 64QAM shows all combinations of 4 amplitude levels with 16 discrete phase conditions, Fig. 20-168.

If each transmitted symbol represents 6 bits (64QAM), a 200kHz wide channel can accommodate the gross data rate of 1.2Mbit/s required for digital signals with 18 bits of resolution at a sampling rate of 44.1kHz. The goal of uncompressed high fidelity audio with frequency response up to 20kHz and dynamic range (*SNR*) of >100dB is thus met!

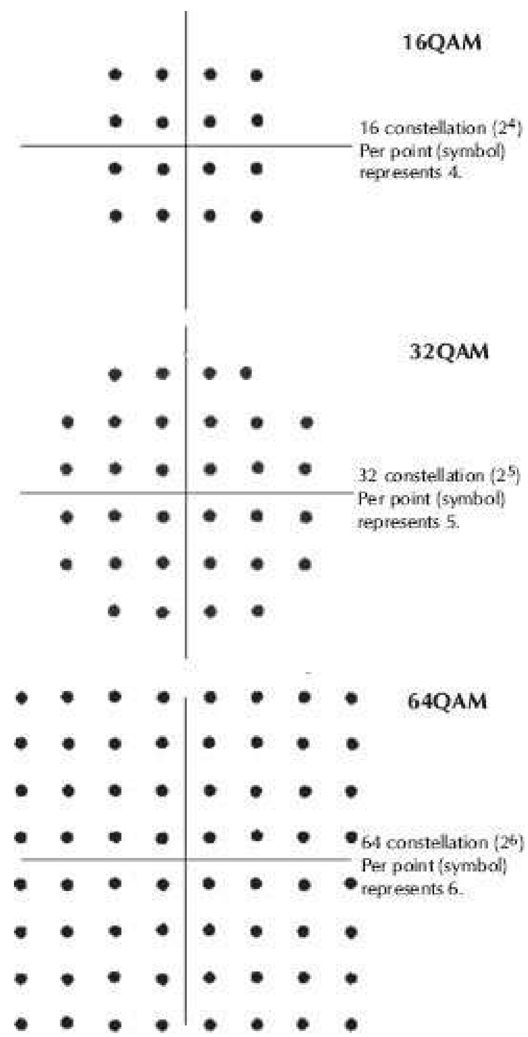


Figure 20-167. QAM constellation diagrams. Courtesy Sennheiser Electronic Corporation.

By comparison, uncompressed digital HD Video plus audio and formatting requires a data rate of ≈ 20Mbit/s. With 4 bits/symbol

(BQAM) this signal can be transmitted over the air within the 6MHz bandwidth of a TV-channel. This shows how the transmission of high quality uncompressed audio, considering the regulatory restraints, presents a greater technical challenge than transmission of a High Definition television signal (1080i HD video plus 5+1 audio).

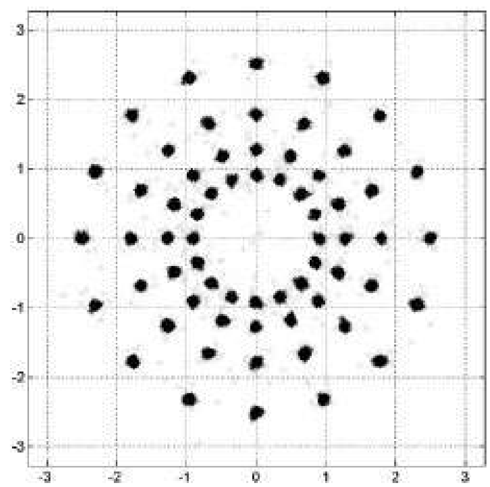


Figure 20-168. 64QAM constellation diagram. Courtesy Sennheiser Electronic Corporation.

Reviewing some properties of this modulation technique:

- Constellation points (CP) are defined by discrete amplitude and phase.
- The number of CPs determines how many bits are defined by a symbol.
- The number of symbols per unit of time is limited by the RF channel BW.
- The CP is determined at specific instances of time (symbol time).

While higher order modulation rates are able to offer much faster data rates and higher levels of spectral efficiency for the radio communications system, this comes at a price. The more bits that are represented by each symbol, the less bandwidth is required, but with a greater likelihood for bit errors and subsequent need for more transmit power. The higher order modulation schemes are considerably less resilient to noise and interference. The modulated RF Carrier (Vector) goes continuously from one constellation point to the other according to the bit sequence (symbol) to be sent. Noise and interference in the received signal make it more difficult to distinguish individual constellation points and decode the corresponding bit sequence. Lower order modulation schemes result in lower bit rate, fewer constellation points and require less carrier- to-noise/carrier-to-inter-ference ratio (CNR/CIR) for acceptable bit error rate (BER). Figs. 20-169 and 20-170 compare the simpler QPSK with the 64QAM modulations.

20.31.2.6 Carrier Frequencies

Manufacturers have classically produced wireless microphones that can tune to a frequency within the TV band. These systems can operate on locally vacant channels (not used for over-the-air TV broadcast), often called "white spaces," per specifications outlined by government, for example, in the United States the Federal Communications Commission (FCC) regulates use of RF frequencies.

When wireless microphones first became widespread in the 1980s, upper-band Very High Frequencies (VHF) (174–216MHz) were used. Between 1990–2010, Ultra High Frequencies (UHF) ranging from 470 to 806MHz were predominantly used. The wavelength of the carrier is inversely proportional to its frequency. Higher frequencies have shorter wavelengths, Antenna size is largely determined by the wavelength of the desired carrier. Therefore UHF systems can use smaller antennas compared to VHF, generally considered a big advantage for mic applications. UHF frequencies (450–960MHz) have a wavelength of less than one meter. They have excellent reflective characteristics. They can travel through a long corridor, bouncing off the walls, losing very little energy. They also require less power to transmit the same distance compared to much higher frequencies, such as microwaves. These excellent wave propagation characteristics and low power requirements make UHF ideal for performance applications.

During 2010 in the United States the TV portion of the UHF was reallocated to accommodate broadband telecom services. The spectrum available for microphone operation was thus truncated to 470–698MHz. There are plans for further repacking of the spectrum. This will result in less UHF spectrum available for mics. Each nation has their own regulations so the specific frequencies available for operation vary from country to country. However, reallocation of the UHF band is the global trend. This is due to the high demand for the favorable wave propagation that the UHF range offers, such as penetration through foliage and walls. Also small antennas can be used and good range (distance) is achieved even with low transmitter RF output power (≤50mW in the case of microphones). These advantages have spurred governments to monetize the spectrum, away from free over-the-air TV broadcast (and use by wireless microphones) to telecom companies willing to pay billions of dollars for the rights to exclusive use of desired frequency bands.

Equipment is also available in frequency bands such as 902–

928MHz and 2.4GHz which have been designated for unlicensed operation. A wide variety of non-mic wireless devices also use these bands. In areas where there is a concentration of such devices there is greater potential for interference. Even higher frequencies have been used for some applications using Ultra Wideband transmission techniques for applications such as wireless conference systems.

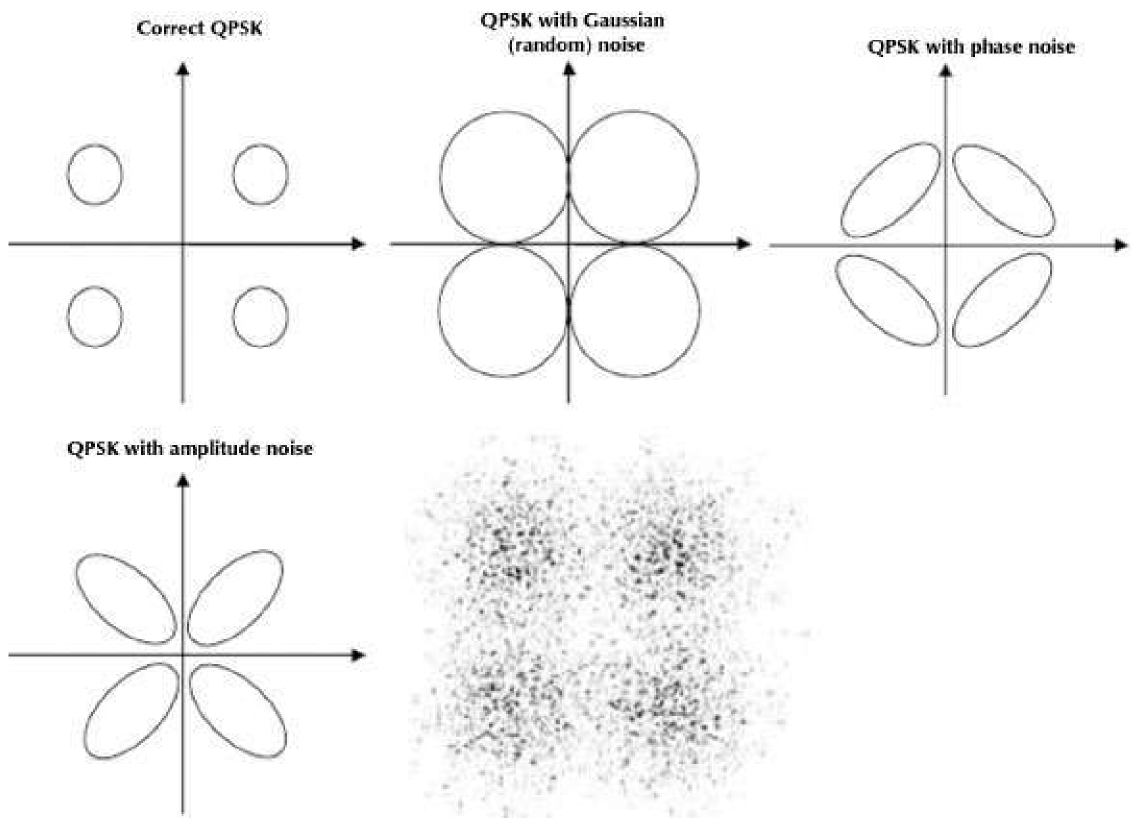


Figure 20-169. Noise in a QPSK system. Courtesy Sennheiser Electronic Corporation.

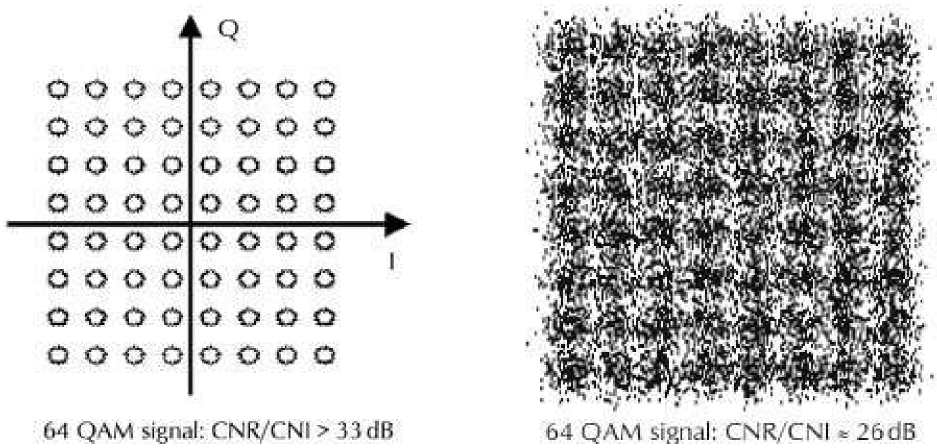


Figure 20-170. 64 QAM signal to CNR/CNI ratio. Courtesy Sennheiser Electronic Corporation.

There is a probability that alternate ranges will become available to mic operators to help alleviate the loss of access to UHF. However, it is likely they will be available on a shared basis through the use of a priority database.

20.31.2.6.1 Spacing

In order to have a defined channel, without crosstalk, a minimum spacing of 300kHz between carrier frequencies is generally employed for classic FM designs. A wider spacing is even more preferable since many receivers often exhibit desensitized input stages in the presence of closely spaced signals. A wider minimum spacing of 400kHz should be used for in-ear monitors when operated in stereo. Certain digital systems with ultra-high linearity that allow equally spaced carriers (discussed below) use 500kHz spacing.

Caution should be used when linking receivers with widely spaced frequencies to a common set of antennas. The frequencies need to be within the bandwidth of the antennas as well as any filtered antenna boosters and distribution systems.

20.31.2.6.2 Frequency Coordination

Multichannel wireless microphone systems can be especially difficult to operate, as they present several extraordinary conditions. Multiple transmitters moving around a stage together with body absorption, shadowing and polarization effects will result in wide variations of field strength seen at the receiver antenna system. This is even more challenging in a touring application since the RF conditions vary from venue to venue. In this case, the mix of frequencies is constantly changing. The daunting task to coax each of these variables makes equipment and frequency selection highly critical.

An important issue is to avoid interference from intermodulation (IM) products. Generally, one can coordinate around them or use top quality equipment that is properly configured so IM products will not be generated.

Coordinating Around IM Products. Intermodulation is the result of two or more signals mixing together, producing harmonic distortion. It is a common misconception that intermodulation is produced by the carrier frequencies mixing within the air. Intermodulation occurs within active components, such as transistors, exposed to strong RF input signals. When two or more signals exceed a certain threshold, they drive the active component into a non-linear operating mode and inter-modulation (IM) products are generated. This usually happens in the RF section of the receiver, in antenna amplifiers, or the output amplifier of a transmitter. In multichannel operation, when several RF input signals exceed a certain level the intermodulation products grow very quickly. There are different levels of inter-modulations defined by the number of addition terms.

In any wireless system with three or more frequencies operating in the same range, frequency coordination is strongly advised. It is necessary to consider possible IM frequencies which might interfere with the desired audio transmission. The 3rd and 5th harmonic, in particular, might raise interference issues. Considering two fundamental (wanted) signals the following signals may be present at the output of a non-linear stage:

- Fundamentals: F1 and F2.
- Second Order: 2F1, 2F2, F1±F2, F2-F1.
- Third Order: 3F1, 3F2, 2F1±F2, 2F2±F1.
- Fourth Order: 4F1, 4F2, 2F1±2F2, 2F2±2F1.
- Fifth Order: 5F1, 5F2, 3F1±2F2, 3F2±2F1.
- Additional higher orders….

As a result, the intermodulation frequencies should not be used, as those frequencies are virtual transmitters. The general rule “never use two transmitters on the same frequency” is valid in this case. However, even-order products are far removed from the fundamental frequencies and, for simplicity, are therefore omitted from further considerations. Signal amplitude rapidly diminishes with higher order IM-products, and with contemporary equipment design, consideration of IM-products can be limited to 3rd and 5th order only.

For multichannel applications (e.g., 30+ channels), the intermodulation products can increase significantly and the calculation of intermodulation-free frequencies can be done by number crunching software. By looking only at the third harmonic distortion in a multichannel system, the number of third order IM-products generated by multiple channels is:

- 2 channels result in 2.
- 3 channels result in 9.
- 4 channels result in 24.
- 5 channels result in 50.
- 6 channels result in 90.
- 7 channels result in 147.
- 8 channels result in 225.

  .

  .

  .

- 32 channels result in 15,872 3rd-Order IM-products.

Adding more wireless links to the system will increase the number of possible combinations with interference potential logarithmically: *n* channels will result in two signal 3rd-order IM-products equal to:

$$(n^3 - n^2)/2 \qquad (20\text{-}33)$$

The situation gets even more complicated if intermodulation among three signals is considered, which would then include the following terms for 3rd order IM:

- F1+F2-F3.
- F1+F3-F2.
- F2+F3-F1.

Equal frequency spacing between RF carrier frequencies inevitably results in two- and three-signal intermodulation products and should be avoided unless using equipment with extreme linearity (discussed below). The RF level and the proximity define the level of the intermodulation product. If two transmitters are close, the possibility of intermodulation will increase significantly. As soon as the distance between two transmitters is increased, the resulting intermodulation product decreases significantly. By taking this into consideration, the physical distance between two or more transmitters is important. If a performer needs to wear two bodypack transmitters, it is recommended to use two different frequency ranges and to wear one so that the antenna is pointing up and the other is pointing down.

With intermodulation factored in, when the number of wireless channels increases, the required RF bandwidth increases significantly, Fig. 20-171.

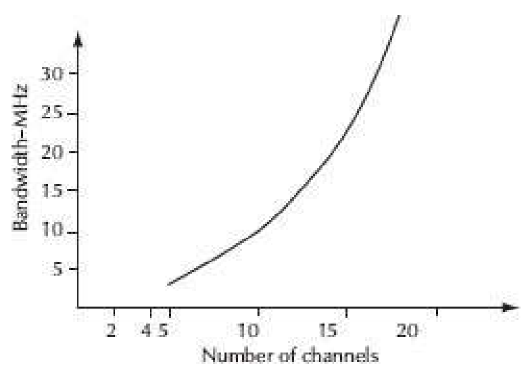


Figure 20-171. Bandwidth required for multi-channel systems. Courtesy Sennheiser Electronic Corporation.

External disturbing sources such as TV transmitters, public safety communications, noise from digital equipment, etc., also have to be taken into consideration. Fortunately, the screening effect of buildings is rather high (30–40dB). For indoor applications, this effect keeps strong outside signals at low levels. A significant problem can occur when poorly screened digital equipment is working in the same room. These wideband disturbing sources are able to interfere with wireless audio equipment. The only solution to this problem is to replace the poorly screened piece of equipment with a better one.

Other RF-systems which should be considered for compatibility are:

1. TV-stations “On-Air.”
2. Wireless intercoms.
3. IFBs.
4. Wireless Monitor Systems.
5. Other wireless systems.

Compatibility between components of a system is achieved if the following conditions are met: each link in a multichannel wireless system functions equally well with all other links active, and no one single link—or any combination of multiple links—cause interference.

If the transmitter of a wireless mic channel is switched off, its complementary receiver should also be switched off or muted at the mixing console. A receiver that does not “see” its transmitter will try to latch onto a nearby signal. That signal may be an intermodulation product. The receiver will then try to demodulate this signal and apply it to the speaker system.

Equipment can be designed to minimize intermodulation. A specification known as intermodulation rejection or suppression is a measure of the RF input threshold before intermodulation occurs.

For a well designed receiver, this specification will be 60dB or greater. An intermodulation rejection of 60dB means that intermodulation products are generated at input levels of approximately 1mV. If high quality components are used, having an intermodulation suppression of 60dB or greater, only the 3rd order products need to be considered.

Avoiding the Generation of IM Products. Intermodulation occurs when active electronic components are driven past their linear operating range. Sophisticated high quality designs can increase the range of linearity. Highly linear designs use top grade electronic components and draw extra current providing “headroom” that prevents harmonics from being created.

Intermodulation between transmitters can be avoided through the use of circulators. An antenna is agnostic in that it will transmit as well as receive signals. If a signal is picked up by a transmit antenna it could be fed into the amplifier of transmitter’s output stage, causing inter-modulation. Circulators allow a signal to be sent out of the transmitter but block unwanted signals from entering in reverse fashion.

20.31.3 Transmitter Considerations

Transmitters are available as portable devices in the form of handheld microphones, bodypacks, and plug-on transmitters, and are produced as stationary units for monitor systems.

20.31.3.1 Range and RF Power

Transmitter power is a rating of its potential RF signal strength. This specification is generally measured at the antenna output. The range of a wireless transmission depends on several factors. RF power, the operating frequency, the set-up of the transmitter and receiver antennas, environmental conditions, and how the transmitter is held or worn, are all aspects that determine the overall coverage of the system. Therefore, power specifications are of only limited use in assessing a transmitter's range, considering these variable conditions. Also, battery life is associated with RF output power. Increased power will reduce battery life with only a moderate increase in range.

Using transmitters with the right amount of RF output power is important to ensure total system reliability. There is a common misconception that higher power is better. However, in many applications high power can aggravate intermodulation (IM) distortion, resulting in audible noises.

First of all, the applied RF output power must fall within the limit allowed by each country's regulations. In the USA, the maximum RF output power for wireless microphones is limited to 250mW for licensed professional broadcast applications and 50mW for unlicensed use. In most of the countries in Europe this figure is 50mW, while in Japan it is only 10mW. Despite the 10mW limitation, many multichannel wireless microphones are operating in Japan. This is achieved by careful attention to factors like antenna position, use of low loss RF cables and RF gain structure of the antenna distribution system.

There are indeed some applications in which more RF output power is an appropriate measure; a perfect example would be a golf tournament, as the wireless system needs to cover a wide area. The risk of intermodulation is low at this type of function since the microphones are generally not in close proximity to each other.

If transmitters with high RF power are close together, intermodulation usually occurs. At the same time, the RF noise floor in the performance area is increased. For these reasons low power transmitters (≤50mW) are recommended for multichannel stage applications.

20.31.3.2 Battery Regulation

Transmitters should be designed to provide constant RF output power and signal processing throughout the event being staged. This can be achieved through the use of a DC-to-DC converter circuit. Such a circuit takes the decaying battery voltage as its input and regulates it to have a constant voltage output. Once the battery voltage drops below a minimum threshold, the DC-to-DC converter shuts off, almost instantaneously. The result is a transmitter that is essentially either off or on. While it is on, the RF output power, signal processing, and other relevant specifications remain the same. Transmitters without regulation circuits, once the battery voltage begins to drop, will experience reduced range and audio quality.

20.31.3.3 Spurious Emissions

Apart from the wanted carrier frequency, transmitters can also radiate some unwanted frequencies known as spurious emissions. For large multichannel systems potential spurious frequencies cannot be ignored. They can be significantly reduced through elaborate filtering and contained by hermetically sealed metal compartments that shield the RF components. A completely metal RF "tight" housing for the transmitter can provide even better protection. Also, an RF tight transmitter is less susceptible to outside interference.

A metal housing not only has advantages for its shielding properties, but also its durability. These devices usually experience much more abuse by actors and other talent than anyone ever predicts, Fig. 20-172.

20.31.3.4 Transmitter Antenna

Every wireless transmitter is equipped with an antenna, which is critically important to the performance of the wireless system. If this transmitter antenna comes in contact with the human body, the transmitted RF energy is reduced and may cause audible noises known as “drop-outs.” This effect of detuning the antenna on contact is called body absorption.

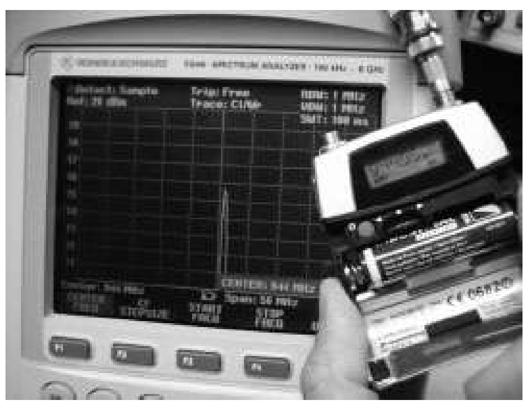


Figure 20-172. Sennheiser SK 5212 bodypack transmitter and a spectrum analyzer. Courtesy Sennheiser Electronic Corporation.

For this reason, talent should not touch the antenna while using handheld microphones. Unfortunately, there is no guarantee that they will follow this recommendation. Taking this into account, optimized antenna set-up at the receiver side and the overall RF gain structure of the system becomes critical.

This same effect can occur when using bodypack transmitters, especially if the talent is sweating. A sweaty shirt can act as a good conductive material to the skin. If the transmitter antenna touches it, reduced power and thus poor signal quality may result. In this case, a possible approach is to wear the bodypack upside-down near or attached to the belt, with the antenna pointing down. Sometimes this measure does not work because the talent will sit on the antenna. In this case, a possible solution is keeping the transmitter in the normal position and fitting a thick-walled plastic tube over the antenna, such as the type that are used for aquarium filters.

20.31.4 Receiver Considerations

The receiver is a crucial component of wireless audio systems, as it is used to capture the desired signal and transfer its electrical information into an audio signal. Understanding basic receiver design, audio processing, squelch, and diversity operation can help ensure optimum performance of the system.

Virtually all modern receivers feature superheterodyne architecture, in which the desired carrier is filtered out from the multitude of signals picked up by the antenna, then amplified and mixed with a local oscillator frequency to generate the difference "intermediate frequency." This "IF" undergoes more controlled discrimination and amplification, before the signal is demodulated and processed to restore the output with all the characteristics and qualities of the original.

Audio signal processing of an FM receiver is the mirror opposite of its transmitter. Processing done in the transmitters often include pre-emphasis (boosting high audio frequencies) as well as compression. These are reversed in the receiver by the de-emphasis and the expander circuit.

An inherent RF noise floor exists in the environment. The squelch setting should be set above this noise level. This acts as a noise gate that mutes the audio output if the wanted RF signal falls below a threshold level. This prevents a blast of white noise through the PA if the RF signal is completely lost. If the squelch setting is too low, the receiver might pick the noise floor and this noise can be heard. If the squelch setting is too high the range of the wireless microphone is reduced.

RF Signal Level

Varying RF signal strength is mainly due to multi-path propagation, absorption and shadowing. These are familiar difficulties also experienced with car radios traveling within cities.

Audible effects due to low RF signals, known as drop-outs, can occur even at close range to the receiver, due to multi-path propagation. Some of the transmitted waves find a direct path to the receiver antenna and others are deflected off a wall or other object. The antenna detects the vector sum, magnitude and phase, of direct and deflected waves it receives at any particular instant. A deflected wave can diminish a direct wave if it has different phase, resulting in an overall low signal. This difference in phase is due to the longer path a deflected wave travels between the transmitter and receiver antennae and any phase reversal occurring when it hits an object. This phenomenon needs to be addressed in an indoor application since the field strength variation inside a building with reflecting walls is 40dB or more. It is less critical outside.

RF energy can be absorbed by non-metallic objects resulting in low signal strength. As stated previously, the human body absorbs

RF energy quite well. It is important to place antennas correctly to minimize this effect.

Shadowing occurs when a wave is blocked by a large obstacle between the transmitter and receiver antennas. This effect can be minimized by keeping the antennas high and distance of ½ wavelength away from any large or metallic objects.

These problems are addressed by a diversity receiver. A diversity system is recommended even if only one channel is in operation. Large multichannel systems are only possible with diversity operation.

There are a variety of diversity concepts available. Antenna Switching Diversity uses two antennas and a single receiving circuit. If the level at one antenna falls below a certain threshold it switches to the other antenna. This is an economical architecture but it leaves the chance that the second antenna could be experiencing an even lower signal than the one that falls below the threshold level. Another approach is switching the audio signal of two independent receiver units where each receiver unit is connected to its own antenna. This is known as TRUE diversity. This technique improves the effective RF receiving level by greater than 20dB. With a true diversity system, frequent switching between the two antennas is a desired result.

The minimum distance between the two diversity antennas is very often an issue of debate. A minimum of 1/4 of a wavelength of the frequency wave seems to be a good approach. Depending on the frequency, 5–6 inches is the minimum distance. In general, a greater distance is preferred.

20.31.5 Antennas

The position of the antenna and the correct use of its related components – such as the radio frequency (RF) cable, antenna boosters, antenna attenuators, and antenna distribution systems—are the key to trouble-free wireless transmission. The antennas act as the “eyes” of the receiver, so the best results can be achieved by forming a direct line of sight between the transmitter antenna and receiver antenna of the system.

Receiving and transmitting antennas are available as omnidirectional and directional variants. For receiving, omnidirectional antennas are often recommended for indoor use because the RF signal is reflected off of the walls and ceiling. When working outside, a directional antenna is most often a good choice since there are usually little to no reflections outdoors, and this directivity will help to stabilize the signal. In general, it is wise to keep an “antenna tool box” that contains both omnidirectional and directional antennas for use in critical RF situations, since they transmit and receive signals differently.

Omnidirectional antennas transmit or receive the signal by providing uniform radiation or response only in one reference plane, which is usually the horizontal one parallel to the earth’s surface. Within that plane, the omnidirectional antenna has no preferred direction and cannot differentiate between a wanted and an unwanted signal.

If a directional antenna is used, it will transmit or receive the signal in the path it is pointing towards. The most common types are the yagi antenna and the log-periodic antenna, which are often wide range frequency antennas covering the whole UHF range. In an outdoor venue, the desired signal can be received and an unwanted signal, from a TV station for example, can be rejected to a certain degree by choosing the correct antenna position. A directional antenna also transmits or receives only in one plane, like an omnidirectional antenna.

Several types of omnidirectional and directional antennas also exist for specific conditions. The telescopic antenna is an omnidirectional antenna and often achieves a wide range (450–960MHz). If telescopic antennas are in use they should be placed within the line of sight of the counterpart antenna. They should not, for example, be mounted inside a metal flight case with closed doors as this will reduce the RF field strength from the transmitter and compromise the audio quality.

System performance will be raised considerably when remote antennas are used. A remote antenna is one which is separated from the receiver or transmitter unit. These antennas can be mounted on a mic stand or similar support. This will improve the RF performance significantly. However, when using remote antennas, some basic rules need to be considered. Placing antennas above the talent increases the possibility the transmitter and receiver remain within line of sight, ensuring reliable transmission. If a directional antenna is used, the position of the antenna and the distance to the stage is important. One common set-up is pointing both receiving antennas toward the center of the stage. Once again, a line of sight between the receiver and transmitter antennas is best for optimum transmission quality.

Directional and omnidirectional antennas do have a preferred plane, which is either the horizontal or vertical plane. If the polarization between the transmitter and receiver antenna is different, this will cause some significant loss of the RF level. Unfortunately, it is not possible to have the same polarization of the antennas all of the time. In a theatrical application, the antenna is in a vertical position when the actress or actor walks on the stage. The polarization of the transmitter may change to the horizontal position if a scene requires the talent to lie down or crawl across the stage. In this case, circular polarized antennas can help. These kinds of antennas can receive the RF signal in all planes with the same efficiency.

Because the polarization of the antenna is critical and telescopic antennas are often used, it is not recommended to use the receiver antennas strictly in a horizontal or vertical plane. Rather, angle the antennas slightly as this will minimize the possibility that polarization would be completely opposite between transmitter and receiver.

One last note: The plural form for the type of antenna discussed in this article is "antennas." Antennae are found on insects and aliens.

20.31.5.1 Antenna Cables and Related Systems

Antenna cables are often an underestimated factor in the design of a wireless system. The designer must choose the best cable for practical application, depending on the cable run and the installation. As the radio frequency travels down the cable its amplitude is attenuated. The amount of this loss is dependent on the quality of the cable, its length and the RF frequency. The loss increases with longer cable and higher frequencies. Both of these effects must be considered for the design of a wireless microphone system, Table 20-4.

Table 20-4. Different types of RF cables with various diameters and the related attenuation for different frequencies. (Source: Belden Master Catalogue)

Cable Type	Frequency MHz	Attenuation db/100ft	Attenuation dB/100m	Cable diameter in/mm]
RG-174/U	400 700	19.0 27.0	62.3 88.6	0.110/2.8
RG-58/U	400 700	9.1 12.8	29.9 42.0	0.195/4.95
RG-8X	400 700	6.6 9.1	21.7 29.9	0.242/6.15
RG-8/U	400 700	4.2 5.9	13.2 19.4	0.405/10.3
RG-213	400 700	4.5 6.5	14.8 21.8	0.405/10.3
Belden 9913	400 700	2.7 3.6	8.9 11.8	0.405/10.3
Belden 9913F 9914	400 700	2.9 3.9	9.5 12.8	0.405/10.3

RF cables with a better specification regarding RF loss are often thicker. These are highly recommended for fixed installations. In a touring application, in which the cable must be stored away each day, these heavier cables can be very cumbersome.

As any RF cable has some RF attenuation, cable length should be as short as possible without significantly increasing the distance between the transmitter and receiver antennas. This aspect is important for receiving applications but is even more critical for the transmission of a wireless monitor signal.

In a receiving application, it is important to consider losses from the cable as well as from any splitter in the antenna system during the design and concept stage of a wireless microphone system. If the losses in the system are small, an antenna booster should not be used. In this case, any drop-out is not related to the RF loss in the antenna system; instead, it is more often related to the antenna position and how the transmitter is used and worn during the performance. An antenna booster is recommended if the loss in the antenna system is greater than 6dB.

If an antenna booster is necessary due to long antenna cable runs, it should be placed as close as possible to the receiving antenna. Antennas with a built-in booster are known as active antennas. Some of these have a built-in filter, only allowing the wanted frequency range to be amplified. This is another measure to reduce the possibility of intermodulation of this amplifier.

Two antenna boosters should not be used back-to-back when the RF cable run is very long. The second antenna booster would be overloaded by the output of the first amplifier and would produce intermodulation.

Special care must be taken when using an antenna booster if the transmitter comes close to the receiver antenna. The resulting strong signal could drive the antenna booster past its linear operation range, thus producing intermodulation products. It is recommended to design and install a system such that the transmitter remains at least 10 feet from the receiver antenna at all times.

Also, signals that come from over-the-air (OTA) TV broadcast—such as a Digital Television (DTV) signal—are unwanted signals that may contribute to intermodulation products in any amplifier stage of your system. Outboard narrowband filters are available that are tunable to a 6MHz (one TV channel) bandpass. This provides added safety when operating on a vacant TV channel in between active TV signals.

This will often work for fixed installations because it is less likely that the RF environment will change. This is especially the case where the RF environment is congested with many TV stations or other wireless systems are operating in the vicinity, such as a broadcast production studio in a major city.

20.31.5.2 Splitter Systems

Antenna splitters allow multiple receivers to operate from a single pair of antennas. Active splitters should be used for systems greater than four channels so that the amplifiers can compensate for the splitter loss. Security from interference and intermodulation can be enhanced by filtering before any amplifier stage. As an example, a thirty-two channel system could be divided into four subgroups of eight channels. The subgroups can be separated from each other by highly selective filters. The subgroups can then be considered independent of each other. In this way, frequency coordination only needs to be performed within each group. It is much easier to coordinate eight frequencies four times than to attempt to coordinate a single set of 32 frequencies, Fig. 20-173.

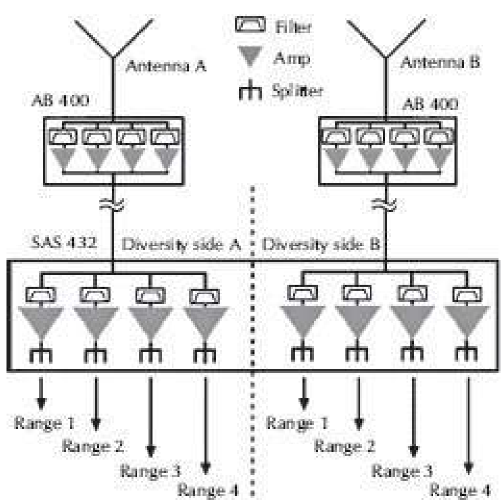


Figure 20-173. Diversity antenna set-up with filtered boosters, long antenna cables, and active splitter with selective filtering. Courtesy Sennheiser Electronic Corporation.

20.31.6 Wireless Monitor Systems

Wireless monitor systems are essential for stage-bound musical productions. Perhaps the biggest advantage of a wireless monitor system is the ability to use an individual monitor mix for each musician on stage. Furthermore, a wireless monitor system significantly reduces the amount of, or even eliminates, monitor speakers in the performance area. This results in lower risk of feedback, and a more lightweight, compact monitor system.

Some special precautions must be taken before using wireless monitor systems. In most cases, this signal is a stereo signal. This multiplexed signal is more sensitive to drop-outs, static and multipath situations. For long range applications, mono operation can improve system performance.

If wireless microphones and wireless monitor systems are used in parallel, those systems should be tuned to separate frequency bands at least 4MHz apart, more is even better. Also, the physical distance between any transmitters and the in-ear receiver on a performer should be maximized. This will reduce the risk of blocking—an effect that desensitizes a receiver and prevents the reception of the desired signal. Therefore, if a bodypack wireless mic transmitter and a wireless monitor receiver are both attached to the same talent, those devices should not be mounted directly beside each other.

When musicians use the same monitor mix, one transmitter can be used to provide the radio frequency (RF) signal to more than one wireless monitor receiver. If individual mixes are desired, each mix requires its own transmitter operating on a unique frequency. To avoid intermodulation disturbances, the wireless monitor transmitters should be combined, and the combined signal should then be transmitted via one antenna. Active combiners are highly recommended. Passive combiners suffer from signal loss and high crosstalk. An active combiner isolates each transmitter by around 40dB from the others and keeps the RF level the same (0dB gain), thus minimizes intermodulation. Again, intermodulation is a major issue within the entire wireless concept. When using stereo transmission, it is even more critical.

When considering an external antenna, one important factor must be taken into consideration: the antenna cable should be as short as possible to avoid losses via the RF cable. A directional external antenna is recommended to reduce multipath situations from reflections, and it will have some additional passive gain that will increase the range of the system.

If remote antennas are used for the wireless monitor transmitters as well as wireless mic receivers, those antennas should be separated by at least 10 feet. "Blocking" of the receivers, as discussed above, is thus avoided. Furthermore, the antennas should not come in direct contact with the metal of the lighting rig. This will detune the antenna and reduce the effective radiated wireless signal.

20.31.7 System Planning For Multichannel Wireless Systems

When configuring a multichannel wireless microphone system, several factors are essential for reliable operation. First, you must understand the environment in which the system will be used.

1. Location. The coordinates of a venue can be determined by using mapping tools on the internet, such as Google Earth. If you figure out the coordinates of the venue, you can simply plug this information into the Federal Communications Commission (FCC) homepage at http://www.fcc.gov/fcc-bin/audio/tvq.html. The result shows all transmitters licensed by the FCC in this area. Valuable information can also be found within an FCC approved TV White Space database, including Spectrum Bridge: http://whitespaces.spectrum-bridge.com/whitespaces/home.aspx; or Key Bridge Global: https://keybridgeglobal.com/whitespace/. Most importantly, these databases will indicate any TV channels that are reserved for wireless microphone use at a specific location. All approved databases share the same information. Note, some channel reservations may be limited to certain dates and times. This information will allow the designer of the wireless system to plan which vacant TV channels can be used for wireless audio devices. If there is a TV transmitter close to the location of the wireless microphone system (<70 miles), this TV channel should generally be avoided. Operators running a large number of mics can apply to reserve locally vacant TV channels for their events through the database system, making those channels unavailable to other white space devices. Licensed operators, such as broadcasters can make reservations directly into the database. Unlicensed operators first need to apply for approval with the FCC. Once one knows which TV channels may be used in the area, the designer can use another software tool that calculates the IM-free frequencies and displays possible set-ups.

2. Quantity and Frequency Coordination. Determine how many wireless microphones, wireless monitor systems, intercoms, etc. are required or in use for your job. With the information you gathered from step one, you can begin the system design. You now have the available TV channels and the number of wireless systems you want to use.

With this know-how you can start the frequency coordination of your system inside the vacant TV channels. This is supported by software that is available from various companies. The key here is to prevent intermodulation products (unwanted frequencies generated by harmonic distortion) from interfering with the wanted frequencies of your wireless systems.

A test at the venue is also necessary. If you have the chance, scan the location with a spectrum analyzer, Fig. 20-174. With this tool, you can verify that the information from the internet is correct. Alternately, you can scroll through the tunable frequencies of your wireless receivers to scan the RF activity in the venue. Many receivers also have an auto scan function to find open frequencies. This cross-check is necessary to find out whether other wireless devices are in use that you do not have on your list, which could interfere with your signals during operation.3. Tune Your Components

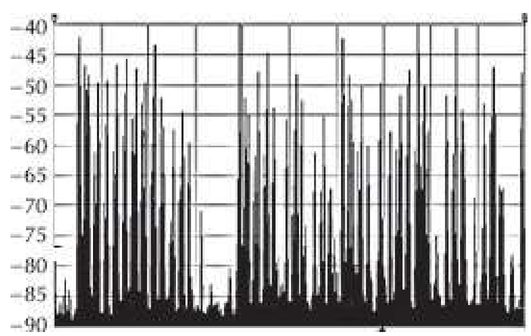


Figure 20-174. Plot of the RF spectrum in Athens outside the Olympic Stadium (450–960MHz). Courtesy Sennheiser Electronic Corporation.

Set your individual transmitters and corresponding receivers to their coordinated frequencies. Switch on all components and perform a final test of compatibility. Physically space the transmitters a couple feet apart and at least ten feet from the receiving antenna. Listen for any interference. Once again, compatibility is confirmed when each link in a multichannel wireless system functions equally well with all other links active and, no single link—or any combination of multiple links—cause interference.

20.31.8 Conclusion

Large multichannel wireless systems demand excellent planning, especially in the initial phase, and good technical support. Observing all the above mentioned items, reliable operation of a system can be achieved, even under difficult conditions.

With gratitude for contributions by Volker Schmitt, Gerrit Buhe, and Peter Arasin.

20.32 Microphone Accessories

By Glen Ballou

20.32.1 Inline Microphone Processors

The overall sound of a microphone can often benefit from signal processing, and most mixers provide some basic equalization as a tool for customizing the sound of the microphone. Digital mixers provide an even greater set of tools, including parametric EQ, compression, gain management, and other automated functions. Dedicated signal processing for each microphone in a system provides a real advantage for the user, and some manufacturers are offering this sort of custom processing on a per-microphone basis via processors that plug inline with the microphone. These include automatic gain control, automatic feedback control, control for plosives and the proximity effect, and integrated infrared gates that turn the microphone on and off based on the presence of a person near the microphone. These phantom-powered processors allow for targeted solutions to many problems caused by poor microphone technique.

20.32.1.1 Lectrosonics HM Digital Hybrid® UHF Plug-On Transmitter

The most common application of this transmitter, Fig. 20-175, is eliminating the cable between a microphone and a sound or recording system. A prime example is acoustic analysis in a large auditorium or stadium where measurements must be made at multiple locations around the sound system coverage area and extremely long cable runs are not practical. In this case, the wireless not only speeds up the process of making measurements, but it also allows more measurements to be taken, which can improve the final sound system performance. Fig. 20-176 is the block diagram of the transmitter.

Figure 20-175. Lectrosonics HM Digital Hybrid Wireless® plug-on transmitter. Courtesy Lectrosonics, Inc.

Digital Hybrid Wireless® is a patented process that combines a 24 bit digital audio stream with wide deviation FM (US Patent 7,225,135). The process eliminates a compandor to increase audio quality and expand the applications to test and measurement and musical instrument applications.

Digital Hybrid Wireless® combines digital audio with an analog FM radio link to provide high audio quality and RF performance.

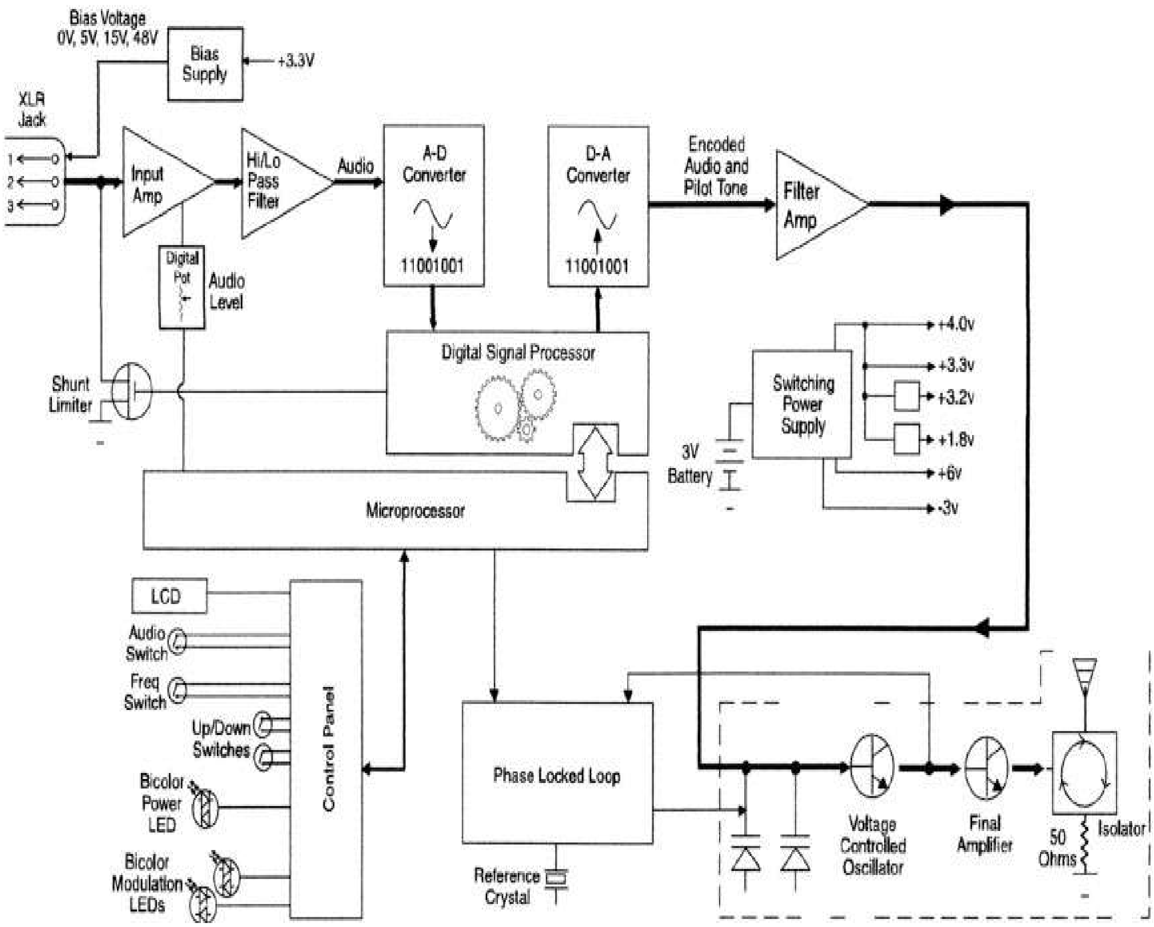


Figure 20-176. Block diagram of the Lectrosonics HM transmitter. Courtesy Lectrosonics, Inc.

Channel noise is reduced by digitally encoding the audio in the transmitter, sending the encoded information via an analog FM wireless link, then decoding it in the receiver. This algorithm is not a digital implementation of an analog compandor, it is an technique which can be accomplished only in the digital domain. The process eliminates companded artifacts, allowing it to be used in the test and measurement of acoustic spaces.

DSP has a frequency stability of ±0.002% eliminating the need for crystals, plus it allows a different pilot tone for each of the 256 frequencies in the tuning range of the systems frequency block. Individual pilot tones virtually eliminate squelch problems and multichannel systems where a pilot tone signal can appear in the wrong receiver via intermodulation products. A circular/isolator in the output stage further insures against intermodulation interference.

An additional benefit of the FM radio link is the ability of the DSP to emulate a compandor for compatibility with analog receivers from Lectrosonics and other manufacturers.

In the native hybrid mode, the FM deviation is ±75kHz to provide a wide dynamic range. This wide deviation combined with 100mW of output power provides a significant improvement in the audio *SNR* and the suppression of RF noise and interference.

Used with a microphone, the antenna is a dipole formed between the transmitter housing and the microphone body. When plugged into a console or mixer output, the housing of the transmitter is similar to the radiator of a ground plane antenna, with the console or mixer chassis functioning as the ground.

The transmitter operates on two AA batteries and can provide 5, 15, or 48V phantom power or it can be turned off for dynamic microphones.

The transmitter is available on 9 different frequency blocks in the UHF band between 470 and 691.1MHz. Each block provides 256 frequencies in 100kHz steps.

20.32.1.2 MXL Mic Mate™ USB Adapter

The Mic Mate™ and Mic Mate Pro, Fig. 20-177, are a USB adapter used to connect a microphone to a Macintosh or PC computer. It uses a 16 bit Delta Sigma A/D converter with $THD + N$ = 0.01% at sampling rates of 44.1 and 48.0kHz and includes a three-position analog gain control. The USB microphone preamp has a balanced low noise analog input, and supplies 48Vdc phantom power to the microphone It includes MXL USB Recorder Software for two track recording. There are three different Mic Mates, one for condenser microphones, one for dynamic microphones, and one for news line feeds, video cameras, etc.

MXL® Mic Mate™ Pro handles microphone gain and headphone volume adjustments with low profile rotary knobs, while the built-in headphone jack allows for zero-latency direct monitoring.

Figure 20-177. MXL® Mic Mate™ and Mic Mate™ Pro USB adapter. Courtesy Marshall Electronics.

20.32.2 Windscreens and Pop Filters

A *windscreen* is a device placed over the exterior of a microphone for the purpose of reducing the effects of breath noise and wind noise when recording out doors or when panning or gunning a microphone. A windscreen's effectiveness increases with its surface area and the surface characteristics. By creating innumerable miniature turbulences and averaging them over a large area, the sum approaches zero disturbance. It follows that no gain is derived from placing a small foam screen inside a larger blimp type, Fig. 20-178A, whereas a furry cover can bring 20dB improvement, Fig. 20-178B. Most microphones made today have an integral windscreen/pop filter built in. In very windy conditions, these may not be enough; therefore, an external windscreen must be used.

With a properly designed windscreen, a reduction of 20–30dB in wind noise can be expected, depending on the *SPL* at the time, wind velocity, and the frequency of the sound pickup. Windscreens may be used with any type microphone because they vary in their size and shape. A cross-sectional view of a windscreen employing a wire frame covered with nylon crepe for mounting on a 1inch diameter microphone is shown in Fig. 20-179. The effectiveness of this screen as measured by Dr. V. Brüel of Brüel and Kjaer is given in Fig. 20-180.

Figure 20-178. Blimp-type windscreen for an interference tube microphone. Courtesy Sennheiser Electronic Corporation.

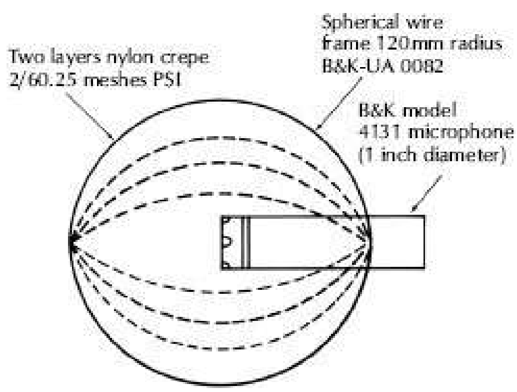


Figure 20-179. Typical silk-covered windscreen and microphone. Courtesy B and K Technical Review.

20.32.2.1 Wind Noise Reduction Figures for Rycote Windshielding Devices

Rycote has developed its own technique for measuring wind noise that uses real wind and a real time differential comparison. The technique compares the behavior of two microphones under identical conditions, one with a particular wind noise reduction device fitted and the other without, and produces a statistical curve of the result corrected for response and gain variations.

A Sennheiser MKH60 microphone, a representative short rifle microphone, is shown without any low-frequency attenuation in a wideband (20Hz–20kHz) test rig, Fig. 20-181.

When a wind noise reduction device is fitted, its effect on the audio response is a constant factor—if it causes some loss of high frequency, it will do it at all times. However, the amount it reduces wind noise depends on how hard the wind is blowing. If there is a flat calm it will have no beneficial effect and the result will be a degradation of the audio performance of the microphone. However, in a strong gale a small deviation from a perfect flat response may be insignificant for a >30dB reduction in wind noise. Wind noise is in the low-frequency spectrum. For a naked Sennheiser MKH60, the wind-produced energy is almost entirely below 800Hz, rising to a peak of 40dB at about 45Hz. It is the effect of a shield at these lower frequencies that is most important. Cavity windshields inevitably produce a slight decrease in low-frequency response in directional microphones but this is not usually noticeable. Basket types have very little effect on high frequency. Fur coverings, while having a major effect in reducing low-frequency noise, will also attenuate some high frequency.

Adding the low-frequency attenuation available on many microphones or mixers (which is usually necessary to prevent infrasonic overload and handling noise when handholding or booming a microphone) may give extra wind noise reduction improvements of >10dB at the cost of some low-frequency signal loss.

The standard (basket) windshield shows up to 25dB wind noise attenuation at 35Hz while giving almost no signal attenuation, Fig. 20-181.

The Softie Windshield is a slip-on open cell foam with an integral fitted fur cover. The Softie reduces wind noise and protects the microphone. It is the standard worldwide in TV. A Softy attenuates the wind noise about 24dB, Fig. 20-181, graph B.

Adding a Windjammer (furry cover) to the basket windshield will give an improvement of about 10dB at low frequency to –35dB, Fig. 20-181, graph C. The attenuation of the Windjammer is approximately 5dB at frequencies above 6kHz although this will increase if it is damp or the fur is allowed to get matted. Overall this combination gives the best performance of wideband wind noise reduction against signal attenuation. To determine the correct windscreen for microphones of various manufacturers, go to www.microphone-data.com.

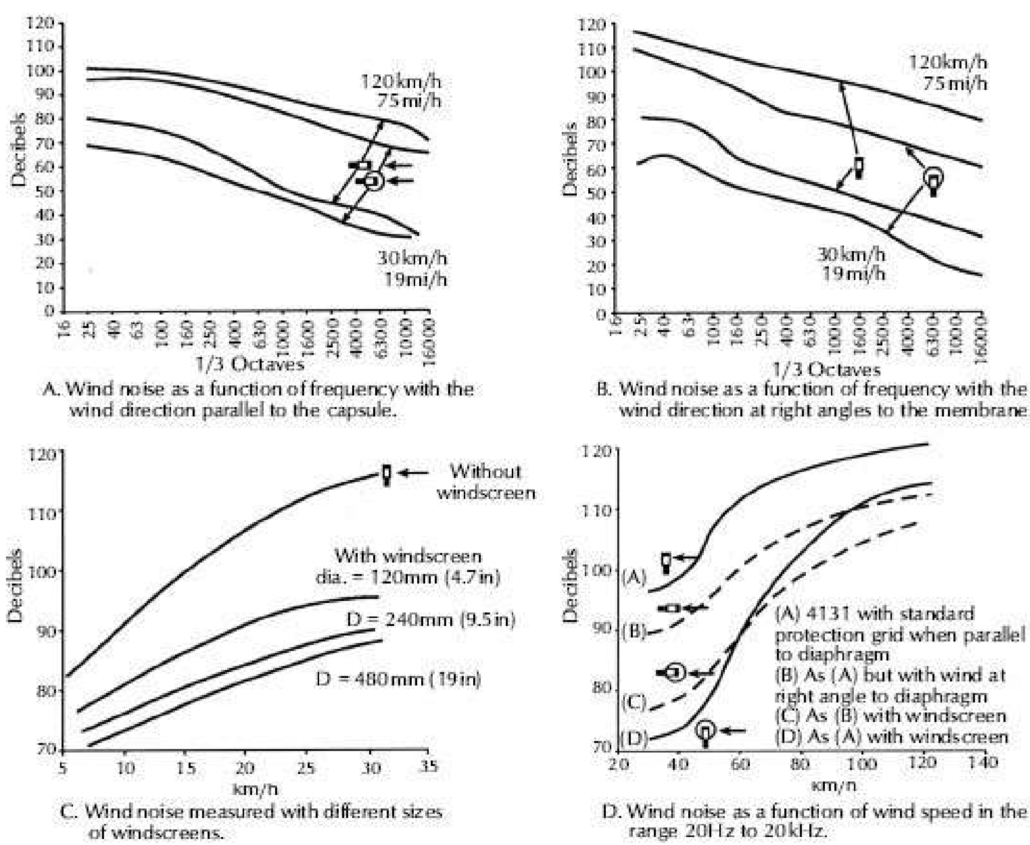


Figure 20-180. The effectiveness of the windscreen shown in Fig.

20-163. Courtesy B and K Technical Review.

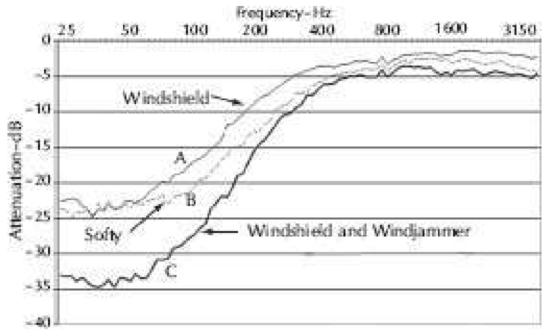


Figure 20-181. Wind noise reduction options for a Sennheiser MKH60 microphone under real wind conditions. Courtesy Rycote Microphone Windshields LTD

Pop protection is best appreciated when close-talking and explosive breath sounds are particularly bothersome. These explosive breath sounds are commonly produced when saying words involving P and T sounds. The phrase *explosive breath sound* is somewhat of a misnomer since these sounds, without amplification, are normally inaudible to a listener.[15]

The electrical output from the microphone is actually the transient microphone response to this low-velocity, high-pressure, pulse-type wavefront. The P and T sounds are projected in different directions and can be shown by saying the letters P and T while holding your hand about 3in (7.6cm) in front of your mouth. Note that the T sound is felt at a considerable distance below the P sound.

For most microphones, pop output varies with distance between the source and microphone, reaching a peak at about 3in (7.6cm). Also the worst angle of incidence for most microphones is about 45° to the microphone and for a glancing contact just at the edge of the microphone along a path parallel to the longitudinal axis.

sE Dual Pro Pop Filter. An interesting pop filter is shown in Fig. 20-182. The sE Dual Pro Pop pop screen is a two-filter device to suit vocal performances. The device has a strong gooseneck with both a standard fabric membrane and a pro metal pop shield on a hinge mechanism. They can be used separately or both simultaneously depending on the application. The metal screen is not simply perforated but louvered at a slight angle to redirect ultra-low frequency breath blasts which are pushed past the side of the screen. This does not attenuate high frequencies as fabric screens do.

Figure 20-182. sE Dual Pro Pop pop screen. Courtesy sE Electronics.

In an emergency pop filters can be as simple as two wire-mesh screens treated with flocking material to create an acoustic resistance.

20.32.2.2 Reflection Filter

The Reflexion Filter by sE Electonics is used to isolate a microphone from room noises hitting it from unwanted directions, Figs. 20-183 and 20-184.

Figure 20-183. The Reflection Filter. Courtesy sE Electronics.

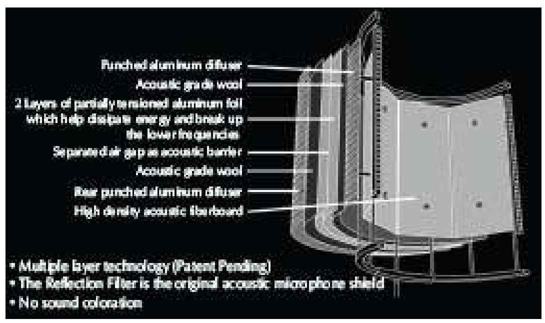


Figure 20-184. Reflexion Filter. Courtesy sE Electronics.

The reflexion filter has seven layers. The first layer is punched aluminum, which diffuses the sound waves as they pass through it to a layer of absorptive wool. The sound waves next hit a layer of aluminum foil, which helps dissipate energy and break up the lower frequency waveforms. From there they hit an air space kept open by rods passing through the various layers.

Next the waves hit an air space that acts as an acoustic barrier. The sound waves pass to another layer of wool and then through an outer, punched, aluminum wall that further serves to absorb and then diffuse the remaining acoustic energy.

The various layers both absorb and diffuse the sound waves hitting them, so progressively less of the original source acoustic energy passes through each layer, reducing the amount of energy hitting surfaces so less of the original source is reflected back as unwanted room ambience to the microphone. The Reflexion Filter also reduces reflected sound from reaching the back and sides of the microphone. The system only changes the microphone output by a maximum of 1dB, mostly below 500Hz.

The stand assembly comprises a mic stand clamp fitting, which attaches to both the Reflexion Filter and any standard fitting shock mount.

20.32.3 Microphone Holders and Shock Mounts

Microphone holders or adapters are used to hold the microphone to a stand or any other surface. They do not include a shock mount. The Shure A27M stereo microphone adapter, Fig. 20-185, is used to hold two microphones in selectable positions for stereo reproduction. It consists of two vertically stacked sections that rotate on their centers. Each section contains a threaded ⅝ in-27 threaded stud and locking ring to accept various microphones. The bottom section includes a ⅝ in-27 adapter for mounting on a microphone stand.

Shock mounts are used to eliminate noise from being transmitted to the microphone, usually from the floor or table.

Figure 20-185. Stereo microphone adapter. Courtesy Shure Incorporated.

Microphones are very much like an accelerometer in detecting vibrations hitting the microphone case. Shock mount suspensions allow a microphone to stay still while the support moves.

Suspensions all use a springy arrangement that allows the microphone to be displaced and then exerts a restoring force to return it to the rest point. It will inevitably overshoot and bounce around, but the system should be damped to minimize this.

As frequency lowers, the displacement wavelength increases so the suspension has to move farther to do the job. For any particular mass of microphone and compliance (wobbliness) of suspension, there is a frequency at which resonance occurs. At this point the suspension amplifies movement rather than suppresses it. The system starts to isolate properly at about three times the resonant frequency.

The microphone diaphragm is the most sensitive along the Z-axis to disturbances. The ideal suspensions are most compliant along the Z-axis, but should give firmer control on the horizontal (*X*) and vertical (*Y*) axes to stop the mic slopping around, Fig. 20-186.

Suspension Compliance. Diaphragm and so-called donut suspensions can work well, but tend to have acoustically solid structures that affect the microphone's polar response. Silicone rubber bands, shock-cord cat's cradles, and metal springs are thinner and more acoustically transparent but struggle to maintain a low tension, which creates a low resonant frequency, while at the same time providing good X–Y control and reliable damping. The restraining force also rises very steeply with displacement, which limits low-frequency performance.

Figure 20-186. A Rycote lyre-type microphone suspension (shock mount) system. Courtesy Rycote Microphone Windshields LTD.

Figure 20-187. Shure A53M shock mount. Courtesy Shure Incorporated.

Shock mounts may be the type shown in Fig. 20-187. This microphone shock mount, a Shure A53M, mounts on a standard ⅝ in – 27 thread and reduces mechanical and vibration noises by more than 20dB. Because of its design, this shock mount can be used on a floor or table stand, hung from a boom, or used as a low-profile stand to place the microphone cartridge close to a surface such as a floor.

Shock mounts are designed to resonate at a frequency at least 2½ times lower than the lowest frequency of the microphone.[21] The goal is simple but there are practical limitations. The resonant frequency (f_n) of a mechanical system can be computed from

$$f_n = \frac{1}{2\pi}\sqrt{\frac{Kg}{w}} \quad (20\text{-}34)$$

where,
K is the spring rate of the isolator,
g is the acceleration due to gravity,
w is the load.

A microphone shock-mount load is almost completely determined by the weight of the microphone. To obtain a low-resonant frequency, the spring rate or stiffness must be as low as possible; however, it must be able to support the microphone without too much sag and be effective in any position the microphone may be used.

The Rycote lyre webs rely primarily on their shape to give different performance on each axis. Typically, a 100g force will barely move a microphone 1mm along the (up and down) Y-axis, whereas it will move about four times that on the (sideways) X-axis. In the critical Z-axis, it will move almost ten times as far, Fig. 20-186.

With a very low inherent tension the resonant frequency can be very low, and the Z displacement can be vast. Even with small-mass compact microphones, a resonance of <8Hz is possible, which means that microphones can be well isolated across almost their entire frequency range.

Damping has to be added to metal spring suspensions, and although integral to rubber band versions, is not very easy to control. With the lyre webs damping can be selected almost independently by choosing a suitable plastic. The Hytrel that Rycote uses not only damps smoothly but maintains its characteristics even down to arctic temperatures.

Most suspension systems are difficult to scale. Springs and elastic bands become thin and fragile, and the range of softness for rubber and foam is limited. However, this does not apply to lyre webs. The tiny InVision suspensions, which are visually unobtrusive, isolate compact and similar sized microphones down to <30Hz, yet are tough enough to be dropped on the floor without risk. Fig. 20-188 shows the actual measured performance of the transfer function for a Schoeps CCM4 microphone being shaken with pink noise in an InVision mount. Trace A shows the output from the microphone with the shaker operating but not touching the mic, revealing the inherent coupling through air and the building itself. Trace B is with the shaker directly coupled to the microphone body to reveal the actual level of vibration input. Finally, the trace C shows the microphone's output with the shaker knocking the bar of the mount, thus demonstrating the effectiveness of the suspension.

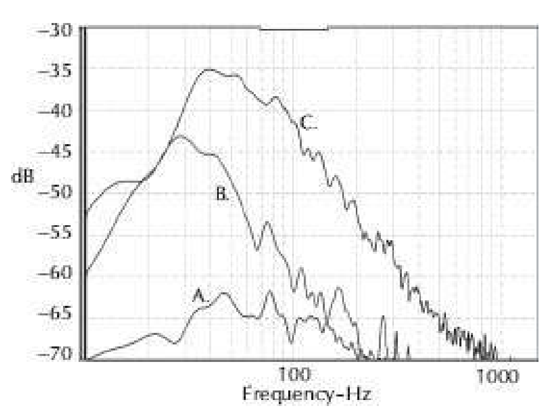


Figure 20-188. Effectiveness of an InVision mount. Courtesy Rycote Microphone Windshields LTD.

To determine the correct suspension systems for microphones of various manufacturers, go to www.microphone-data.com.

20.32.4 Stands and Booms

20.32.4.1 Microphone Stands

Microphones are mounted on microphone floor stands or table stands to place the microphone in front of the sound source. The floor stands are usually adjustable between 32 and 65in (0.8–1.6m) and incorporate a ⅝ in – 27 thread for mounting the microphone holder or shock mount. They may come with 10–12in (25.4–30.5cm) round bases and weigh 8–14lbs (3624–6342g), may have a 15in (380cm) triangular base and weigh 18lbs (8154g), or incorporate three legs with a 25in (380cm) spread for stability. If the three leg units are used with a boom, it is important that the boom extends out over a leg for stability.

The Shure Model 522, Fig. 20-189, is a dynamic base-station microphone for radio communications, paging, and dispatching systems. Its cardioid (unidirectional) pickup pattern suppresses unwanted background noise from nearby dispatchers, ventilating equipment, or office machines. It also can reduce feedback in public address paging applications.

It includes dual impedance; Low (150Ω) or High (30,000Ω) through an impedance change switch. A finger-tip control bar (locking or non-locking action) actuates the microphone circuit and can actuate an external relay or control circuit. Press and hold the bar switch to talk, and to lock the switch, press and pull it forward. To unlock the switch, move the bar back and release. The microphone height is adjustable from 9.75–12.25in (248–312mm).

Figure 20-189. Shure 522 table microphone stand with push-to-talk switch. Courtesy Shure Incorporated

20.32.4.2 Booms

Small booms, which are mounted on the standard microphone floor stand, are normally used to put the microphone in a place where it is difficult to reach with a straight floor stand, Fig. 20-190. They are also useful when micing from above the source. Combination booms and stands are often on wheels or flat tripod legs and adjustable from 60–90in (1.5–2.3m) vertically and 90–110in (2.3–2.8m) horizontally, Fig. 20-191.

Figure 20-190. Small adjustable microphone boom.

Figure 20-191. Adjustable microphone stand/boom. Courtesy Atlas Sound.

It is important that the boom and/or microphone stand be easily adjusted and that the clutch/brake system has a positive lock. Better microphone stands incorporate a piston-type air suspension system for effortless height adjustment and microphone protection.

The Audix MicroBoom™ system, Fig. 20-192, provides a solution for installations requiring continuous placement flexibility. They are compatible with all standard microphone stands and fixed mounts making set up both effortless and effective. The boom is a carbon fiber boom arm with a high-performance condenser microphone, the Audix MicroBoom™ offers the good solution for micing choirs, music assemblies, presentations, and plays. They come 24in, 50in, and 84in long to allow precise placement and positioning of the microphone. Every boom is wired internally with a shielded cable for maximum induced noise rejection. The MicroBoom™ system supports the entire range of Audix Micros™ Series condenser microphones. The clutch assembly holds the carbon fiber boom and includes a slot for the cable to exit, Fig. 20-193.

Figure 20-192. Typical MicroBoom™ setup. Courtesy Audex Corporation.

Figure 20-193. MicroBoom™ clutch. Courtesy Audex Corporation.

The Galaxy Audio MST-C Standformer tripod microphone stand, Fig. 20-194, is an interesting approach to a convertible boom/straight microphone stand. The center pivot clutch mechanism allows the boom arm section to slide down inside the fixed vertical section underneath. The five disc anti-slip cam allows the boom arm to hold up to 2 pounds fully extended.

Figure 20-194. Galaxy MST-C Standformer microphone stand. Courtesy Audio, Inc.

Figure 20-195. Røde PSA1 studio boom arm. Courtesy Røde Microphones.

Studio boom arms are used for radio, broadcast, studio, and home use. They allow the user to move the microphone toward him when in use and to push it away when required. They also free up the table area, Fig. 20-195.

Small booms, which are mounted on the standard microphone floor stand, are normally used to put the microphone in a place where it is difficult to reach with a floor stand, Fig. 20-196. They are also useful when micing from above the source. Combination booms and stands are often on wheels or flat tripod legs and adjustable from 60–90in (1.5–2.3m) vertically and 90–110in (2.3–2.8m) horizontally, Fig. 20-196.

Figure 20-196. Atlas BB-44 microphone boom. Courtesy Atlas Sound.

It is important that the boom and/or microphone stand be easily adjusted and that the clutch/brake system has a positive lock. Better microphone stands incorporate a piston-type air suspension system for effortless height adjustment and microphone protection.

Large booms, as used in television and motion-picture sound stages, are motorized and often include a stage for the microphone sound person.

Large booms, as used in television and motion-picture sound stages, are motorized and often include a stage for the microphone sound person.

20.32.4.3 Drum Microphone Mounts

The Kelly SHU Pro and Kelly SHU Composite microphone mounts are a horseshoe-shaped design to support any dynamic microphone that has standardized microphone threads; i.e., ⅝-27 thread count, Fig. 20-197. The Pro series mounting unit is made of aluminum for heavier and larger microphones such as the Shure BETA 52A, EV RE20 or AT 2500. The Composite series is made from 30% fiberglass-filled high-density nylon resins. Both the utilize the same isolating support system which is designed for internal mounting, however they will also work being mounted on the front of the drum. Recording studios is the best situation for external installation. Installation is adjustable on site to find the best position for the microphone inside the drum, and it will not move out of position during play or transport. No drilling is required to install the system. The product's support cord system is custom built by the user enabling exact microphone position. Self-crimping hooks are used to attach to the isolation cords. The product utilizes existing hardware screws as attachment points for the solid rubber support cord system thereby providing what is essentially a large shock-mount platform for the microphone. Stage vibration and crosstalk is virtually eliminated and the removal of the mini-boom stand for the kick drum mic is an added bonus. The microphone is always ready to simply plug and play at any given time with a consistent microphone output and signal.

Figure 20-197. Kelly SHU Pro microphone mount. Courtesy Kelly Concepts, LLC.

The Kelly SHU FLATZ, Fig. 20-198, is designed for use with boundary microphones. It is made from the same material as is the Kelly SHU Composite mounting unit. There are three different

FLATZ mounting plates to accommodate four of the most popular boundary mics used in the industry for kick drum amplification; The Shure BETA 91 and 91A, The Shure SM91 and the Sennheiser E901. Once installed, the boundary microphone can be safely left in the drum during transport.

20.32.5 Attenuators and Equalizers

Attenuators, equalizers, and special devices from Electro-Voice, Shure, and others are available to reduce the microphone output level or shape the response to roll off the low or high end, increase the 3–5kHz articulation region, or reverse polarity. These units normally have standard input and output male and female XLR or ¼inch phone plug connectors. Attenuators are also available to be installed between the capacitor capsule and the condenser microphone electronics to eliminate overload from high-level sources.

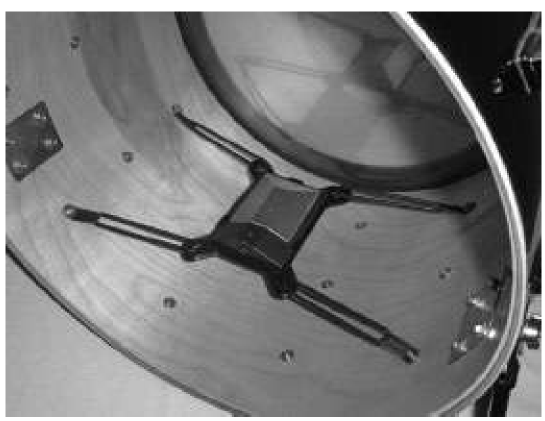


Figure 20-198. Kelly SHU FLATZ for use with boundary microphones. Courtesy Kelly Concepts, LLC.

References

1. Don and Carolyn Davis, *Sound System Engineering*, Indianapolis, IN: Howard W. Sams & Co., Inc., 1975.

2. A.M. Wiggins, "Unidirectional Microphone Utilizing a Variable Distance between the Front and Back of the Diaphragm," *ASA*, Vol. 26, no. 5, September 1954. (Parts copied with permission.)

3. R. Schulein, Shure Incorporated, "Development Consideration of a Versatile Professional Unidirectional Microphone," *Journal of the AES*, Vol. 18, no. 1, p. 44, February 1970. (Parts copied with permission.)

4. W. R. Bevan, R. B. Schulein, and C. E. Seeler, Shure Incorporated, "Design of a Studio-Quality Condenser Microphone Using Electret Technology," *J. Audio Eng. Soc.*, vol. 26, no. 12, p. 947, December 1978.

5. H. Tremaine, *Audio Cyclopedia*, Indianapolis, IN: Howard W. Sams & Co., Inc., 1969, pp. 148–150.

6. Ibid., pp. 150–151, 152.

7. A. P. G. Peterson and E. E. Gross, Jr., *Handbook of Noise Measurements*, General Radio Co., p. 33.

8. Don and Carolyn Davis, *Sound System Engineering*, Indianapolis, IN: Howard W. Sams & Co., Inc., 1975.

9. Henning Gerlach, "Stereo Sound Recording with Shotgun Microphones," *Sennheiser News*. (Parts copied with permission).

10. Yuri Shulman, "Reducing Off-Axis Comb-Filter Effects in Highly Directional Microphones," *J. Audio Eng. Soc.*, Vol. 35, no. 6, June 1987.

11. Y. Ishigaki, M. Yamamoto, K. Totsuka, and N. Miyaji, Victor Company of Japan, Ltd., "Zoom Microphone," *AES preprint 1718 (A-7)*. (Parts copied with permission.)

12. Manfred Hibbing, “XY and MS Microphone Techniques in Comparison,” *Sennheiser News*. (Parts copied with permission.)

13. Wesley L. Dooley and Ronald D. Streicher, “M-S Stereo: A Powerful Technique for Working in Stereo,” *AES preprint 1792 (J-6)*.

14. David A. Ross, “A Practical Approach to Applying the MS Stereo Microphone,” Shure Brothers, Inc.

15. C.P. Repka, “A Guide to Coincident Mikes,” *Audio Magazine*, November 1978.

16. Klaus Genult and Hans W. Gierlich, “Investigation of the Correlation between Objective Noise Measurement and Subjective Classification,” HEAD Acoustics GmbH, Aachen, Federal Republic of Germany.

17. Dr. Klaus Genuit and Wade R. Bray, “Binaural Recording for Headphones and Speakers,” *Audio Magazine*, December 1989.

18. H.W. Gierlich and K. Genuit, “Processing Artificial-head Recordings,” *AES preprint 2460 (G-5)*.

19. Don and Carolyn Davis, “In-the-Ear Recording and Pinna Acoustic Response Playback,” *AES preprint 2874 (W 3/5-E)*.

20. Joe Ciaudelli, “Large Multi-channel Wireless Microphone Systems,” *Sennheiser Electronic*. (Parts copied with permission.)

21. “RF-Transmission Technique, Wireless Microphones and Reporters Sets,” *Sennheiser Electronic*.

22. G.W. Plice, “Microphone Accessory Shock Mount for Stand or Boom Use,” *J. Audio Eng. Soc.*, Vol. 19, no. 2, February 1971. (Parts copied with permission.)

Bibliography

J. Eargle, *The Microphone Book*, Focal Press, Burlington, MA, 2004. *Microphones*, New York: AES, 1979.

Don Davis and Eugene Patronis, Jr., *Sound System Engineering*, Focal Press, Burlington, MA, 2006.

David Miles Huber, *Microphone Manual*, Howard W. Sams & Co., Indianapolis, IN, 1988.

* Ina space where the microphone will be exposed to both direct and reflected sound, even though the source of the direct sound is "off axis", the reflected sound will not be: hence less than absolute rejection.

Chapter 21

Loudspeakers

by Tom Danley and Doug Jones

21.1 Introduction

There are many ways that the subject of loudspeakers can be handled. In previous editions of this handbook, the authors of this chapter have treated the topic with considerable detail and precision. The reader is referred to the previous edition for a more mathematical approach. In this edition, we are hopefully no less precise, but perhaps somewhat less mathematics and more descriptive about how loudspeakers behave and the principles that govern their behavior. This version should not be seen as a replacement for the previous. See it as another perspective on a fascinating and complex subject.

21.2 The Medium, the Sea of Air

We live in an ocean of air and as the late audio great Richard Heyser said, "sound is what happens when air gets pushed" so this seems like a good place to start. The ocean of air is actually an immense collection of gas molecules. This collection of molecules is attracted to the Earth's exterior by the force of gravity. We live at the bottom of that ocean. To the first order, it is not difficult to imagine all these molecules acting like a huge collection of tiny but springy balls, all in contact with each other. Each ball is very light but it does have mass as well as a spring effect. It is easy to picture the transfer of kinetic energy that takes place between the molecules in one dimension. Newton's cradle is a good way to visualize what happens when you "push air", Fig. 21-1. Add force to the first ball and you can see that the energy is transmitted to the last one. Or, line up a dozen pennies all touching and in a straight line. If you flick a penny into one end, a penny will project off the other end. Each penny obviously has mass but they are also very stiff. Pennies are not very compressible which means that their “spring” is very strong. With the pennies or the spheres in Newton’s cradle, the speed of the kinetic transfer is faster than you can see but it has a finite or measurable speed. The air too, has a “speed of sound” governed by the same factors, springiness, mass and proximity.

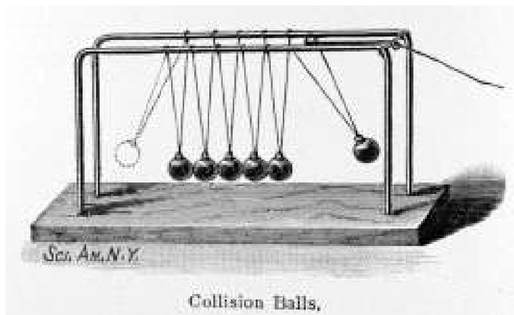


Figure 21-1. Newton’s cradle.

This sea of air is about 80mi (129km) thick with three-quarter of it located below 36,000ft (11,000m) or about 7mi. Although the molecules are very light, an 80 mile thick layer ends up producing a static or steady pressure of about 14.7 pounds per square inch (psi) (1033g/cm^2) measured at the earth’s surface. Picture yourself standing on the ground with a 80 mile thick layer of molecules above you. Like a fish in the water, you are unaware of the static pressure. But imagine a sensor that would convert air pressure into a proportional voltage. If the reference was 0psi (vacuum) the sensor would produce a voltage proportional to 14.7psi. Microphones (of the pressure sensitive type) normally do not measure down to dc or 0Hz relative to 0 psi, otherwise the microphone voltage would be offset proportional to 14.7psi static air pressure. The sound we hear is actually a very small change in pressure plus and minus, superimposed on top of the static pressure offset of 14.7psi. To appreciate how small these variations in pressure are, consider that at 132dB SPL (very loud), the peak pressure is only around 0.0138psi ($0.97 g/cm^2$) greater than the static pressure.

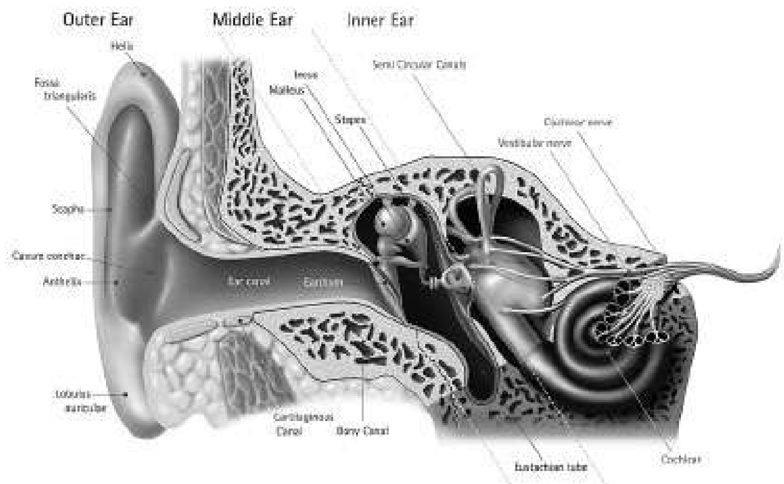


Figure 21-2. Image of inner ear.

The inner workings of your ears are pressure sensors too. Your Eustachian tubes act like a “high pass” filters by equalizing the pressure at low frequencies, Fig. 21-2.

The departure from “flat response” can be seen by looking at the ears’ response curve or equal loudness curves. This low frequency “roll off” is why we don’t hear the changes in barometric pressure with weather and other large very low frequency sounds all around us. The ear’s natural roll off in sensitivity seen in the equal loudness curves is also one of the things that makes low frequency loudspeakers difficult to produce and physically large as we will see, Fig. 21-3.

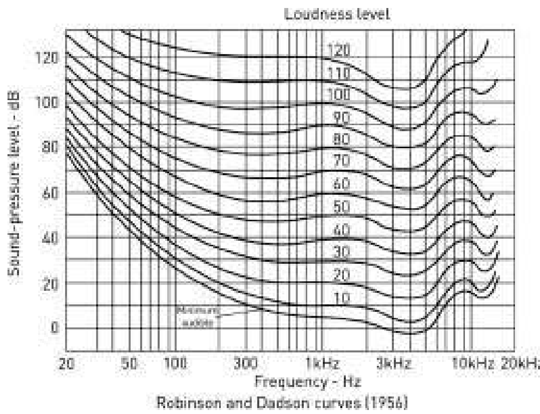


Figure 21-3. Equal loudness curves.

Imagine a concrete bunker with air tight steel doors. On one wall there is a piston which is in a cylinder recessed into the wall. If the piston is pushed in, the air in the cylinder is added to the fixed space in the bunker and so the pressure in the bunker rises according to Hydraulic theory and Pascal's law. Conversely, if the piston is withdrawn to the bottom of the stroke, it slightly evacuates or lowers the pressure in the room again following these same laws. This is easy to visualize, but let's take it one step further. Imagine you could see air pressure in super slow motion. You would see that the pressure in the room didn't rise instantly, rather the pressure increase occurred as the piston moved and that pressure radiated away or propagated into the room at "the speed of sound". The piston in this analogy is a transducer. The pressure wave moving into the room is an example of radiation. But before we consider radiation, let's examine the transducer.

21.3 Transducers

Transduction has a number of meanings but in audio it refers to the mechanism which converts one form of energy to another, normally from mechanical motion to an electrical signal and vice versa. Often transducers are bi-directional, for example some loudspeaker can be microphones and some microphones can be (weak) loudspeakers.

There is a temptation to use the words energy and signal interchangeably but actually there is an important distinction. Energy, work or power is always the product of two things. In the mechanical domain, the power delivered by a motor shaft is the product of the speed of rotation AND the torque or twisting force being exerted. The term Horsepower was coined to describe doing a specific amount of work in a specific time and was based on measured output of a horse turning a mill wheel. One horsepower is the equivalent of lifting 33,0001bs (15,000kg), a distance of 1ft (30.5cm), in 1min or any other combination of weight and distance with the same product. So technically, you could choose to generate 1hp by lifting 10lbs (4.5kg) a distance of 3,300ft (1006m), but you would have to do it in one minute as well! In the electrical domain, "power" delivered to a load is the product of the voltage across the load and the current flowing through it, Fig. 21-4. From Ohms law we can see that the relationship or proportion of voltage and current being delivered is set by the load resistance. To be clear, one can have a thousand volts with a current of 1 A, and dissipate 1000W, or one could have 1000A and 1 V and dissipate 1000W. By the way, horsepower and watts are two ways of quantifying the same thing. One horsepower of energy = 746W.

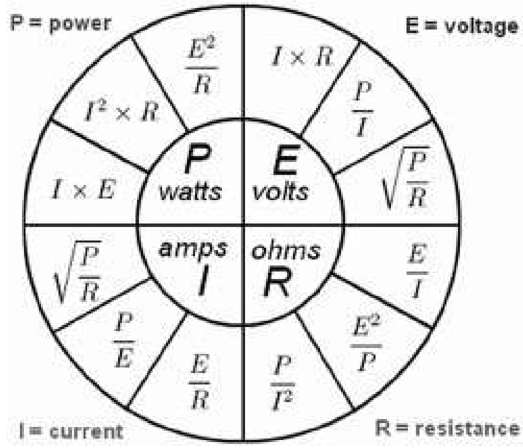


Figure 21-4. Ohm's Law wheel.

In the mechanical domain we can use a lever to adjust the ratio of the distance moved to the force applied, or we use a pulley or gear reduction to match the fast turning but lower torque motor to slower turning load requiring greater torque. If you ever started off in 10th gear on a ten speed bicycle, you find your legs are a poor match for the load. But if instead of legs you had slow moving but very high force hydraulic cylinders, the match would be fine. That is the point of the gears on a 10 speed bike. They allow the rider to match the force and speed of pedaling to the load. Another analogy that may be helpful, consider a professional baseball pitcher. If you hand him a whiffle ball and ask him to throw it at 90mph (145km/h) he may actually harm his arm. The mass of the whiffle ball is not appropriately matched to the force in his arm. The load is too small. Hand him a baseball, and he is able to transfer a great deal of energy to the load. Hand him a bowling ball and now the load is too big for the energy available.

In the electrical domain we use a transformer to match the source to the load to the required proportions of voltage and current in a similar way. Our power distribution system is a voltage referenced system. This means that the voltage doesn't change with a changing load and so the current drawn from a transformer can be anything up to the maximum rating for the transformer.

In the “acoustic domain,” i.e., when dealing with sound radiation, the impedance can be transformed with either a Helmholtz resonator over a narrow band or with a duct or pipe at a series of length related frequencies. For more broad band applications a horn can be a very effective acoustic transformer.

In the radio wave/electromagnetic domain, the radiation impedance of free space can be transformed with an resonant antenna or one that is tunable over a wider bandwidth, like a horn, log, rhombic or biconical, forms which are, in effect, resonant over a wider span of frequencies.

The loudspeaker driver radiates “power” in the form of sound. Sound energy like all the other types of energy, is transmitted in two forms, alternating pressure and alternating particle velocity. The load, is the mass and spring effect of air that must be “pushed” along with a small amount of resistance or absorption. This combination of properties is called the acoustic impedance. This small amount of resistance or absorption is the dissipation or the resistive loss part of the process involved in the continuous exchange between velocity and pressure, the kinetic and potential energy of the sound wave. There is some energy loss in the viscosity and other issues in of our less than perfect sea of air and it is those losses which explain why sound doesn’t travel forever, it gets absorbed. A tuning fork, struck mechanically in a vacuum, will ring much longer because only the mechanical losses are present, the radiation resistance was removed when the air was pumped out. Technically the power dissipated in the air, heats it slightly.

At the beginning of this section we referred to the fact that signal and energy are different. Here is the difference. In analogue audio a sound is converted to an electrical signal by representing it as a voltage. A measurement microphone for example will produce a specific voltage when exposed to a given sound pressure level or reference, and will do so over a very wide frequency range. What is captured is a representation of sound pressure. The voltage that is produced is proportional to, or a direct result of, the variations in pressure above and below the 14.7psi (1033g/cm^2) “reference” over a large range of frequencies.

When we drive a typical loudspeaker, we are primarily concerned with the voltage that is applied to it. In a perfect world, a loudspeaker would produce sound that would be directly related to the voltage applied to it. So 1V at 100Hz would produce a 100Hz signal of X sound pressure, and 1V at 1000Hz would produce the same sound pressure at 1000Hz and so on. We would call this a “flat frequency response”. It would radiate an acoustic pressure that is proportional to the voltage applied to it. If resistance remained constant, we could also say that the acoustic pressure was proportional to the current and also to the power, which is the product of current and voltage. However the audio signal is alternating current, where the current flows in both directions alternately, just as the pressure we seek to produce is alternating above and below the ambient air pressure. In addition we also find there are “reactive” components like capacitors and inductors which do not act like resistors and can temporarily store energy like a spring or moving mass. In these ac systems we find that R does not remain constant with frequency. The R like quantity which changes with frequency we call impedance. Since the impedance changes as a function of frequency, the power will fluctuate dramatically with frequency even if the drive voltage remains constant.

There are two sides to the transducer itself, an electrical side and a mechanical side and each of these sides has two terminals or outputs. For example, to connect a voltage across the device, there needs to be a complete circuit and so there are two connections. Acoustically a loudspeaker, like a typical woofer, has a radiator which is driven by the motor. The moving part, or radiator has two sides which are always equal, on one side of it there will be a positive pressure, that is greater than 14.7psi, but at the same time there will be an equal and opposite pressure on the other side. In the case of a woofer, allowing the two equal but opposites to combine around an open driver results in most of the sound pressure from one side being canceled out by that on the other side. To increase the output especially at the lower frequencies the output from the two sides must not be allowed to mix. There are many strategies, but most involve the use of a baffle, which can be anything from a hole in the wall to a sealed box.

Here are four examples of transduction mechanisms that show some of the variety of what is possible. Electrically, there are transduction mechanisms which produce force proportional to voltage and others which produce force proportional to current.

21.3.1 Voltage to Force Transducers

First, let's consider voltage causing a force. There are two good examples of this. One is the electrostatic loudspeaker and the other is the piezoelectric element.

21.3.1.1 Electrostatic Transducers

In the case of the electrostatic transducer, for a fixed spacing between elements, the electrostatic force is proportional to voltage. The greater the voltage the stronger the force. This sounds like the thing that makes a balloon sick to your hair on a dry day and it is, that same attraction is the force used here but applied over a large surface area. In this type of transducer, Fig. 21-5, the diaphragm is a thin sheet of plastic on which has been deposited a very thin layer of conductive material. It is supported by multiple small elastic elements that hold the diaphragm in place but permit it to follow audio-signal waveforms. The thin film is charged or polarized through a very high value resistance to a voltage typically of a thousand volts or more. The signal is applied as a differential signal on the grid or electrodes on either side. The result is the audio signal applied to the electrodes causes the film to be attracted more to one side or the other in proportion to the voltage. The electrodes on each side of the diaphragm are acoustically transparent to avoid pressure effects from trapped air as well as to permit acoustic energy to propagate away from the diaphragm. This type of construction permits the diaphragm to be of arbitrary size. The performance per unit area is the same for any area of the diaphragm. The actual loudspeaker is a thin surface often curved in the horizontal, forming a section of a cylinder. A surface that is large with respect to wavelength becomes increasingly directional at high frequencies.

Since an electrostatic loudspeaker is designed to couple directly with the acoustic resistance of air, the mass of the diaphragm is quite small and can be neglected with little effect on the accuracy of predictive models. The velocity of the diaphragm is directly proportional to the electrostatic force applied, except as altered by the stiffness of the diaphragm suspension.

Measurements indicate that for a constant voltage applied to the electrodes, the acoustic response is uniform (flat) to well beyond the range of human hearing.

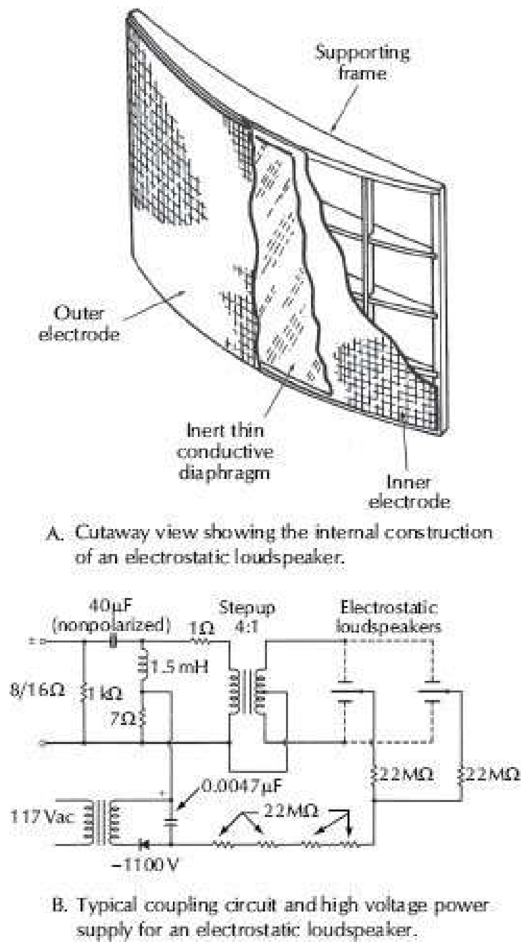


Figure 21-5. Typical electrostatic loudspeaker.

Output at low frequencies is limited by the maximum linear amplitude of the diaphragm motion, which is determined by spacing between the diaphragms and damping in the suspension. The maximum power output from an electrostatic loudspeaker of a given diaphragm area is determined by the strength of the electrostatic field that can be produced between the diaphragm and the electrodes. Some designs actually placed the driver within a “plastic bag” filled with an electrically insulating gas which allowed higher voltages before arcing over.

An electrostatic loudspeaker is seen by an amplifier as a capacitor with a value typically on the order of 0.0025μF from electrode to electrode. Thus, the magnitude of the impedance presented by the loudspeaker to the output of the amplifier falls off at 6dB/oct as the frequency is increased. This presents some problems for driving electrostatics, as many amplifiers are not designed to drive purely capacitive loads.

Because electrostatic loudspeakers are relatively large in area compared to cones, their directivity is high in comparison to cone systems. Various schemes have been used by designers of electrostatics to address this issue. The Quad ESL63, Fig. 21-6, is one example. In this loudspeaker, the diaphragm is broken into circular rings with each driven through an electrical delay line so that the center element radiates first, then the smallest ring, sequentially outward. The object is to radiate spherically, as if the sound originated from a point well behind the loudspeaker, thereby making them wider in dispersion than a large single panel.

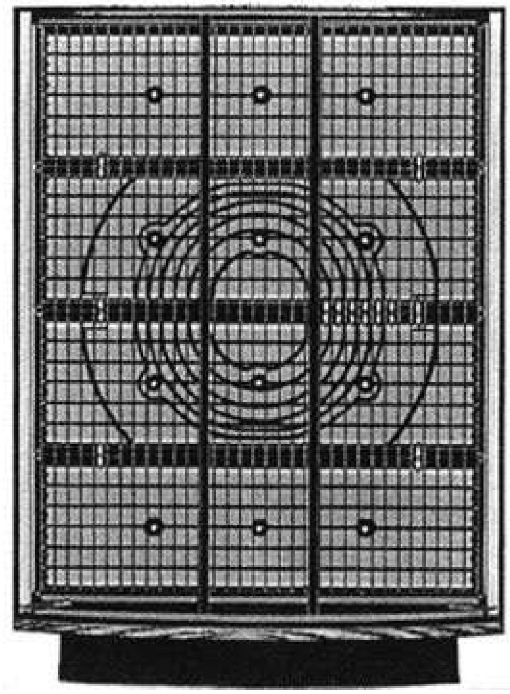


Figure 21-6. Quad ESL63 loudspeaker.

21.3.1.2 Piezoelectric Transducers

The other kind of voltage transducer is the piezoelectric transducer. *Piezoelectricity*, or pressure electricity, was discovered in the 1880s by the Curies. Certain materials have the remarkable property of deflection or mechanical strain when a voltage is applied to it. Just as remarkable, this piezoelectric material works both ways. When the material is deformed or stressed by an outside force it produces a voltage! Piezoelectric material can therefore be used to create both loudspeakers and microphones or pickups. It is today a feasible motor drive mechanism for loudspeakers. In a piezoelectric material, a voltage applied to the material will result in a mechanical strain or deflection. The reverse is also true, and piezoelectric elements can be used in microphones. This characteristic is attractive for direct-drive units such as ultrasonic devices. For loudspeakers, however, some means must be applied to mechanically amplify the inherently low excursion so that a loudspeaker diaphragm may be driven properly.

One of the earliest discovered piezoelectric substances is was Rochelle salt. Although Rochelle salt is still widely used, it suffers from poor mechanical strength, low temperature breakdown (55°C [131°F]), and extreme sensitivity to humidity. Quartz is also piezoelectric and was widely used as a tiny mechanical resonator locking radios on to their assigned frequencies. It is the time keeper in "quartz" watches and the tiny electronic part that sets the clock frequency in computers and other digital gear.

Barium titanate is was the first piezoceramic to be developed. Although it is not as electrically sensitive, it is still widely used, exhibiting many superior characteristics over Rochelle salt. The most widely used piezo material today is lead zirconate titanate, developed first in Japan in the 1950s. This material (PZT) is now highly refined and exhibits the best properties of any piezo material for loudspeaker use.

In the case of the piezoelectric element, the voltage potential across the internal lattice produces a large internal force, causing its dimension to change. One near-optimal application of piezoelectric drive is underwater use in sonar systems. This is due to the excellent impedance match of the piezoelectric material to water via a waterproof barrier. Although swimming pool loudspeakers using standard electromagnetic drivers are also available, the piezoelectric configuration is more efficient due to its mechanical impedance match to water. The loudspeaker is fixed to the side of the pool and driven like a conventional loudspeaker.

On the mechanical or acoustic side, the electrostatic transducer and the piezoelectric transducer look very different, Figs. 21-5 and

21-7. The electrostatic transducer radiator and motor has a low moving mass, the thin film while the piezoelectric element is a solid material so each has a very different proportion of motion to force when delivering power into an ideal load, its acoustic impedance. Both of these are reversible forms of transduction. An electrostatic driver when polarized will act like a microphone when exposed to sound and a piezoelectric element will generate a voltage when under changing stress. Warning!! Piezoelectric transducers can do something others types can't do. A large well-made piezoelectric transducer, especially the type frequently used in submarines, can really hurt you. The internal stress caused by changing temperature can generate and then store a charge of hundreds to thousands of volts. If you pick it up carelessly it can zap you just like a charged capacitor.

21.3.2 Current to Force Transducers

21.3.2.1 Electrodynamic Transducers

Transducers which user current to create force are much more common. Most loudspeakers are the electrody-namic type. These loudspeaker motors are based on another reciprocal transduction which we call Faraday's law. Faraday's law describes the transformation of changing magnetic flux into voltage and changes in voltage into magnetic flux. In these speakers, the stronger the current, the greater the current induced magnetic field, and the greater the permanent magnetic field, the greater the force the interaction with the wire produces. The typical voice coil loudspeaker driver has a coil of wire in an annular magnetic gap so that a long length of wire can be placed within an efficiently shaped magnetic gap. Actually there are many “laws” which govern the electro/magnetic behavior of the transducer. The reader is encouraged to review Faraday, Lenz and Lorentz as a way to understand the interaction between magnetism and electrical current. It should be noted here that only the portion of the wire in the magnetic gap produces the force. However the current passes through the entire coil and so heats all of it. Commutation is used in rotary motors to reduce the R part. As the motor rotates, the commutator continuously selects which sets of coils the current is flowing through. That way, only the portion of the wire in the magnetic field which is actually producing force is conducting the current and being heated.

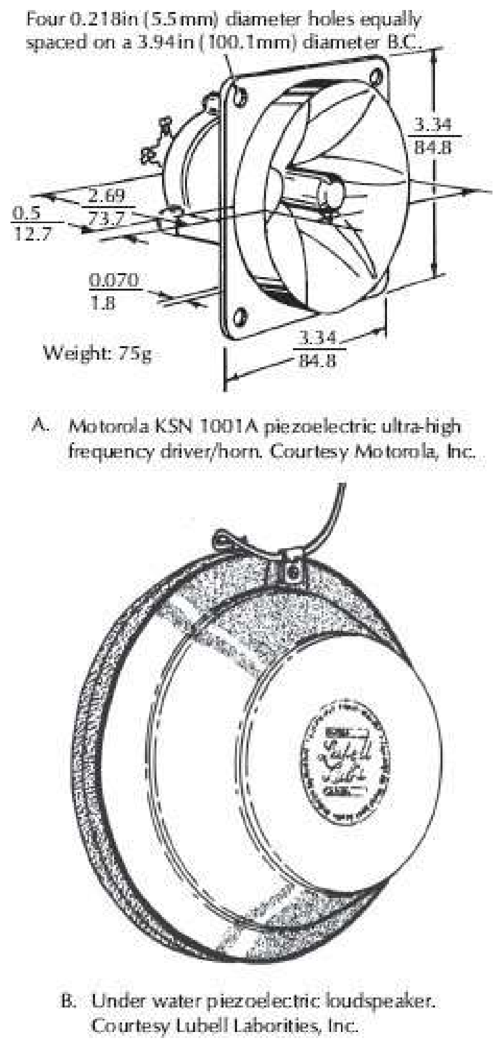


Figure 21-7. Piezoelectric loudspeakers.

The standard measure of this part of a transducer's force conversion is called the *BL* factor which is the strength of the magnetic field applied to the voice coil. It can be read interchangeably as the force in Newton's per ampere of current or Tesla/meters (i.e., the length of wire in meters) in a 1T (10,000 Gauss) magnetic field).

An interesting dual property exists here too. When a voltage is applied to the voice coil and it moves because of the opposing magnetic field in the gap, the force it exerts is proportional to the current. When the conductor is put into motion, it becomes a generator and produces a voltage proportional to the velocity. This is seen in Lenz's law. The greater the *BL* or force per ampere, the greater the voltage it produces when moving. This property is called "back emf." Emf is of course electromotive force, and it is another way of saying voltage. We use the term "Back emf" to distinguish this voltage which is the result of the motion of the transducer from the voltage which is *applied* to the transducer. We will return to back emf shortly.

While the *BL* describes the force per ampere that motor produces, a figure of merit for such a motor might be better expressed as Force per unit of dissipation and is more like either $BL^2/\sqrt{\hat{R}}$ or $BL/\sqrt{R}$ since the power dissipated in the wires resistance is proportional to $I^2 \times RI^2 \times R$ or the voltage across it times the current flowing through it.

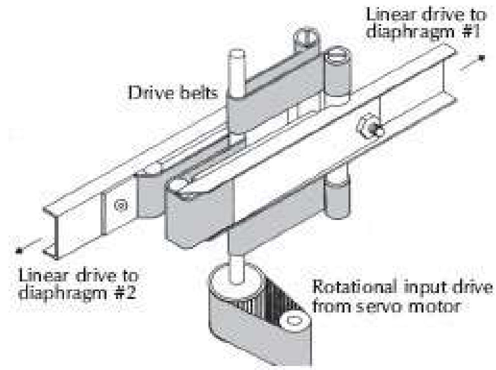
Here is another dual property. A coil can have a few turns of heavy wire or many turns of fine wire, as long as there is the same amount of wire in the gap and the magnetic field is the same, there will be the same figure of merit but wildly different load resistances. The standard voice coil loudspeaker is this kind of transducer.

Most brush type permanent magnet dc motors are rotary equivalents to a wire in a magnetic field. The force or torque can be specified, as well as the voltage per rpm. These motors will run in one direction when connected to a battery or dc source. However if they are driven with an alternating signal like audio, they will oscillate back and forth following the signal.

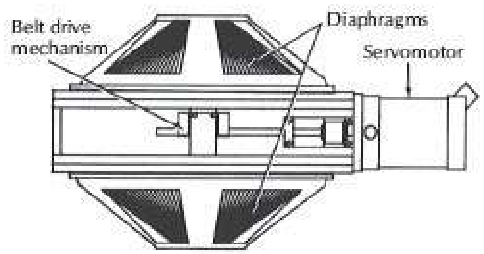
Motors of this sort were used in the 1980s and 1990s to produce low frequency audio in the Servodrive loudspeaker shown in Fig. 21-8. The low inertia dc servomotor used the in Servodrive had a torque constant of 9.1oz in (6.4N-cm) of torque per ampere and generated 6.7 V/1000 rpm when turning. Earlier we mentioned that a commentator has a distinct advantage in that the coil is much more efficiently situated in the magnetic field. This was one reason that the Servo-Drive was developed.

21.3.2.2 Back Emf

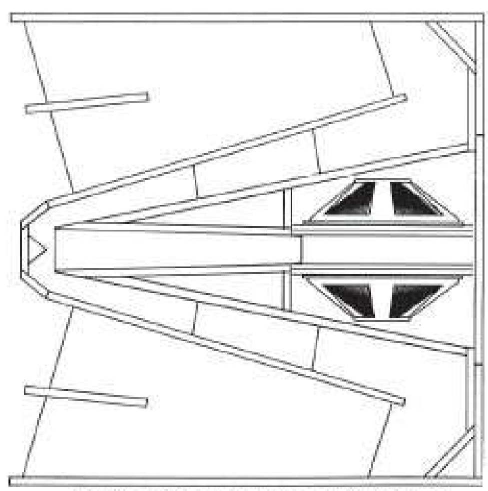
Before going on, an understanding of back emf may give us a better understanding of how the electrody-namic transducer works. It is perhaps easiest to consider a simple permanent magnet dc motor. This will allow us to more easily examine the phenomenon. The principles are exactly the same, but maybe easier to comprehend by considering a spinning motor than an oscillating voice coil.



A. Belt driven system.



B. Position of the belt drive and opposing diaphragms.



Position of the diaphragms on a folded horn.

Figure 21-8. Servo-Drive Loudspeaker (SDL).

If we connect our dc motor to a battery, it will run continuously in one direction at one speed, for as long as the battery supplies current. Note that it runs at a *constant* speed. It does not run faster and faster until it blows up. It is not under constant acceleration. It looks like there is a governor somehow setting the speed limit. If we reverse the polarity of the voltage, the motor will switch direction, but it will also run at a constant speed. Now, if we drive the motor with a slowly changing voltage, the speed will slowly change proportional to the voltage. If we drive it with a slowly changing, i.e., low frequency, ac signal, the motor would rotate one way, slow down, stop and reverse direction for each cycle of the drive signal. It is the voltage which determines the speed, while the current applies the force to make the motor move. The governor is actually the back emf.

So consider what happens the moment the switch is closed, connecting the battery to the motor, Fig. 21-9.

We can plot time on the horizontal axis and the voltage across the motor, the current through the motor and the speed the motor is turning and plot them all on the vertical part of graph, Fig. 21-10.

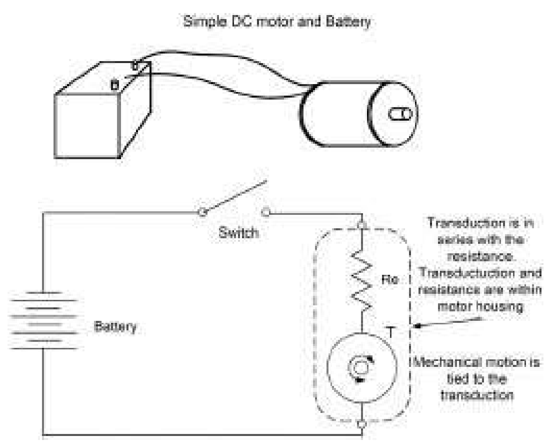


Figure 21-9. Simple circuit switch battery motor.

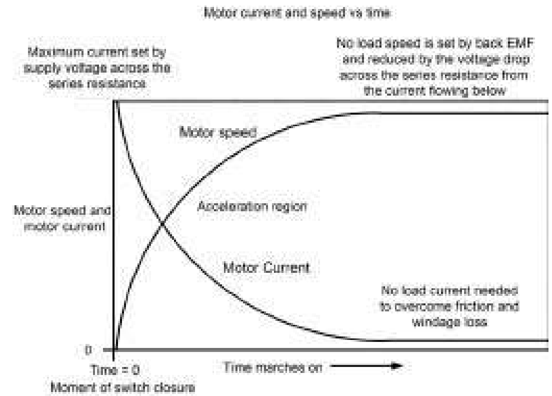


Figure 21-10. Motor speed graph.

The part of the motor where the actual transduction is taking place is a coil of wire (the rotor) suspended in a magnetic field. This wire can be seen as a dc resistance in series with the actual transduction process; the resistance due to the wire itself. When the switch is closed, the battery current flows through the transduction mechanism including the dc resistance. A magnetic field is produced. But in that first instant, since the rotor has not yet started to move, there is no back EMF…Yet.

Of course the amount of current that will flow is governed by the voltage and the load which in this case is the dc resistance of the wire in the rotor. To put numbers on it, let's imagine a 12V battery and the dc resistance of 10 Ω From Ohms law we know there would be 1.2A flowing. That 1.2A is converted into a torque or twisting force (or "push/pull in a voice coil). That force begins to accelerate the mass and rotates the motor. As soon as the rotor begins to move, voltage or back emf is produced by the transduction mechanism. The back EMF effectively reduces the voltage that the battery is applying to the load, so the current and therefore the accelerating force begins to fall. After some time, the motor has accelerated to the speed where the back EMF equals the voltage from the battery. The voltage required from the battery drops to 0, no current is flowing, no torque or force is being exerted. OK, in the real world, the current does not actually fall to zero, there is force required to overcome the friction in the bearings, windage (internal wind resistance) and other losses which will always require some current/torque to overcome, but it will be very small. If you place an ammeter in series with the motor this phenomenon can easily be observed. Now if you put a load on the motor, say you connect it to a fan, if you watch the ammeter, you will see the current increase. As the current increases the force will increase as well. You might well ask how the battery feels the mechanical load on the motor. As the mechanical load on the motor increases, there is small drop in speed or rpm. When the rpm drops, the back EMF drops and the voltage that the battery is supplying increases, and the current and therefore the force, in this case torque, increases. Under a load, the motor now runs at a slightly slower speed. But this is a chapter on loudspeakers, not dc motors. Well, within the mechanical limitations of a voice coil, (X_{max}) the exact same thing happens in a loudspeaker. As the coil exerts force first one way and then the other and accelerates the moving mass of the system, the force is governed by the BL product, the back emf and dc resistance. When driven with a typical ac audio signal, and you observe the voltage and current at the electrical terminals, the moving mass of both the motor and the voice coil transducer behave like a series resistor and capacitor to ground over much of its range. We will address this more later.

These types of transducers exhibit behavior seemingly opposite to the piezoelectric transducer. When not connected to an amplifier are easy to move and depend on external spring forces to return them to the center of motion, the rest position. The loudspeaker spider and edge suspension provide this function and when in an enclosure, the “air spring” from the box is in parallel with or adds to the stiffness of the spider and suspension spring.

21.3.2.3 Magnetostrictive Transducer

The magnetostrictive transducer is rarely used in audio but is but very useful in mechanical sound and underwater use. Iron, Nickel as well as many proprietary metallic compounds have the property that the internal stress from passing a magnetic field causes the dimension to change slightly but like the piezoelectric, producing a large force and small motion. Since the effect is proportional to the magnetic field, passing a signal through a coil wound around the material, causes it to produce the second harmonic of the signal since the magnetic field goes up and down twice per cycle. To make these devices produce the true input signal, a magnet is usually placed in the circuit that biases the material to about half of its saturation level. The current in the driver coil now modulates the field up and down relative to the bias point set by the magnet. This bias point could also be set using dc. Like the ceramic piezoelectric element, this is a solid material in comparison to the loudspeaker transducer but like the piezoelectric makes up for a tiny motion with a very large force. The place one is most likely to hear this type of transduction is the inadvertent hum or buzz that some power transformers make as a result of the laminations being slightly magnetostrictive.

21.3.3 Diaphragm types

Regardless of the type of motor, and the type of transduction, there needs to be a mechanism that connects the driving force to the sea of air. Generally this is by means of a diaphragm. The ideal diaphragm would have very low mass to allow it to move and change direction quickly. The ideal diaphragm would also be rigid so that it will not distort as it pushes against the air. The most common direct-radiation device is the paper cone driven by a cylindrical voice coil. The cheapest cone to make is the folded cone, which is cut from a sheet of paper, rolled, and bonded at the seam. A more expensive and difficult to make cone is the molded-paper cone. These are one piece, molded by straining a slurry of water and paper pulp through a strainer mold in the shape of the desired end product. The formed wet mat of pulp is then pressed and baked to remove residual moisture, bearing a dry, strong one-piece cone, free of joints. Ribs and concentric rings are sometimes molded into the cone to strengthen or dampen resonant modes within the cone, and the cones can be formed with straight or curved sides of varying depth. These and many other options are all available from suppliers of cones. At first it might appear that the lightest stiffest materials might be the best and they would be except for another important property, internal damping. A material that rings like a tuning fork or bell is not desirable. So to summarize, the ideal cone should cheap, as light and stiff as possible, have sufficient damping and be easily glued.

A soft dome, Fig. 21-11, is a common type of radiator which operates on a different principal than a ridged piston. It is generally a cloth dome shape that is impregnated with a damping material that is often tacky or sticky. At some frequency near the top of its operating range, a piston radiator will stop acting like a piston. At this point the behavior becomes hard to predict as it may transition into one or many resonant modes as the frequency climbs. All of these behaviors are undesirable and are why internal damping is an important feature of piston/cone radiator design. A soft dome is a good choice, as the dome has very little strength or stiffness. The motion from the voice coil at its perimeter travels as a wave through the damping material and depending of the frequency is partially absorbed internally before reaching the tip. This can provide a very well behaved low efficiency source. The down side is that both temperature and age can reduce the effectiveness and properties of the damping material itself.

Figure 21-11. Close-up of a soft dome tweeter.

An annular radiator is another shape that is somewhat common in high frequency drivers. Since moving mass is a limitation on the high frequency response of a horn driver, the annular radiator allows smaller lighter radiator system for a given motor strength.

As shown in Fig. 20-12, when a current is passed through a conductor, in this case a flat ribbon, a magnetic field is produced which is either attracted to or is repulsed from the permanent magnet surrounding it on two sides. This ribbon motion causes the broad side of the ribbon to move in and out of a close fitting magnetic structure. When it moves, the ribbon displaces the air along its length in the perpendicular or in and out direction. In the ideal case, the entire length of the ribbon moved uniformly with a single motion. One novel approach to motor design involves printing or etching a conductor onto a thin sheet of Mylar™ (0.0005in [0.0127mm]) then folding it to produce and pleated diaphragm that is forced in the magnetic field. The ribbon loudspeaker is a special case in which the voice coil serves as both conductor and diaphragm In another implementation, continuous lengths of wire are bonded to a large panel of Mylar™, which is operated over a field of bar magnets. The leaf tweeter, Fig. 21-13, is similar, etching a conductor field on Mylar™. They are identical in principle to the ribbon loudspeaker.

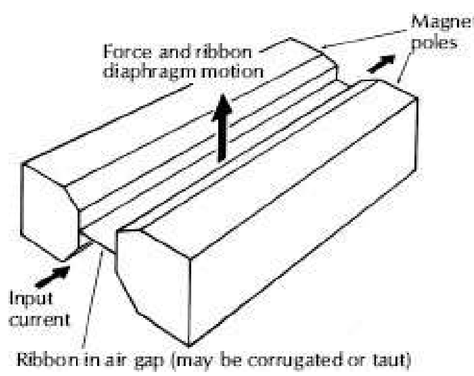


Figure 21-12. Ribbon loudspeaker.

Suspension Methods

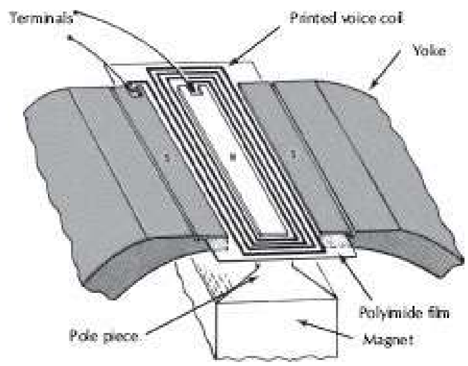


Figure 21-13. Technics Leaf Tweeter diaphragm detail. Courtesy Panasonic Industrial Corp.

The suspension of a cone driver serves the primary purpose of maintaining the voice coil alignment. It also provides a restoring spring force or centering force to the cone and motor in the axis of motion. It is comprised of two distinct components: the surround and the spider. Occasionally there are two spiders. The surround is attached to the periphery of the diaphragm or cone and is itself attached to the support structure (the basket in the case of a cone driver). The spider is attached to the voice coil former (or to the cone in the vicinity of the former) and is also attached to the basket on its periphery. Because they affect cabinet sealing, surrounds are designed to be nonporous. Surrounds and spiders both contribute to the damping of the motion of the diaphragm in the form of corresponding mechanical losses. The most popular surround construction is heat-formed, open-weave, resin-impregnated linen with formed-in convolutions and sealed with damping dope. Other surrounds are made of foam or butyl rubber formed in a half-roll. On some loudspeakers, a viscoelastic (never-drying) dope is applied to the surround.

Spiders are usually made of a heat-formed, open-weave, resin-impregnated cloth that is formed into convolutions. They are usually not treated with a sealing material (dope). The unsealed fabric is needed for venting, since the air beneath the spiders can otherwise be trapped. This also tends to damp the spider. The spider is not required to seal the edge of the cone to its enclosure as is the surround. In a typical cone driver, the spider contributes the majority of the stiffness in the suspension and in maintaining voice coil alignment.

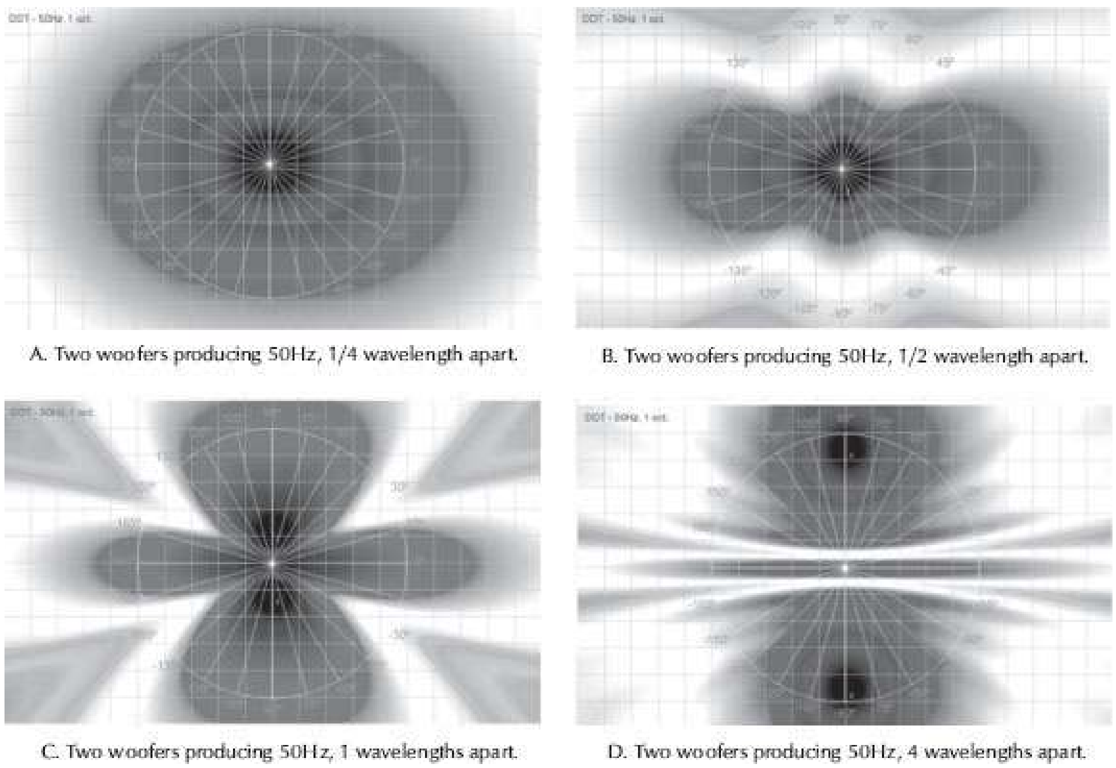
21.4 Radiation

In section 21.2, we discussed the properties of the sea of air that surrounds us and how a pressure disturbance propagates away from the source at the speed of sound. We then looked at the transducers which convert an electrical signal into sound. We now turn our attention to radiation, or what happens to the pressure wave after it leaves the transducer. To best understand this, let's return to the extreme case we introduced in section 21.2, the concrete bunker with an airtight door and a piston embedded in one wall. When the piston is moved inward towards the room, there is a change in pressure but no sound power being radiated away. Very nearly all the energy is contained within that bunker. The air in the bunker is effectively a spring, which stores the potential energy applied by the piston. If the force applied to the piston is removed, that room pressure would move the piston back to the rest position, and the net transfer of energy would be 0. No work will have been done, no power dissipated. If this is difficult to picture, consider this. If it were my hand that moved the piston to begin with, I would have burned calories in order to move the piston. When the pressure in the room pushed the piston back at me, my hand and body has no means of recovering that energy. From my perspective it is work that I have done and the energy was transferred into the system. But what if the device that moved the piston was able to recover that energy? As we will see, a loudspeaker does just that. In real life of course nothing we deal with is “perfectly anything” so there is always some energy being dissipated, but it’s very small in this special case.

Put furniture and carpet in the room and it is a different story. To the degree these items are porous, each change in pressure around them causes some of the air to move in and out through the pores of those materials. They sort of “breathe” in response to changes in pressure. The air moving though the fabric encounters resistance to flow, viscous drag, according to materials’ porosity and other properties and this *is* dissipation, some of the energy is absorbed and takes away energy from the formerly “pure” spring properties of the air in the bunker. That energy cannot be returned to the piston. It has been dissipated in the material, turned into heat. If you had a very sensitive thermometer, you could actually measure the temperature increase of an acoustic absorber like a big fuzzy couch. The temperature would increase according to how much sound energy it absorbs. Thankfully at sound levels we can tolerate, the energy per area is minute and the temperature rise would be infinitesimally small. An old acoustician once put this in perspective; he said if you had a room full of politicians all arguing with each other and were able to capture all that acoustic power, at the end of a week, you could heat a cup of tea.

One of the things that makes the design of loudspeakers so interesting and complex is the extremely large bandwidth that we have to cover. Since the bandwidth of our hearing is wide, around 3 decades or 10 octaves, it is very difficult to produce the entire audible bandwidth with a single driver. As the required acoustic power increases, the difficulty is multiplied. The frequency range must be divided up and assigned to high power drivers which have been optimized for each frequency range. One might assume that when two sources are added, the result is always more. That is not always the case. When two sources add, what you get depends on any phase difference between the two, and the distance from each source to the observer or microphone. For example, the reason we enclose one side of a woofer because one side while equal in displacement is opposite in pressure. When they add, they cancel each other out because they are 180° in phase apart, equal but opposite. When two equal sources like subwoofers are in phase, their combined magnitude depends on how far apart they are. Starting at a very low frequency and up to the point where the radiators are about 1/4 wavelength apart edge to edge, the sources combine into one new source with the same radiation pattern. The sources also “feel” each other’s radiation pressure and so they “move up” the radiation resistance curve and as a system they have a larger effective radiator and so, a higher efficiency. When you add two equal sources in this condition, what you get is 4 times more acoustic power; 2 times more because you have two drivers and two times the input power and 2 more times because the combined radiator is larger and 2 times more efficient. This condition can also be modeled another way. It can look like signals adding via resistors where there is only one resulting value. Once a radiator spacing of about 1/2 wavelength apart has been reached, the two sources begin to radiate independently with each one’s individual directivity. What you get depends on where you observe it from. As the loudspeakers move farther apart (relative to the wavelength) they produce an interference pattern marked by a pattern of lobes and nulls in its radiation pattern or polar plot, Fig. 21-14.

This pattern is the result of the two sources having a unique distance to each from any given point in space and so wherever the difference in the distances is equal to 1 or more odd number of half wavelengths or N odd times 180°, there will be a cancelation or null because they are opposite pressures, Fig. 21-15.



A. Two woofers producing 50Hz, 1/4 wavelength apart.

B. Two woofers producing 50Hz, 1/2 wavelength apart.

C. Two woofers producing 50Hz, 1 wavelengths apart.

D. Two woofers producing 50Hz, 4 wavelengths apart.

Figure 21-14. The effect of multiple woofers on coverage.

Conversely, where the two source pressures are in phase or an equal number of half wavelengths apart, they will add together and make a lobe. If two sources have no individual directivity or radiate in all directions by themselves like subwoofers, the interference pattern will have lobes and nulls over an entire sphere of radiation. A large array of individual sources can be constructed and by the same process, some of the sound will combine into a “beam”. It will also radiate a good deal of energy outside the beam due to the interference pattern as the array does not increase the directivity of the individual self-interfering sources. With enough sources the pattern of addition and cancelation average out when looking at the amplitude response. However if a broadband impulse were used as an input, what arrives at the listener is individual impulses separated in time from the closest source to the farthest source and not a single impulsive event in time like the original input signal.

It is rarely possible to place the upper and lower drivers in a multi-way system close enough together to avoid creating an interference pattern. Most try to fix it with crossover design. While a proper crossover design is well beyond the scope of this writing, to be sure in part it will require addressing the main lobe to be sure it points towards the listener as the upper and lower speakers in any multi-way system transition though crossover.

So far, we have not considered time, except in the horsepower formula and of course inferred whenever frequency was mentioned. Our audio signals also add like the sound in the interference patterns we just examined in that they are ac, half the time, they are a positive voltage and the other half negative. Unlike adding simple dc voltages, an extra consideration is in play. Since it is possible to have one signal which is at a given moment in time positive and another signal which is at the same moment in time negative, we can't simply take the sum of the two voltages. We must consider the difference in time between where each passed through 0 or changed from plus to minus. The convention is to divide one complete cycle into 360° and call this difference the phase angle.

With our sealed bunker experiment, as we move the piston back and forth more rapidly we can produce changes in pressure that actually become audible sound. From zero Hz upward the pressure our source produces is only tied to displacement, it produces the same sound pressure at any frequency for a fixed displacement. As the frequency of the piston increases to the point where the largest dimension of the bunker is about one quarter of one wavelength, of the frequency the piston is producing, things begin to change as it is no longer a sealed pressure system. The pressure that radiates away from the piston returns after reflecting off the far wall. It will return delayed 1/2 cycle or 180° and so it arrives out of polarity. Its pressure will be opposite from the source. A reflection from an obstruction 1/4 wavelength away, causes a cancellation notch where the returning signal is out of phase with the source pressure.

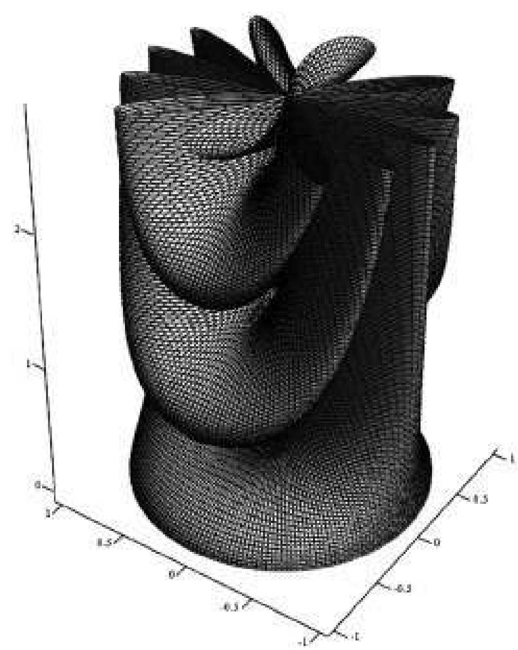


Figure 21-15. Stack of polars.

Increase the frequency further and at the point where the farthest wall is about 1/2 wavelength away, the pressure that reflected off the far wall, will returns in sync with the driving source, a condition which is additive or a resonant standing wave. Increase the frequency further and at the distance roughly equal to 3 × 1/4 wavelengths and you have another cancelation, move to 4 × 1/4 wavelengths and there will be another additive condition. As you go ever higher, there become so many of these addition and cancellations (every multiple above) that they average out in a measurement. If you have ever walked around in a room and heard a lot of bass in one place and very little somewhere else, it is likely that you are hearing room modes. These patterns of nulls and hot spots, are the result of interference patterns produced by the direct and reflected sound, see Chapter 6 *Small Room Acoustics*.

At the higher frequencies, much of that furniture becomes acoustically absorptive as the internal air paths become larger and larger compared to the wavelengths. Now the energy the source radiates is being absorbed and dissipated. Work is being done.

Consider the opposite extreme case. Let's move out of the concrete bunker to outdoors. Now *all* of the sound pressure from the moving piston radiates away and none of it returns. There is no pressure spring effect like the trapped air in the room and all of the energy is radiated away. This second case is actually much closer to reality even in a room at least at high frequencies. Usually we want to radiate sound to listeners and not pressurize concrete bunkers, although this same pressure mode or "room gain" is how very low bass can be produced within a car, when the windows are rolled up.

If we can understand the load, that is to say, understand what the transducer is radiating into, we can optimize both the transducer and the device which connects or couples the transducer to the air, called the radiator. Like all of the forms of energy discussed previously, radiated sound energy also has a dual property of particle velocity and pressure. We are normally only concerned with the pressure part as it is easiest to measure, but the velocity part is present as well as can be easily observed with a velocity microphone. Fig. 21-16 shows the radiation resistance of air for a piston radiator. The horizontal scale on the graph shows the size of the radiator relative to the wavelength being produced, the vertical scale shows the radiation resistance that the piston “feels” when oscillating back and forth.

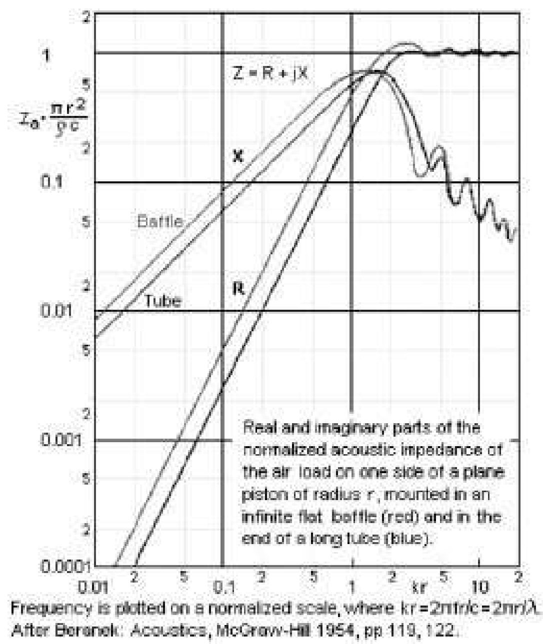


Figure 21-16. Radiation resistance of air.

When the radiator is small compared to the wavelength, it “feels” less resistance from the radiation and so would radiate less power for a given radiator velocity. At some point this relationship becomes optimal and no further efficiency is gained. This occurs roughly when the radiator circumference becomes larger than the wavelength it is producing. On the other hand when the radiator is large compared to the wavelength, there is no further advantage to be gained in radiation resistance.

Woofers generally operate in the range where the radiator is tiny compared to the wavelength. If you look at Fig. 21-16 you can see that the radiation resistance is such that if velocity of the radiator were kept constant, there would be a +6dB/octave rising frequency response. In fact, to make the response flat facing this changing radiation resistance, the radiator must be driven as a constant *acceleration* rather than constant velocity. The woofer then is a transducer which converts current into force, but is driven by a source which is referenced to voltage. If we want a flat response the signal must be decreased as frequency increases. This is accomplished by sizing both the moving mass, and the motor strength and taking into consideration the effects of the dc resistance in the wire of the coil.

Now, thanks to an electrical resistance (the resistance of the wire itself) in series with the transduction mechanism (the conductor in the magnetic gap) the constant voltage input signal has been converted to a current signal flowing to the coil. This is accomplished within the driver by choosing the moving mass, decreasing the motor strength and including the effects of the dc resistance of the voice coil. As discussed earlier, the moving mass when converted by transduction, appears as a parallel capacitor, driven by that resistance.

Electrically, a parallel capacitor driven by a series resistance forms a "low pass" filter, Fig. 21-17. Above the filter's corner frequency, the response will roll off at a rate of –6dB/octave. The mechanical equivalent shows the same behavior only it is the velocity which is decreasing. This is exactly what is needed to compensate for the increasing radiation efficiency as the frequency climbs.

Note here too that the radiator velocity represented by the voltage is what appears across the capacitor, not across the series resistor from the winding resistance.

When a diaphragm pushes the air, a positive pressure slightly higher than 14.7psi is formed on one side, and a negative pressure, slightly less than 14.7psi is formed on the other.

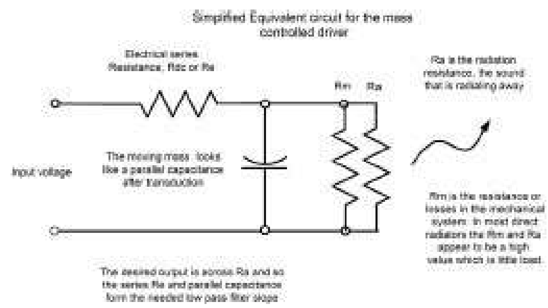


Figure 21-17. Equivalent circuit midband simple RC low-pass filter.

As seen earlier, when these two exactly equal but opposite signals are added together, they cancel each other out. +1 added to –1 equals zero. Obviously the way to avoid this cancellation is to prevent the pressure wave from the back side of the radiator from mixing with the wave radiating from the front. The simplest way to achieve this would be to place the loudspeaker in a sealed box to contain the radiation from the back, so that it can't mix with the desired radiation from the front. Placing a driver in a box does a number of things. It keeps the two out of polarity signals from ever mixing, it changes the restoring force on the driver by increasing the spring force and the energy radiated out of the back of the driver is not really used.

Earlier we discussed the restoring force or spring effect of the driver's suspension which returns the driver to the neutral or center position when there is no signal. In our equivalent circuit, this spring force appears as an inductor in parallel with the capacitance after being converted by the transduction. The air in a sealed enclosure acting on the movable cone “looks like” a spring force in parallel with the suspension spring force in the driver, making it stiffer.

In the mechanical world there are two forms of energy. Energy can be stored in things which move. We call this moving energy kinetic. Energy can also be stored in things due to their position. This positional energy is called potential.

Both can be seen in a simple pendulum, Fig. 21-18.

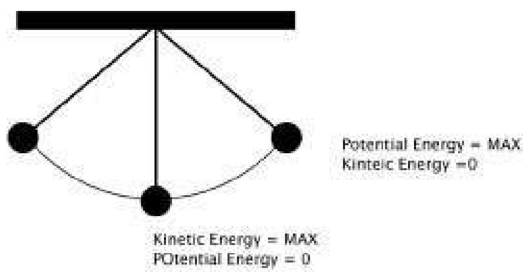


Figure 21-18. A simple pendulum.

In order to set a pendulum into motion, energy must be applied to the weight to move it away from its rest position. The energy is transferred from my body to the weight when I move the weight a few feet away from the center. That energy is now stored in the weight as potential energy. Note that as pure potential energy it is not really doing anything. It is potential. When the weight is released and is allowed to swing freely, the energy is traded back and forth between potential and kinetic. The kinetic energy is greatest when the pendulum crosses over the down or rest position. At this point there is no potential energy at all. Conversely, the point where the mass has the greatest potential energy is when it stops momentarily at the top of each swing, where there is no kinetic energy at all. In the pendulum, the stored energy is constantly exchanged back and forth between kinetic and potential, where one is the highest, the other the lowest and vice versa. This process will continue until all the stored energy is dissipated due to frictional losses. Another way to visualize a mechanical system storing and releasing energy is a weight suspended on a spring. The weight will oscillate up and down at resonance or its natural frequency if properly energized. As the spring compresses, the restoring force opposes that motion. As the spring stretches, the restoring force again tries to bring the weight back to the rest position. In this system, like the pendulum, the energy is constantly being exchanged between kinetic and potential. One tine of a tuning fork is really like a tiny pendulum. The mass of the tip is moving fastest as it passes the rest position while the internal spring force is greatest at maximum travel. Eventually the tuning forks tone will die down as the energy is radiated away in the form of sound and internal losses (friction). The reference to "*Q*" in electronics and acoustics generally refers to the quality factor or sharpness of the resonance.

In the electrical domain there are two basic components that store energy; inductors and capacitors. The inductor is the electrical counterpart to the spring force and the capacitor is the electrical counterpart to the moving mass. We call these parts reactive components.

In our loudspeaker in a sealed box, the mass and spring force play off each other and the system has a fundamental resonance. Just like the tuning fork, the resonance is determined by the spring and the mass. Electrically we can see in Fig. 21-19 that the inductance (the spring force) is in parallel with the capacitance or mass and it is driven through the resistance. In a loudspeaker this parallel resonance circuit is formed by the inductance (reflected spring force) and capacitance (the reflected moving mass). We can actually observe it by looking at the proportion of voltage and current at the input terminals. When we increase the stiffness of the total spring force (putting the driver in a sealed enclosure) the resonant frequency increases. Notice the curve changes its value depending on the frequency, Fig. 21-20A and 21-20B.

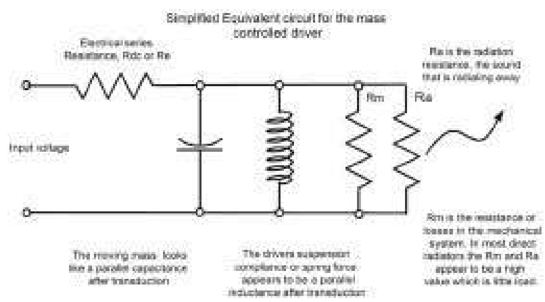
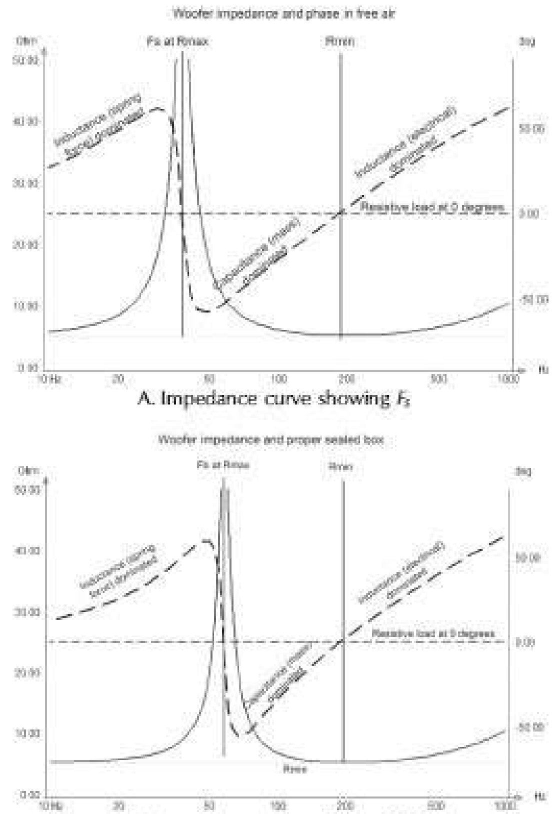


Figure 21-19. RC low-pass filter.



A. Impedance curve showing F_S

B. Impedance curve of a sealed box.

Figure 21-20. Impedance vs. frequency graph of a loudspeaker.

Remember when dealing with dc the load is called “resistance” or dc resistance. However with an ac or alternating current the resistive values are called impedance and the values change with frequency. When you measure a loudspeaker and plot the impedance as a function of frequency, the result is this characteristic impedance curve.

Recall that being acoustically small, the woofer’s radiator is operating on the sloped part of the radiation resistance curve. If the radiator were driven in such a way that the velocity remained constant, it would produce a rising frequency response. Remember, the velocity of the moving system is proportional to the voltage across the moving part of transduction mechanism. In our electrical equivalent circuit it is unavoidably in parallel with the moving mass, *C*, and the combined compliance or spring forces, *L*, while connected to the voltage source through the resistance of the same wire the coil is made with. The moving mass once above the resonance, is acting like a capacitor driven by a resistor, so a RC "low pass" filter is created which perfectly off sets the increasing radiation efficiency by rolling off the radiator velocity to an acceleration response. Therefore, as the frequency increases and one moves up the slope on the resistance curve, the low pass filter compensates and the driver exhibits a flat frequency response to a great degree.

When we move down in frequency, the radiator excursion must increase 4 times for each octave in order to maintain a flat response. If it were a constant velocity system, the velocity would double with each octave. As we approach the bottom of the loudspeakers' response, the spring force becomes more and more of a governing factor and the mass less so. At the frequency associated with the resonant peak in the impedance curve is reached, the spring and mass have properties that are equal but opposite. The peak value where the reactive parts cancel out is determined by the sum of various mechanical losses *and* by a tiny bit of radiation resistance. Since these losses are small but in parallel with the moving system, when reflected through the transduction process, the value for the impedance peak is high.

Continuing down below resonance, the system is totally governed by the spring forces, both the air in the box and the driver suspension spring force acting to return the radiator to the rest position. This extra force resisting the radiator motion as the frequency falls acts like a first order high pass filter, rolling off the low frequency velocity with decreasing frequency. As the spring force is tied to displacement and the transduction motor is driven as a current source through the resistance, its displacement becomes constant as the frequency falls.

When these two slopes are considered in light of the radiation resistance slope, the net result is a flat response above F_b or resonance, and a 12dB/oct roll off or a constant displacement response below resonance. In the sealed bunker, if the low frequency corner of the sealed box system mirrored the onset of the room's pressure mode region, then as the frequency is lowered, the constant displacement source compliments the displacement equals pressure response in the room and the response is flat to theoretically dc or 0Hz. Take a deep breath, that is perhaps the only "free lunch" in audio... and sadly doesn't exist outdoors or even to a strong degree in most living rooms but is a factor in car audio!

The shape of the low frequency corner is a major concern. Normally a large peak in the response is not desirable, and neither is a response which rolls off too sharply.

If the moving mass and spring forces are held constant, an important feature of resonance can be explored. Very often is it not so much that a system has a resonance but rather how sharp that resonance is. The degree of sharpness is referred to as Q. The higher the Q, the more energy is being stored in the moving system, much like a tuning fork. A system with a lf corner Q approximately 2 or more would be boomy and "one note" sounding, Fig. 21-21.

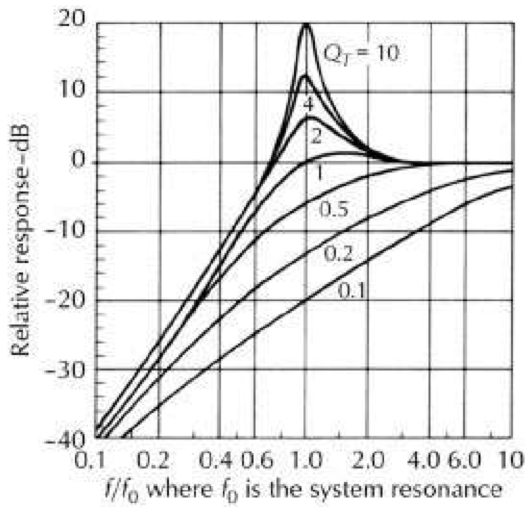


Figure 21-21. Response of a closed-box system vs. Q and normalized frequency relative to system resonance.

The thing that governs the shape of the loudspeakers response *Q* is primarily the result of the motor damping. Modern amplifiers when switched on but not given a signal, represent a very low impedance source. To the loudspeaker that source appears to be essentially a short circuit across the driver's terminals. To be clear, the amplifier also looks the same way when producing a signal except the output voltage is whatever the signal commands with a very small resistance in series.

A demonstration of damping can be seen if you have a woofer outside its enclosure and you get down close and gently tap on the cone. You can (usually) hear the resonance or rather you can hear the energy you imparted into the system ring and decay at the drivers "free air" resonant frequency. The *Q* of this resonance is fairly high and is determined by the driver's suspension losses and radiation resistance damping or dissipating the radiator velocity. With an alligator lead, wire, paper clip, short the loudspeakers terminals together and repeat the tapping process. Notice now that the sound is completely different, the quiet boom is gone and is now a dull thud with none of the decay of a tuning fork or plucked string.

The resonant frequency is the same, the mass is the same and the spring is the same, the difference is the amount of damping the transduction system adds.

Recall that the motor velocity produces a corresponding generator voltage, a back emf in the voice coil wire. With the terminals shorted, the wire's resistance is now across the motor or transduction mechanism. That generated voltage or back emf across that resistance produces a current flow. That current flow in the wire acting against the magnetic field produces a force in the voice coil wire. That induced force is in opposition to the moving system's velocity and so acts to damp motion. This is more fully explained by studying Lenz' law.

Thiele and Smalls' (T&S) standardized loudspeaker parameters were a significant contribution to the design of both enclosures and drivers. In this day and age, armed with T&S parameters and a freeware computer program, the proper box volume can be determined to obtain a desirable response shape.

While there are too many aspects to cover in depth here, if you tap on an un-shorted woofer, the resonant "boom" you hear is related to Q_m or the mechanical Q limited by the mechanical and acoustic losses. When you short the woofers' terminals, the resulting damped thud you hear is due to the Q_t or the combined effect of Q_e (electromagnetic damping) and the mechanical damping. When the driver is put in a smaller enclosure than outdoors or a room, it raises the resonant frequency because of that added spring force on the radiator. The two springs are in parallel, but raising the resonant frequency also raises the response shape Q in the box. The routine in effect is to put the driver in a box whose volume raises the F_b or resonance in the box to where the response shape Q is between 0.5 and 1, usually around 0.7. Another T&S parameter of particular interest is the V_{as} or compliance volume. This parameter specifies the driver suspension spring force on the radiator but as expressed as an equivalent volume of air. This is easily seen in Fig. 21-20 A and B.

At the high frequency end there is still more to the driver's simple equivalent circuit. This we can see in the previous figures more clearly when we add the response curve along with the impedance curves. In series with the input side *R*, is another smaller value inductance. With a real driver, this inductance may be somewhat dependent on the position of the voice coil in the gap and may not be a simple single inductance over a wide frequency but for now, consider it to be a fixed value. This inductance in series with the input eventually begins to roll off the current flowing onto the driver and electrically forms an *L-R* series low pass filter. The lowest point in the driver's impedance curve in the working band (called *R-minimum* or R_{min}) is the point where the series inductance and parallel capacitance form a series resonant circuit.

Both R_{min} and F_b are resonant conditions, one parallel one series. One has its lowest impedance at resonance, the other its highest impedance at resonance. Interestingly, at both resonances, the reactive parts are canceled out leaving a resistive load where the current is in phase with the input voltage, Fig. 21-22.

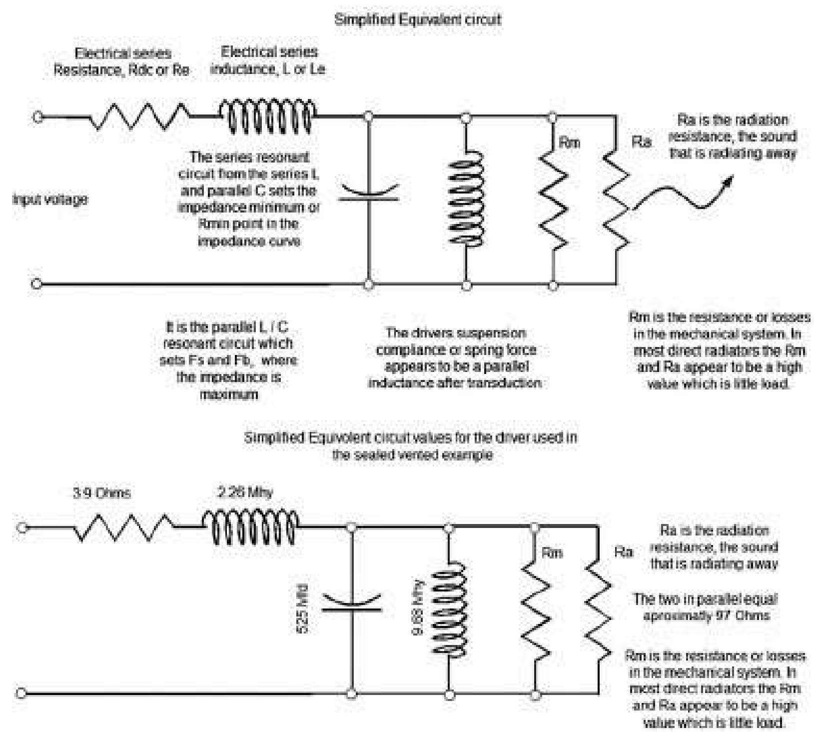


Figure 21-22. Expanded equivalent circuit now showing series L and two resonant paths, and the same with values.

As frequency is increased, the inductance increases the load impedance, reducing the current flowing and rolling off the driving force high frequency end.

Fig. 21-23 shows an imaginary long excursion woofer with the basic elements that were discussed previously, now called out. Note that many of the elements are involved in more than one action.

For example all of the moving parts contribute to the total moving mass, the suspension serves as a part of the radiating area but also part of the suspension spring and mechanical losses. Also, while it might not look like it, the displacement of the back side of the radiator is the same as the front side one is more familiar with.

That does not mean one reached the end of useful response however. Assuming the radiator continues to behave like a simple piston, as the radiator is larger than K=1 in the radiation curve, it begins to become directional. By increasingly focusing a falling amount of energy into a smaller angle, the two can offset each other, over part of the spectrum until some other issue impacts the system, like the radiator no longer being a simple piston or any number of other things, Fig. 21-24.

If the falling output from the series *L* is offset by the increasing directivity from the acoustically larger radiator, the flat response on axis can be extended until the radiation pattern becomes too narrow. So far, the load from the radiation resistance has been small enough to ignore, which reflects the very low efficiency of the typical direct radiator loudspeaker. That is not always the case, as we shall see with the horn loaded driver.

Efficiency

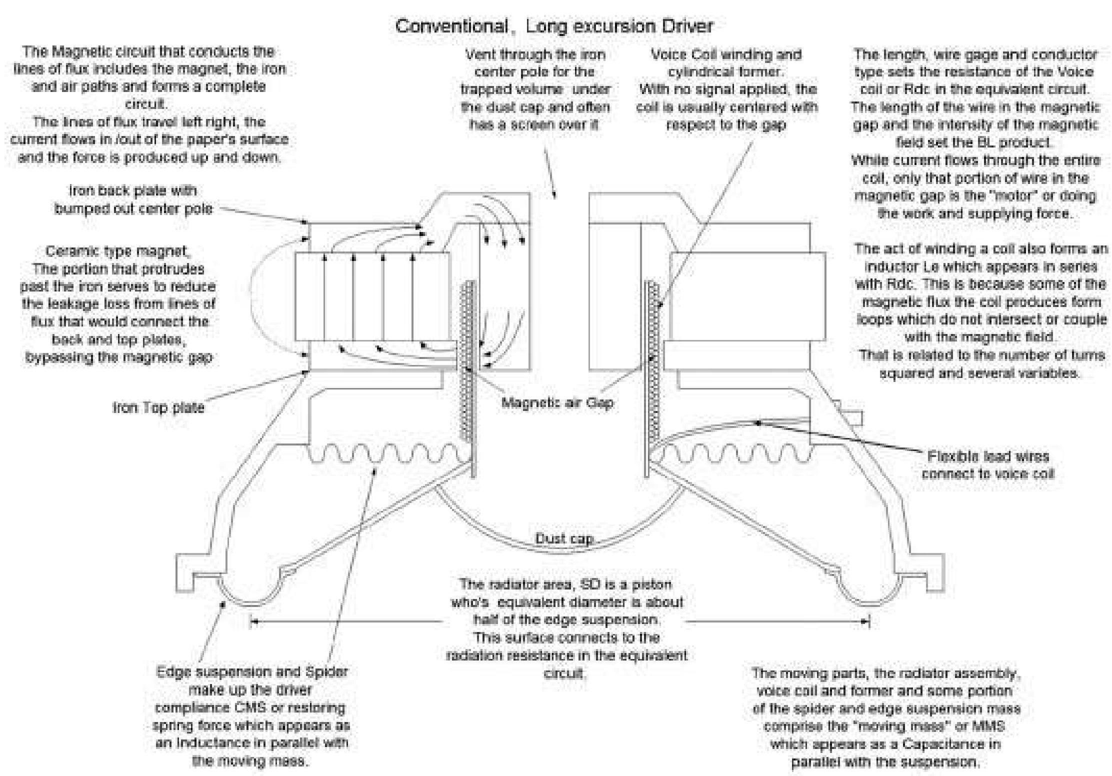


Figure 21-23. Parts of a typical long excursion woofer.

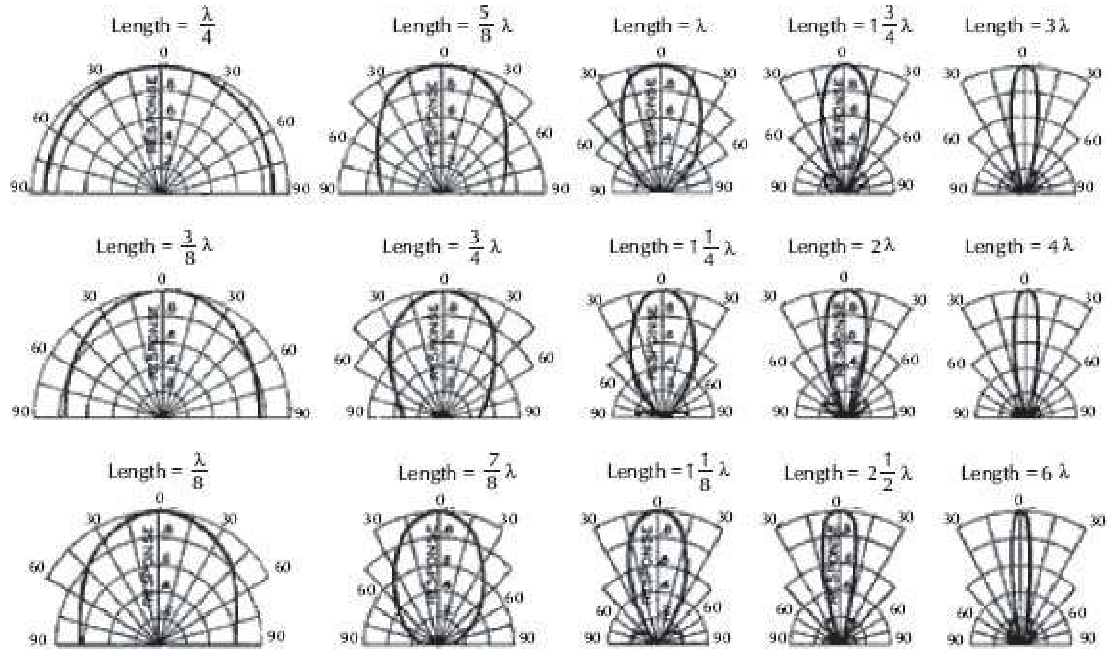
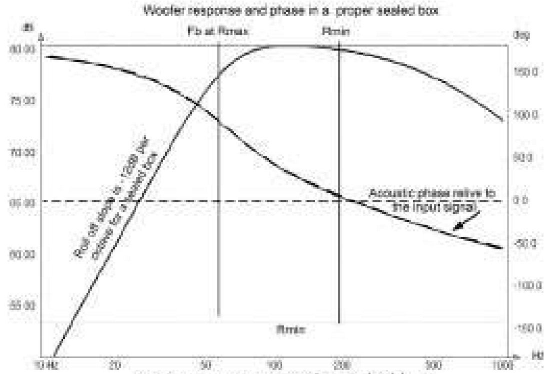


Figure 21-24. Radiation balloons for increasing radiator size.

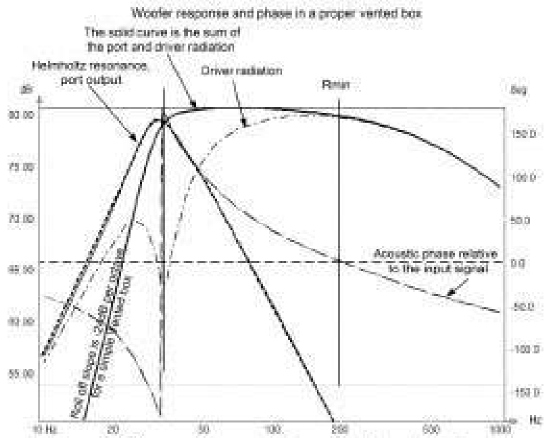
The nominal efficiency of a direct radiator is generally quite small often less than 1%. This means that if 100W is dissipated by a direct radiator, only 1W is dissipated as sound energy, the rest is essentially heat lost in the transduction process. Being strictly correct, the radiated sound is also eventually heat when it is fully absorbed in the atmosphere. As can be seen from the radiation resistance graph, the radiation efficiency is lowest at the lowest frequency of interest. For a sealed box, the low frequency corner in the frequency response is coincident with the box resonance and the peak in the impedance curve. The shape of the low frequency corner is also set by the amount of damping and from the graph of corner shape Q's, somewhere around 0.7 is usually considered ideal. What is not shown is that when this Q is set by the motor strength, then increasing the damping, lowering the corner Q shape also raises the efficiency well above the low corner. At the same time, as the motor strength is increased, the impedance curve becomes wider and higher in magnitude.

21.4.1 The Vented Box

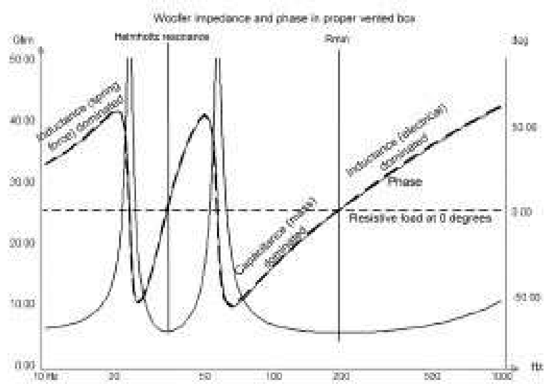
The vented box has improved efficiency compared to the sealed enclosure, for a number of reasons, but most importantly this type of enclosure permits the use of considerably more powerful motors and also by having as much or more vent displacement than the cone itself can produce. The vented system can (all other things being equal) radiate more sound for a given driver displacement with in an octave or so above the low corner frequency relative to a sealed box but curiously less than a sealed box somewhat below the low corner due to its steeper response roll off. The vent resonance system has a resonant shape defined by a *Q* or sharpness of the resonance. With the driver on the input side and the port radiation resistance on the output side, the *Q* is set by the box spring and the port mass. The Thiele Small formulas represent a formulaic way to specify drivers and design sealed and vented boxes where the result has the desired response shape. When the Thiele Small relationships are correctly chosen, one can have the increased sensitivity / efficiency of a stronger motor but instead of the earlier LF response roll off of a “too strong (for sealed box) motor” the vent resonance adds an additional acoustic load at the low corner and the result is flat response, Fig. 21-25.



A. Response curve of a sealed box.



B. Response curve of a vented box showing
showing driver, port, and total response.



C. Response curve of impedance and phase
in a proper vented box.

Figure 21-25. Response curve of a sealed and vented box.

The vented box can be thought of as a resonant phase inverter acting on the radiation from the rear of the driver. The resonant system in the vented box which acts as a phase inverter can be seen more clearly in an analogy. If you put a weight on the end of a spring or rubber band, by moving your hand up and down, you can find the “resonant” frequency where the motion of the weight is maximum and the motion of your hand, minimum. More importantly, if you observe carefully your hand will be moving in the opposite direction from the weight. If you move your hand just right (at resonance), you can get the weight to move much farther than your hand. In the vented box, the weight is the air in the port or passive radiator. The spring is the box volume and your hand is the back side of the driver’s radiator. At the low corner, the output from the port is much greater than that from the driver front radiation. The load from this resonance is normally placed centered on the sealed boxes impedance peak where the radiator has maximum mobility and the result is an impedance curve with two lower peaks. The load from that phase inverter/resonance is usually placed at or near the middle of that curve where the additional radiated power comes from. Consider a much larger load, or even an infinite load. If the voice coil were glued in place in the magnet, there would be no moving system resonance, no mobility, no motion and so no impedance curve peak. This can be seen more clearly going back and forth between the vented box impedance curve and the “mound” that is the port contribution in the prior curve.

Once the pressure created by the back side of the radiator is inverted, it can add to and augment the front radiation if only over a narrow band. There are two common ways to build a resonant phase inverter. The most common resonator consists of the “spring” compliance of the box volume and the acoustic mass coupled to it from the mass of air in the port. A less common resonator utilizes a passive radiator. The passive radiator is a driver of sorts but there is no electrical or transducer part to it. The passive radiator can substitute for a port by lumping the effective port mass and area into a freely movable piston with an equivalent value. The advantages to a passive radiator are considerable. They avoid port losses and noise at low frequencies, they do not have organ pipe length mode resonances and do not pass extraneous noise the driver may produce on the back side. The primary disadvantage is cost. For the passive radiator to be effective its own suspension resonance must be well more than an octave below the desired low response corner. They must have enough displacement as to not be the limiting factor and they shouldn't make noise of their own and so on. On the other hand, a vent is simple, cheap and will never break.

In either case as mentioned earlier, the vent system's resonant frequency should be set to near the center of the impedance curve. The impedance curve shows the moving systems ability to move in the velocity domain and the peak in the impedance curve is the place it is easiest to put an acoustic load on the driver. This is the point where the driver has the most mobility or the least constraint.

We can see this looking at the impedance curve of the vented system that a velocity filter was needed to achieve a flat response. In a sealed enclosure in a sealed enclosure going below the box resonance, the driver gradually becomes a constant displacement system. The constant force from the motor vs. frequency, pitted against a constant spring force vs. frequency. The combination of operating on the slope of the radiation resistance curve plus being on the slope below the mechanical resonance, the combined acoustic roll off is −12dB/oct. In a vented enclosure this second resonance increases the slope of the high pass filter to a fourth order or −24dB/oct. The reason for the added roll off is perhaps more easily visualized by following the frequency response curve down in frequency. As you go below the low corner, the vented system acts like a low pass filter. As the frequency falls, the "phase inverter" becomes less and effective as the frequency moves away from resonance. More of the radiators back radiation emerges from the port and progressively cancels out the front radiation. With the sealed box, the rear radiation is contained under all conditions and so the only fall off is related to the drivers constant spring controlled displacement. In the concrete bunker example, at low frequencies where displacement equals pressure, it would be flat to a very low frequency from "room gain" or pressure containment. If the box parameters are kept constant, but a vent is added to the system the response would not be flat below the corner frequency but would still be rolling off at 12dB/oct.

Band pass enclosures exchange bandwidth for gain over a narrower bandwidth using other forms of acoustic resonators. There are too many varieties to generalize, except to say, as one adds complexity and resonances, the alignments become more sensitive to driver variations, cabinet flexing and unexpected level dependent losses.

21.4.2 Transmission Lines

A transmission line is an older design which has benefitted greatly in more recent times from computer modeling and measurements. Transmission lines use the rear radiation by delaying it with a tunnel or duct, Fig. 21-26. The duct has significant resonances tied to its length and the design is dependent not only on the driver parameters but also internal damping within the transmission line to suppress the undesirable higher frequency modes.

21.4.2.1 Horns

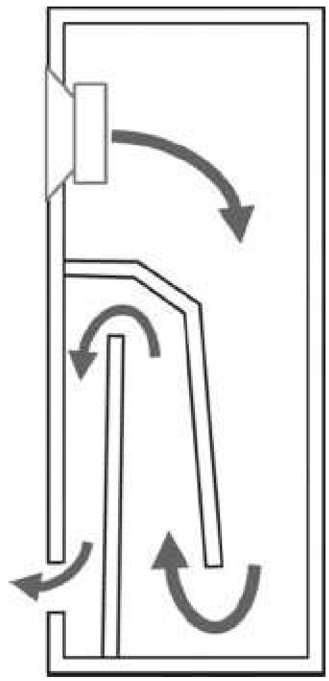


Figure 21-26. Transmission line.

Horn loudspeakers are more complicated and more subject to misunderstanding than the more common types and so some extra explanation is in order. The “horn” added to the loudspeaker is a ridged boundary which constrains the sound pressures radiation as opposed to radiating into free space in all directions. It could be argued that an acoustically small driver flush mounted on a flat baffle fits this descriptor and it does. This is a semantic distinction. We don’t usually refer to a flat baffle as a horn. When a driver is located in the corner boundary of a room we don’t usually use the term horn either. We talk about these drivers radiating into fractional space. Either way, horn or fractional space, the important thing is what happens to the sound pressure as it radiates away from the source. In a horn, the fraction of the space is much smaller than a driver in a room corner and so is considered differently. Confining the sound with a horn serves two main purposes, the first relates to the radiation resistance curve.

The horn is often referred to as a “transformer” which improves the load on the loudspeaker radiator and so improves its efficiency. This was of critical interest when producing 2W of power required say 50lbs of expensive electronics! This transformer effect increases the acoustic load, which appears as a resistor in parallel with the *L* and *C* formed by the moving mass and spring. This load appears as a lower resistance which conducts more current for a given voltage and so delivers more power to the acoustic load so more energy is radiated away. Note that for the horn, we no longer wish to roll off the radiator velocity or the voltage across the moving mass. The reason is seen in an old “thumb rule” which suggests that when radiating into “free space” (away from the ground or any boundary) the horn mouth circumference needs to be about 1 wavelength at the lowest frequency of concern. Unlike the low frequency direct radiator that spends its time operating on the lower, sloped part of the radiation resistance curve, the horn’s task is to couple the radiation resistance of the flat portion of the curve to the driver at the small end of the horn. The “thumb rule” relationship of one wavelength circumference at the lowest frequency assures that down to the low corner frequency and anywhere above, the horn mouth will be on the flat part of the radiation resistance curve, proving a constant radiation resistance vs. frequency.

The math that describes the relationships can be found elsewhere but what happens can be seen referring back to the impedance curve of the sealed box and the sealed box with a proper horn attached. In a perfect situation, some generalizations can be made about these curves, Fig. 21-27.

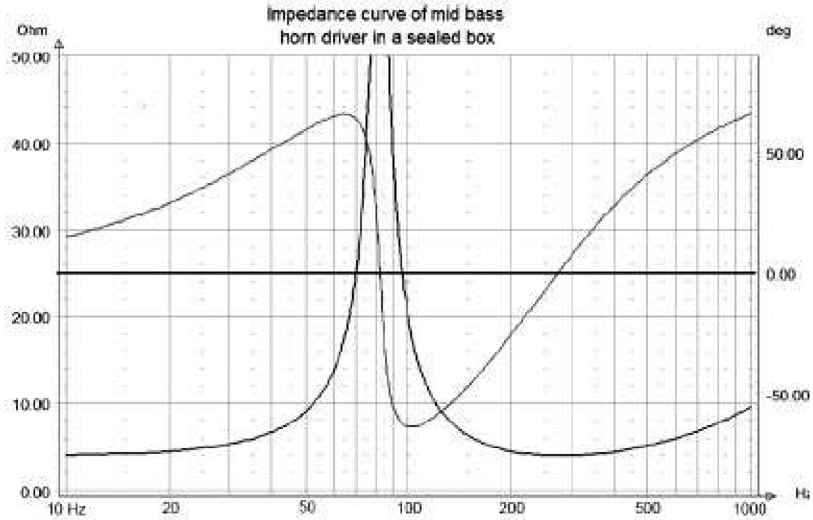


Figure 21-27. Sealed box impedance.

The impedance curve shows the moving systems ability to move in the velocity domain as a function of frequency. At the high frequency end, its ability to move is governed by the mass which can be seen as a parallel capacitance. At the low frequency end, the mobility is governed by the spring force, which can be seen as an inductor (*L*) parallel with the *C*. In the case of the direct radiator, the peak in the curve is where the effect of the spring and mass are equal but opposite and so cancel each other out, leaving the peak value to be limited by the mechanical losses and radiation resistance, both of which are rather small in the direct radiator. This explains why the magnitude of the peak is relatively high in direct radiators, Fig. 21-28.

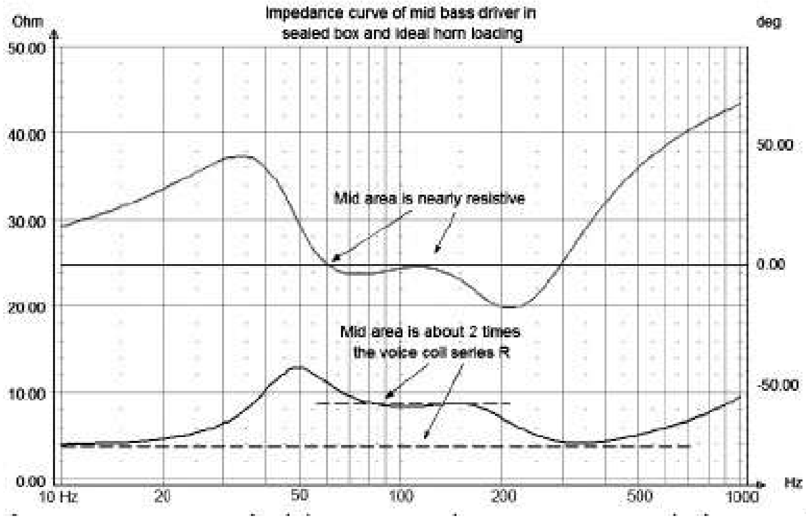


Figure 21-28. Sealed box impedance curve with lowered plateau.

If the acoustic load provided by the horn is increased so that it appears as a resistance equal to the driver's electrical resistance, the efficiency can be in the ball park of 50%, a big improvement over the 1% delivered by a direct radiator using the same driver! The bandwidth where this efficiency is possible, however, is limited. The limits are defined by the mass and spring effects as seen in the impedance curve but can be offset somewhat because in addition to resistance, the horn has some reactance. See section 21.4.2.2 *Reactance Annulling*. Horns are, to a greater degree, confined to a narrower bandwidth than direct radiators because the mass which gives the direct radiator its flat response, rolls off a horns response with increasing frequency. See Fig. 21-28.

There is a way around this, at least to a degree. As the motor strength increases, the width and height of the peak in the impedance curve increases as well and so too the efficiency bandwidth and upper frequency limit. If the moving mass can be reduced, leaving everything else alone this will also have a similar effect on bandwidth and high frequency response. Unfortunately, this is truly a complicated business! When you alter these parameters, you have changed the acoustics of the driver and so will need to alter other aspects to get the best results. Also, just as with the direct radiator, the series electrical inductance with the resistances acts as another “low pass” filter which rolls off the high frequency response.

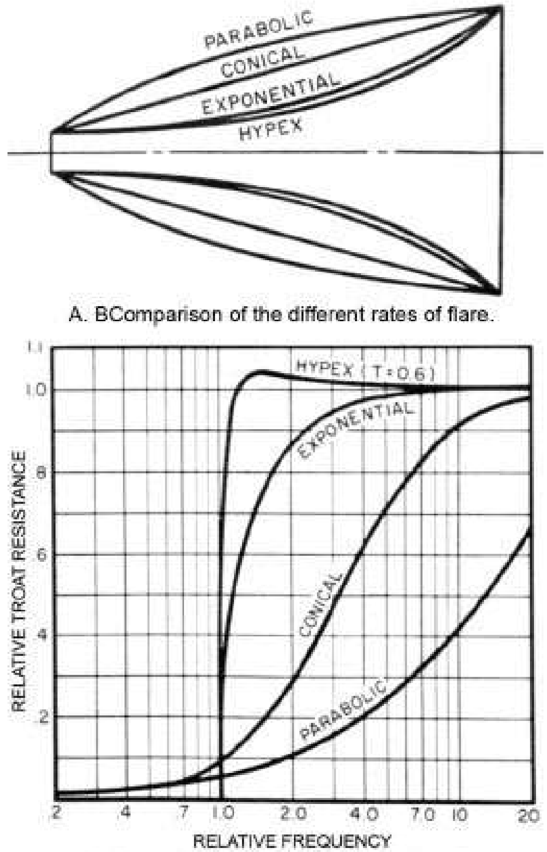
Compression drivers are most often seen coupled to horns because they typically combine the strongest motor budget and physics will allow, with very small, lightweight radiators attached. Even so, a typical 1 inch compression driver usually has a mass governed high frequency roll off beginning around 2–4kHz and an added inductive impedance roll beginning an octave or more higher. Some can still operate at frequencies much higher but the efficiency is much lower. They are called “compression drivers” because the area of the radiator is much larger than the throat end of the horn. The beginning of the horn, the acoustic throat, actually begins within the driver where the cross sectional area is smallest and starts beginning its outward expansion. The compression driver mounting system is simply a standardized way of interfacing the driver to the external horn. It is not uncommon for the radiator area to be 10 or more times larger than the actual horn throat. Inside the driver where the acoustic dimensions are very small compared to the highest frequency of concern, sound behaves much like a fluid. It passes through tiny spaces and around sharp bends within the compression driver’s phase plug. The front air volume between the radiator and throat acts as a spring and the air in the horn throat acts like a mass. If you get the relative sizes right, this spring/mass creates yet another “low pass” filter to the system. This acoustic filter if chosen properly, and/or the physical reality permits, can actually extend the high frequency response for a little bit, in exchange for a sharper ultimate roll off. This acoustic low pass filter also has the useful property of reducing the harmonic distortion any driver will produce as these added sounds are always multiples in frequency of the real signal and will be attenuated to a degree before entering the horn.

A side note; given the thumb rule for where the acoustic impedance transformation stops within the horn, this suggests that at 20kHz, the acoustic transformation would be well over and done, before the sound even reaches the exit of a one inch driver.

The compression ratio is another degree of freedom in the design, it is in effect using “Pascal’s Principal” a.k.a. hydraulic theory to accomplish an independent impedance transformation between the driver and the horn. When the compression ratio is higher than 1:1, a given radiator motion produces an air motion larger than that within the horn throat.

For those pursuing horn design, *On the Specification of Moving-Coil Drivers for Low-Frequency Horn-Loaded Loudspeakers* is an excellent starting point dealing with the design of such a system.

All horns are not created equal, Fig. 21-29. The shape of the horn governs how the radiation impedance at the mouth is felt or transferred to the throat end and its length and shape tied to its low frequency corner. The most common expansion profile, the exponential horn, exhibits a “high pass” effect to the transformation where the high pass corner frequency is set by the rate of expansion. For example, for a 30Hz horn expansion, the area has to double about every 24in (61cm) and no faster! For a 300Hz horn, the area has to double in 2.4in (6.1cm). You can’t have a small bass horn! There are other types of expansions available including arbitrary shapes and each will have a unique shape to the “high pass” corner.



A. BComparison of the different rates of flare.

B. Comparitive performance of the horn flares.

Figure 21-29. Radiation resistance for some various flares.

21.4.2.2 Reactance Annulling

At the low frequency end of the horn's response, there is an additional effect which can be used to extend the corner frequency downward a bit. The horn adds acoustic loading to the driver, but it also has reactance. The hyperbolic expansion profiles are expansion rates which initially expand more slowly. These shapes can present a significant acoustic reactance to the driver end. When reflected back through the transducer, the reactance appears as a mass or additional capacitance in parallel with the moving portion of the driver. However, unlike an electrical capacitance which has a fixed value, this C increases in value as the frequency falls and approaches the cutoff frequency. By choosing just the right amount of acoustic reactance and optimizing the enclosure and driver spring L and the combined mass of the driver and horn “C” the effect of the resonance can be extended over a wider range and can extend the range formerly limited by the spring compliance. There is more to it that is beyond the scope of this writing, but when everything is right this can be used to extend the low frequency response somewhat.

21.4.2.3 The Synergy Horn™

As the desire for constant directivity horns became more pronounced in commercial sound, the desire for a wide bandwidth source became more important as well. The horn adds acoustic loading to the driver but it also has resistance. Musical reproduction for an audience requires both a wide bandwidth and constant directivity. Furthermore within a room, the extraneous lobes that always accompany loudspeaker interference are not desirable as they often send sound energy to walls and ceilings where there are no listeners, effectively raising the background sound level. Voice intelligibility is predominantly governed by the ratio of direct to reflected sound. Loudspeakers which increase both the direct and reflected fields simultaneously cannot improve intelligibility. However a loudspeaker, which is a full range yet highly directional horn, can have a positive impact on intelligibility.

In the curve showing the radiation resistance for various horn flares, Fig. 21-29, it is clear that the conical shaped horns have a more gradual onset of the acoustic loading. As described in Don Keele’s paper *What’s So Sacred About Exponential Horns*, the up side is the more conical shaped horns have a more constant directivity than the more curved wall horns and the down side is inferior loading in the lower portion of the driver's response. The Synergy Horn™ and its predecessor the Unity Horn™ are approaches that recognized that the reason for inferior low frequency loading was that the rate of expansion near the apex was very rapid. This rapid expansion is a "high pass" for impedance transformation at a high frequency. Further down the horn towards the mouth the expansion rate gradually slows to that of a lower and lower frequency expansion. This design is typically a single large continuous horn passage driven with a high frequency driver at the apex but also driven at different frequency ranges, from different points along the sides of the horn. The locations for the mid and lower range drivers are governed by the expansion rate of the horn passage. Drivers are selected which are suited to horn loading in their respective frequency ranges. By placing all the drivers which cover a specific band less than 1/4 wavelength apart, they couple together forming one new radiation within the horn. The front to back spacing present in the physical layout, can be exploited and a passive crossover can be implemented. The drivers covering the lower frequencies couple to the horn body through a ports. The combination of the port size, length, shape and trapped air volume under the cone forms an acoustic low pass filter. Just like in the vented box example, the vent and box form a low pass filter. Here that low pass is not the bottom of the response but the top of it, with the working range below. This "low pass" acoustic filter is placed to be somewhat above the electrical low pass filter in the crossover network and the result is that the inevitable harmonic distortion is produced by the drivers which is not part of the actual signal, is attenuated by the acoustic low pass filter in front of the driver before entering the horn passage.

The Synergy crossover actually has little or none of the "all pass" phase shift that the "named" crossover slopes above 1st order all produce. The Synergy crossover is another design that couldn't have been arrived at without computer modeling because the shapes that are used are based on the systems measured behavior and are unlike the text book formulas or named, e.g., Butterworth, Bessel, Linkwitz-Riley, etc., ones. The ideal implementation of the Synergy Horn™, Fig. 21-30, will measure as if it only had one single crossover-less driver and produces a main lobe with no other lobes or nulls in the polar pattern and so radiates less energy outside the intended pattern.

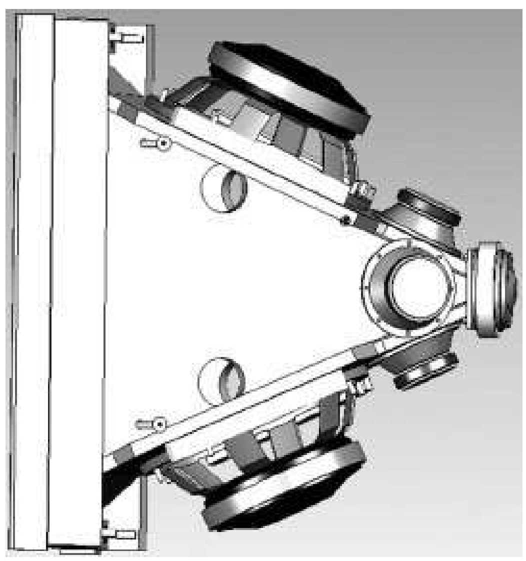


Figure 21-30. Side view Synergy Horn™.

21.4.2.4 "Compromised" Horns

Some horns are far enough away from the theoretical ideals so that at first glance it is difficult to see how they might work. And yet with careful measurements informing careful modeling, it is possible to cheat and get away with it! Bass horns are often "compromised", as they often do not have a large enough mouth to reach the knee in the radiation resistance curve. When the mouth is too small, the radiation resistance, which should be damping the duct resonances in the horn, is reduced. To the degree resonances are not damped at each end, the radiation resistance at one end the driver damping at the other, the Q of each can grow to an unacceptably large value. This subjectively limits “how small” one can make the bass horn. The horns length is also an issue here as the efficient bandwidth of the horn begins when it is about a half wavelength long. At a half wavelength, both the driver and mouth ends are at a velocity maximum. Reality forces most to use horns that are a quarter wavelength long at the low corner. That condition is very different for a driver as that location is a velocity minimum and pressure maximum. It can’t be emphasized enough that compromised horns, even more than normal horns are best designed using computer modeling combined with careful measurements. The more the horn design has departed from dimensions considered ideal, the more important it is to model such designs before making a pile of sawdust.

A very popular “compromised” bass horn is the “Scoop” or back loaded horn, Fig. 21-31.

Figure 21-31. Scoop horn.

This design originated long before computer modeling was available. Modern low frequency drivers have rendered the scoop somewhat obsolete, but it is an interesting study. The idea was to make the driver radiate directly from the upper portion of the enclosure while the rear of the driver connects to something between a horn and vented box, where the vent is the horn and radiates less and less above the low frequency corner.

The Tapped Horn™ is another small or compromised bass horn which only became possible and practical with the advent of computer modeling. Tapped Horns™ are typically much smaller than conventional bass horns with similar low frequency cutoff. These horns are built using some of the impedance matching relationships in bass horns but the configuration is more like a transmission line combined with a Klein bottle, Fig. 21-32.

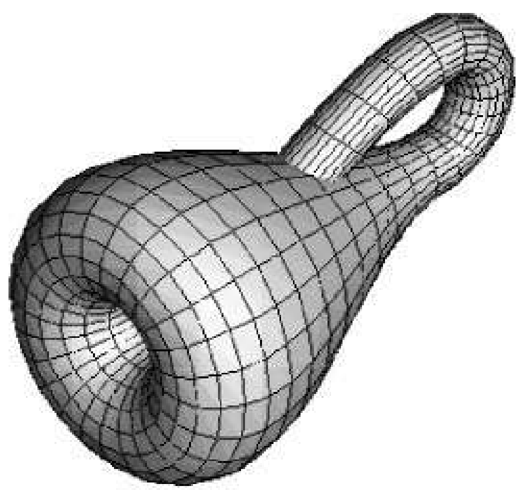


Figure 21-32. Klein bottle.

Here the driver is mounted at the interface between the small end of the horn and the side of the mouth end and unlike the Klein bottle picture, the end of the horn is open, Fig. 21-33.

Figure 21-33. Tapped Horn™ cutaway.

Both sides of the driver are now connected to the horn, one at the small end farthest from the mouth and the other near the mouth. While the two sides of the driver produce and equal but opposite displacement, this arrangement puts a time delay between the two radiations which imposes frequency dependent phase shift where the radiation from each side adds together in the horn. In the region where the first large dip would show up in the response of a regular small horn, both sides of the driver are additive, filling in that dip. Unlike a full size conventional horn, there is no back volume "spring" to tune to get reactance annulling. Some extension can be achieved by having a compatible driver "spring" or suspension compliance. When the various relationships and lengths are correct, the efficiency of a given driver can be increased from 6–9dB or more over the same driver used as a direct radiator with a usable bandwidth approaching two octaves. The roll off below the low corner is similar to a vented alignment. It is greater than −12dB/oct, while it can be less initially, ultimately it reaches the same 24dB/oct roll off as a vented alignment.

21.4.2.5 The Other Part of a Horn's Job

It can be seen by looking at the radiation resistance curve that increasing the size of the radiator beyond the "knee" causes no further increase in loading. Effectively, the point where the horn circumference is equal to the frequency produced is where the impedance transformation stops. As the frequency increases above the low cutoff and as the point where impedance transformation stops moves into the horn towards the throat, the large end of the horn still plays an important role as it will govern where the sound goes once it leaves the horn. It has been observed that at high frequencies many of the traditional types of horns exhibited a radiation pattern which narrowed with increasing frequency, Fig. 21-34.

To combat beaming, the multi-cell and other types of horns were developed in order to keep the radiation pattern a more constant shape over a wide frequency range. Don Keele, Jr. realized the cause of the narrowing and developed both a horn design strategy and a guideline which described the frequency at which the angle of the walls no longer govern the radiation angle, Eq. 21-1.

$$D = \frac{K}{a \times f_o} \tag{21-1}$$

where,

D is the horn dimension in inches.

K is the constant 10^6,

a is the horn wall angle,

f_o is the frequency where the pattern control is lost, By extension, this rule can also hint at why the horns of old became narrower in dispersion angle as the frequency increased. In the same way that the acoustic impedance transformation point retreats into the horn, moving towards the throat as the frequency increased, so does the point within the horn where the horn walls govern the radiation angle.

21.4.2.6 Pattern Flip

When discussing the phase plug used in compression drivers, it was mentioned that sound can act behave a fluid and flow around corners, if the dimensions are small compared to the wavelength. Conversely, in the discussion of pattern control in a horn, it is also clear that once the controlling dimensions are large compared to the wavelength, that the horn wall angle can govern the angle of the radiated sound and it can even behave more like light when the reflector is very large compared to the wavelength.

Don Keele, Jr.'s pattern loss thumb rule, Eq. 21-1, describes at what frequency a given horn angle and mouth width would lose pattern control. A little discussed or understood behavior was discovered as different kinds of horns were explored. This counter intuitive behavior is called "pattern flip". Pattern flip is most often observed in simple shaped horns which are not symmetric, for example a 20° by 60° horn or other combination where one plane has a significantly different horn angle than the other. As the name suggests, pattern flip infers that at some point, the narrow direction becomes wide and the wide direction becomes narrow.

Starting with high frequencies, the horn radiates just as its internal structure governs. However as the frequency is lowered, according to the pattern loss relationship, the horn operation will reach a frequency where it loses control. When the horn angles are very different, the horn will lose control in one plane significantly before the other. When one plane has lost pattern control relative to the other, the radiation pattern will be compressed into the plane where pattern has been lost.

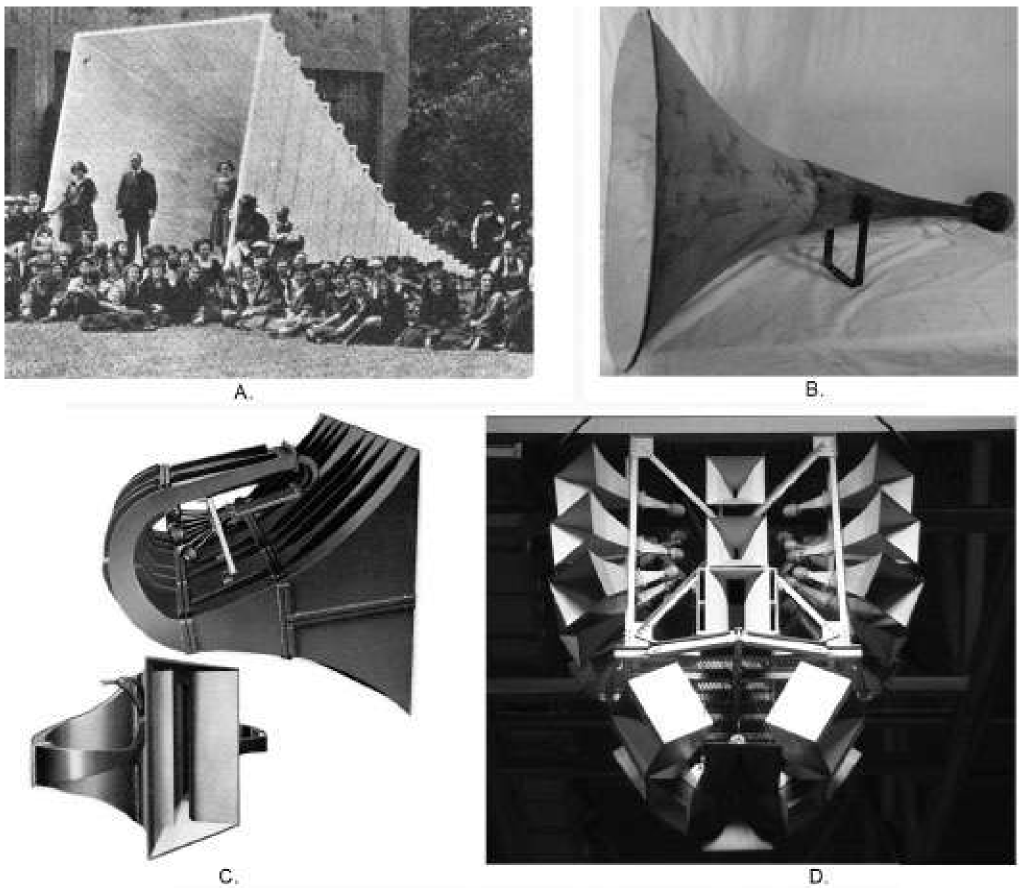


Figure 21-34. Various one-cell horns.

The historical Electrovoice TW-35/T350, Fig. 21-35, is a great example of pattern flip being used to advantage. Mounted with the long axis vertical the T350 lost pattern control in the horizontal at just below 12kHz. It exhibited a very wide horizontal coverage due to the fact that the horn had no control over the horizontal at all. The T350 is a fascinating study as many used this driver in larger systems and mounted it with the long axis horizontal because it looks like it should be mounted that way.

The root problem with simple shape horns is that once the horn wall has become very asymmetric, the aspect or mouth conditions conflict with the shape and cause pattern flip. So to build a 60° H by 20° V horn that avoids pattern flip, the mouth height needs to be about 3 times the width. This is exactly the opposite of what is usually done with simple horn shapes which are wide and short.

The exploration of this area brought about another generation of horn shapes all designed to provide constant directivity or in other words ideally the same frequency response across a listening plane.

Sometimes these old horns find their way into the hi-fi area where the directivity and dynamic capacity can make a dramatic difference as well as an impressive visual statement, Fig. 21-36.

21.4.2.7 Output Limitations of Loudspeakers

The maximum usable output of an electromagnetic loudspeaker is a function of a number of parameters, including diaphragm displacement, heat transfer, sound quality, i.e., maximum acceptable nonlinearity, and/or wear life due to fatigue of moving parts.

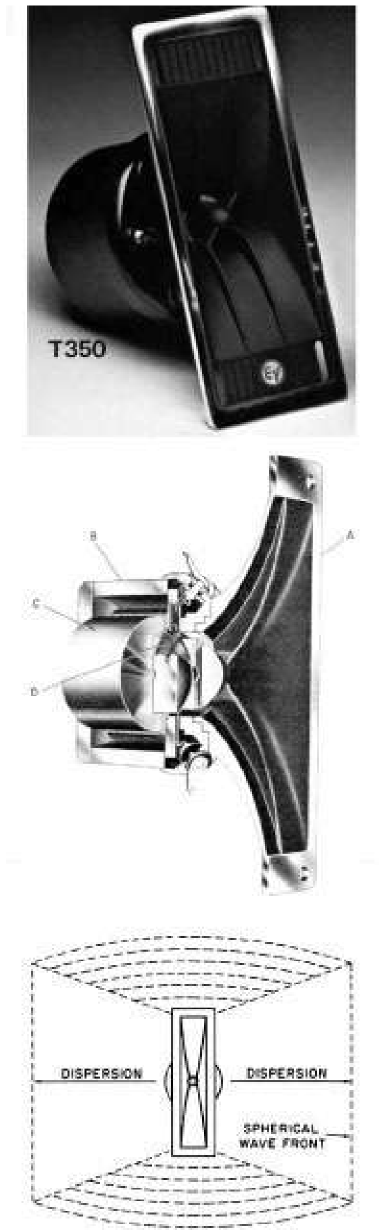


Figure 21-35. Electrovoice TW-35/T350.

There are two fundamental limitations on a magnetic driver, a displacement limit and a thermal limit. Displacement limits may be caused by mechanical or electrical factors. Mechanical displacement limiting occurs when a moving part contacts a stationary one or when a suspension element is deformed beyond its design range. Electrical displacement limiting occurs when the motor is operated outside its range of linear travel. This is a function of the length of the windings on the voice coil and the thickness of the plates that form the magnet gap. Fig. 21-37 shows three typical voice coil configurations: equal length, overhung, and under-hung coils. When any of these coils reaches a displacement that causes a reduction in the current sensitivity of the motor, higher distortion will result. Linear transduction requires the force per current stay essentially constant with position.

It has been empirically determined that an excursion of 15% farther than the gap length results in a reasonable distortion level of approximately 3% harmonic distortion at low frequencies. Each one of the three voice coil types has its own strengths and weakness. The equal length voice coil, Fig. 21-37C, has the greatest potential for motor-generated distortion. However, it also yields the highest motor strength because it has the greatest total conductive mass in the highest density magnetic field. The equal length voice coil is a common configuration for compression drivers, where maximum excursion is intrinsically low. The underhung coil, Fig. 21-37B, allows greater excursion but requires a larger magnet due to the longer gap. For moderate flux density levels (10,000–15,000G), this design, as compared to the equal length design, requires approximately twice the magnet weight (twice the area and the same length) for a doubling of the gap length. This approximately doubles the excursion capacity, giving four times the acoustic power output capability (6dB) for a doubling of magnetic weight (3dB). The overhung coil, Fig. 21-37A, is capable of the greatest motor linearity, all else being equal. It is commonly seen on woofers used as direct radiators, where higher excursion is required. The major disadvantage here is that the coil that is not in the gap does not participate in transduction. The extra coil length does add both mass and dc resistance, however, reducing motor efficiency. In spite of this, there are numerous examples of successful commercial woofers using overhung coils.

Figure 21-36. Image courtesy John Kalinowski, Horn Loudspeaker Forum.

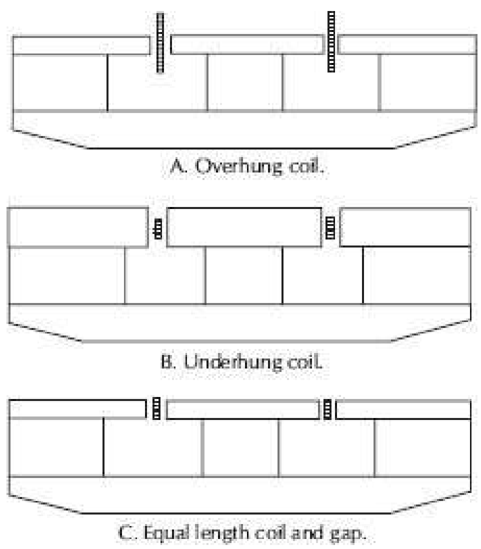


Figure 21-37. Typical voice coil configurations.

The thermal limit of a magnetic loudspeaker motor is a function of the temperature limits of the materials used and heat transfer from the coil assembly to the outside world. Most adhesives used in the loudspeaker industry have an upper limit between 120°C and 177°C (248°F and 350°F). Some adhesives will tolerate higher temperatures, exceeding 354°C (670°F), but they can require special curing processes and are therefore potentially more difficult to use. Anodized aluminum wire has the melting point of aluminum as a limit. Voice coils operated at high temperatures have higher resistance. A 1°C (1.8°F) rise produces approximately a 0.4% rise in dc resistance in both copper and aluminum. Therefore, operating a voice coil 100°C (212°F) above ambient (127°C or 292°F) will cause the voice coil resistance to increase to 40% above its ambient value, and has approximately doubled by 230°C (446°F).

The operating temperature of the voice coil, T_{VC}, is determined by ambient temperature, the amount of power being dissipated in the coil, and a parameter called *thermal resistance*, expressed in degrees Celsius per watt, °C/W. The thermal resistance is a measure of the ability of an object to transfer heat away from itself. The lower the value of the thermal resistance, the more effective the object is at this transfer. As power is doubled, final temperature rise above ambient is doubled. Heat transfer in a loudspeaker is a function of the air gap design, voice coil design, and the ability of the loudspeaker frame and magnet to dissipate heat to the surrounding or ambient air.

As the air gap length is decreased and the area increased, heat transfer increases (or, equivalently, thermal resistance decreases). Making the voice coil former of aluminum will increase effective heat transfer area; the thicker the aluminum, the greater the effect. Voice coils wound on aluminum formers with large diameters in magnets with large gap areas and very tight coil to gap tolerances are capable of handling high electrical power due to good heat transfer in the air gap. In short, large, accurately constructed loudspeakers can usually handle more power. As the loudspeaker moves, it may be able to pump the air in the gap to improve heat. The loudspeaker designer may be able to exploit this behavior. Given voice coils of the same length, the underhung and equal-length configurations will have greater heat transfer capacity. The overhung coil would only conduct heat well in the gap region, while the coil ends remaining out of the gap would be more likely to suffer damage at high power level because of relatively poor heat transfer.

Temperature rise in voice coils is not instantaneous. It is directly related to mass as well as time and program material. As one might suspect, light voice coils have short thermal rise times, and vice versa. The time constant of the magnetic structure and frame can be on the order of hours. For this reason, long duration power tests are required to evaluate the maximum power tolerance of transducers. Heat transfer from the frame and magnet to the air is another important consideration. Although the rise time is long, the final temperature may vary greatly due to the enclosure. A vented enclosure with vents at the top and bottom with no fiberglass insulation might provide adequate ventilation for a hot loudspeaker. The same loudspeaker in a closed box stuffed with fiberglass might be subject to a dangerously high temperature rise. Attention to the final thermal path is warranted in applications that will demand maximum output from enclosed loudspeakers.

The efficiency of a loudspeaker has a direct bearing on the thermal load it must withstand for a given acoustic output level. The more efficient the loudspeaker, the lower the self-heating for a given output level; all else being equal, a loudspeaker with 3dB higher overall sensitivity for a given impedance will experience one-half the thermal load for a desired output level.

In concert touring use, loudspeakers are routinely operated at and even beyond their design limits. If the resistance of the voice coil doubles as a result of heating the driver will be 6dB less sensitive For this reason, sound quality can vary greatly over the course of a performance. In failure situations, the nature of the input signal will usually determine the type of failure mode. Thermal failure can be precipitated by compressed high-frequency content material (low dynamic range). Mechanical failure is often due to dynamic, percussive material, such as might occur in a recording studio with drum channels set to solo, as well as other signals that do not limit dynamic range. Another cause of mechanical failure, most often in high-frequency transducers, is the application of a highly clipped signal that has been passed through a high-pass filter. Such a signal will contain a peak-to-peak voltage that is twice that of the input signal.

21.4.3 Heat Transfer Designs for High-Power Woofers

Of all the components in a sound reinforcement system, more heat is generated in low-frequency devices than in any other. These devices are typically only 2–8% efficient. The remaining 92–93% of the power goes directly into producing heat. For some forms of music this is not a problem. However some forms of music most notably EDM or electronic dance music contain very high levels of extremely low frequencies and these signals last for extended periods. Heat dissipation becomes a critical concern. Adding to the problem is the fact that much modern program material is bass-heavy.

As understanding of heat transfer mechanisms in loudspeakers grew, designs appeared that improved heat transfer from the voice coil and gave improved thermal power handling ratings.

21.4.4 Line Arrays

Another type of loudspeaker system is a line array. Although line arrays have much in common with other types of loudspeaker systems, they have some attributes that are unique enough to justify their separate treatment. A line array may form a complete full-range loudspeaker or one or more bands thereof. In a line array, individual radiators are arranged in a straight line or an arc segment. It is also possible for a number of complete loudspeaker systems to be configured as a line array. It is this configuration that has come into fashion in recent years. In the simplest form of line array, each of the elements—usually a small cone transducer—is supplied an identical full-range signal. This type of array, also called a *sound column*, was popular in this country through much of the 1970s and is still in common use in installed sound systems.

Recent developments in DSP technology, combined with the constant pressure on the touring concert reinforcement industry to minimize weight, blockage of audience sight lines by speakers, and truck space, have resulted in a resurgence of interest in line arrays. As attractive as some of their perceived performance characteristics may be, they have inherent limitations. First, the directivity attributes associated with line arrays are present in the vertical plane (along the length of the array) only. The horizontal directivity is only as good as the horizontal performance of the individual devices used to form the array. Secondly, line arrays invariably comprise discrete elements, as opposed to a continuous line source.

This periodicity of an array of individual sources exacerbates problems with nulls and lobes, and it causes the off-axis impulse response of a line array to contain multiple discrete arrivals. That is to say when driven with a signal like a single impulse (a click or something like a recording of a firecracker), what arrives is train of arrivals beginning with the one from the closest source and ending with the farthest source.

It is often incorrectly asserted that a line array behaves, or can behave, as a line source. A line source is largely a theoretical construct. It consists of a infinitely long, narrow radiator that radiates sound with perfect uniformity at every point on its surface. In practice "acoustic infinity" is often said to be a size of around 40–50 wavelengths. This assumption of perfect uniformity, while impossible to achieve in practice, simplifies the mathematics required to model the behavior of a line source. When used for illustrative purposes in texts, line sources may additionally be assumed to have infinite length, making possible even further simplification of the mathematical model. The same model has been employed in texts on electromagnetic theory, for the same reasons.

The two assumptions—continuous radiation and infinite length—lead to two interesting results. First, due to symmetry, the frequency response of an infinitely long, continuous line source is not a function of observation position along the line. For example, if the line is assumed to be coincident with the Z-axis in a cylindrical polar coordinate system, its response will not vary with changes in the Z-coordinate of an observation position (i.e., for movement in a direction that is parallel to the line). Second, due to the infinite length of the source, the wavefront (a collection of isophase points) will form a cylindrical, rather than a spherical, shape. For this reason, the intensity of radiation in the outward direction falls off as the inverse of the first, rather than the second, power of the distance from the line, i.e., 3dB for a doubling of distance.

As interesting and attractive as the two above results may be, they are not achievable in any physically realizable array. The effects of radiation that is neither continuous nor uniform, and of finite array length, cannot be neglected in discussing the behavior of real-world systems. Unfortunately, these issues have been glossed over or completely ignored in the information that is provided regarding the performance of commercially available line array products.

Full-range line arrays characteristically have relatively narrow vertical radiation patterns. The details of these radiation patterns vary widely with frequency and typically contain undesirable off-axis nulls (deep response notches) and lobes (response peaks). The same phenomena that produce off-axis response variations in a non-coaxial, multi-way loudspeaker—interference caused by variations in the relative distances between multiple sources and the listener—create this directivity. At high frequencies, the angular separation between the first two nulls—and therefore the useful coverage angle—may be on the order of 5° or less.

A number of remedies to the problem associated with line arrays have been implemented over the past 50 years. There are two primary areas in which the line array intrinsically poses challenges to the designer: total array length and individual device spacing. Both must be addressed in order to produce a well-behaved system. Interestingly, both of these make the line source behave more like a point source.

One means to address the issue of total array length is to implement a tapered array. In this type of array, only the innermost elements carry the highest frequencies. The signals applied to the more outwardly placed elements in the array are low-pass filtered at successively lower frequencies. The goal of this approach is to make the effective length of the line array become shorter at higher frequencies. An alternative way of stating this goal is that the ratio between the effective length of the array and the wavelength of sound should ideally be invariant. With the ability via DSP processing to create filters of essentially arbitrary amplitude and phase response, it has become relatively straightforward to create tapered arrays. Additionally, the availability of frequency-independent delay makes lobe steering possible. In all of these cases, the directivity of each individual source is what limits the total directivity as the sources all still radiate independently, Fig. 21-39.

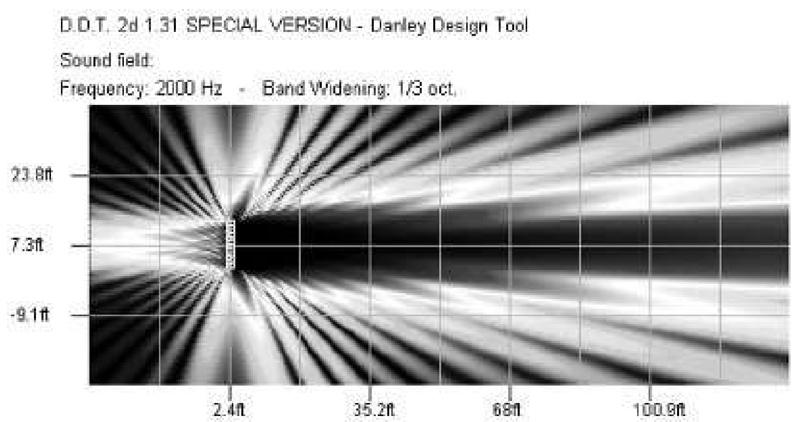


Figure 21-38. Twelve box line array at 2kHz.

The matter of device spacing poses another set of challenges. The smaller the spacing can be made relative to wavelength, the better a line array or point source array can approximate the behavior of a continuous radiator. When device spacing becomes large relative to a wavelength—roughly in the range of a 1/2 wavelength or more—the off-axis response of the array will contain many lobes and nulls. It is likely that one or more of these off-axis lobes will approach the level of the on-axis radiation. Consider Fig. 21-14 which are the polar plots from two sources which begin less than 1/4 wavelength apart and end several wavelengths apart. When one considers the small wavelengths of the higher audible frequencies—the wavelength of 10kHz is 34.4mm (1.35in)—the challenge of achieving optimal device spacing for higher frequencies becomes apparent. The continued reduction in size of motor assemblies through the use of high-powered magnetic materials has been helpful in addressing this issue, as has the development of combiners which allow many high frequency drivers to be coupled to the same horn, Fig. 21-40. Fig. 21-41 shows the vertical coverage of a Danley J4 loudspeaker.

Figure 21-39. Danley Sound Labs J4 astigmatic point source horn array with 64 compression drivers on one horn.

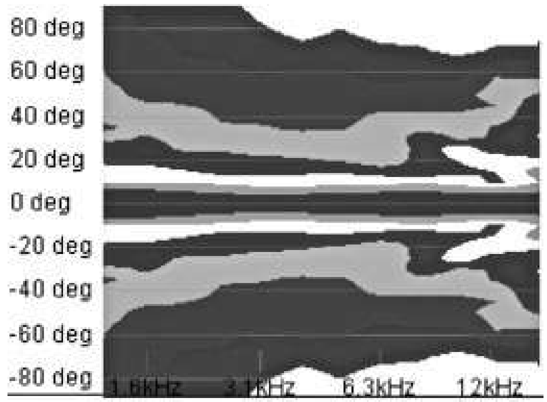


Figure 21-40. Danley L4 vertical coverage contour.

Bibliography

A. A. Janszen, "An Electrostatic Loudspeaker Development," *J. Audio Eng. Soc.*, Vol. 3, no. 2, April 1955.

A. H. Benade, *Fundamentals of Musical Acoustics* Oxford, University Press, 1976.

A. N. Theile, "Loudspeakers in Vented Boxes: Part I & II," *J. Audio Eng. Soc.*, Vol. 19, no. 5, and 6, May and June 1971.

A. Wood, *Acoustics*, New York: Dover Publications, Inc., 1966.

Allan D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, New York: McGraw-Hill, 1981.

B. N. Locanthi, "Application of Electric Circuit Analogies to Loudspeaker Design Problems," *J. Audio Eng. Soc.*, Vol. 19, no. 9, October 1971.

C. A. Henricksen, "Heat Transfer Mechanisms in Loudspeakers: Analysis Measurement and Design," *J. Audio Eng. Soc.*, Vol. 35, no. 10, October 1987.

C. A. Henricksen, "Phase Plug Modeling and Analysis: Radial versus Circumferential Types," presented at the 59th Convention of the AES, preprint 1328 (F-5).

C. R. Hanna and J. Slepian, "The Functions and Design of Horns for Loudspeakers," *J. Audio Eng. Soc.*, Vol. 25, no. 9, pp. 573–585, September 1977.

D. B. Keele Jr., "Low-Frequency Loudspeaker Assessment by Nearfield Sound-Pressure Measurements," *J. Audio Eng. Soc.*, Vol. 22, no. 3, pp. 154–162, April 1974.

D. B. Keele Jr., "What's So Sacred about Exponential Horns?" AES preprint 1038 (F-3).

D. B. Keele Jr., "Low-Frequency Horn Design Using Thiele/Small Driver Parameters," presented at the 57th convention of the AES, preprint no. 1250 (K-7), May 1977.

D. B. Keele Jr., "Effective Performance of Bessel Arrays," AES preprint 2846 (H1)

presented at the 87th Convention of the AES, October 1989.
G. L. Augsburger, "Electrical versus Acoustical Parameters in the Design of Loudspeaker Crossover Networks, *Journal of the Audio Engineering Society*, Vol. 19, no. 6, June 1971.
H. F. Olson, *Acoustical Engineering* Princeton: Van Nostrand, 1976.
H. F. Olson, *Music*, Physics and Engineering, New York: Dover Publications.
H. Suzuki and J. Tichy, "Diffraction of Sound by a Convex or a Concave Dome in an Infinite Baffle," *Journal of the Acoustical Society of America*, preprint 0 (5), November 1981.
J. L. Bernstein, *Audio Systems*, New York: John Wiley & Sons, 1966.
J. M. Kates, "Radiation from a Dome," *J. Audio Eng. Soc.*, Vol. 24, no. 9, November 1976.
J. R. Ashley and A. L. Kaminsky, "Active and Passive Filters as Loudspeaker Crossover Networks," *J. Audio Eng. Soc.*, Vol. 19, no. 6, June 1971.
J. R. Ashley and M. D. Swam, "Experimental Determination of Low-Frequency Loudspeaker Parameters," *J. Audio Eng. Soc.*, Vol. 17, no. 5, October 1969.
J. R. Wright, "Fundamentals of Diffraction," *J. Audio Eng. Soc.*, May 1997.
John Borwick (Editor), *Loudspeaker and Headphone Handbook* London: Butterworths, 1988.
John Vanderkooy, "A Simple Theory of Cabinet Edge Diffraction," *J. Audio Eng. Soc.*, December 1991.
K. O. Johnson, "Single-Ended Wide-Range Electrostatic Tweeters with High Efficiency and Improved Dynamic Range," *J. Audio Eng. Soc.*, July 1964.
L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, New York: John Wiley & Sons, 1962.
L. L. Beranek, *Acoustics*, New York: McGraw-Hill, 1954.
N. W. McLachlan, *Loudspeakers; Theory, Performances, Testing, and Design*, Corrected Edition, New York: Dover Publications, 1960.
P. M. Morse, *Vibration and Sound*, New York: McGraw-Hill, 1948.
P. W. Klipsch, "A Low-Frequency Horn of Small Dimensions," *Journal of the Acoustical Society of America*, October 1941.
P. W. Klipsch, "Modulation Distortion in Loudspeakers," *J. Audio Eng. Soc.*, Vol. 17, no. 2, April 1969. (See Parts II and III, *J. Audio Eng. Soc.*, Vol. 18, no. 1, February 1970 and Vol. 20, no. 10, December 1972.)
P. W. Klipsch, "Modulation Distortion in Loudspeakers Part II," *J. Audio Eng. Soc.*, Vol. 18, no. 1, February 1970.
R. F. Allison, "The Influence of Room Boundaries on Loudspeaker Power Output," *J. Audio Eng. Soc.*, Vol. 22, no. 5, June 1974.
R. G. Brown, *Lines Waves and Antennas*, New York: The Ronald Press Co., 1961.
R. H. Small, "Constant-Voltage Crossover Network Design," *Journal of the Audio Engineering Society*, Vol. 19, no. 1, January 1971.

R. H. Small, "Direct-Radiator Loudspeaker System Analysis," *J. Audio Eng. Soc.*, Vol. 20, no. 5, pp. 383–395, June 1972.

R. H. Small, "Closed Box Loudspeaker Systems; Part I: Analysis," *J. Audio Eng. Soc.*, Vol. 20, no. 10, pp. 798–808, December 1972.

R. H. Small, "Vented Box Loudspeaker Systems; Part I II III and IV," *J. Audio Eng. Soc.*, Vol. 15, nos. 5, 6, 7, and 8; June, July, August, and September 1973.

R. H. Small, "Phase and Delay Distortion in Multiple-Driver Loudspeaker Systems," *J. Audio Eng. Soc.*, Vol. 19, no. 1, January 1971.

R. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion Part 1," *J. Audio Eng. Soc.*, Vol. 17, no. 1, p. 30, January 1969.

R. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion Part 2," *J. Audio Eng. Soc.*, Vol. 17, no. 2, p. 130, April 1969.

S. H. Linkwitz, "Active Crossover Networks for Non-Coincident Drivers," *J. Audio Eng. Soc.*, Vol. 24, no. 1, p. 2, January/February 1976.

S. Ishii and K. Takahashi, "Design of Linear Phase Multi-Way Loudspeaker System," AES 52nd Convention, October 1975, preprint 1059.

Chapter 22

Loudspeaker Cluster Design

by Ralph Heinz

22.1 Why Array?

For the purposes of this discussion we can define a loudspeaker array *as a group of two or more full-range loudspeaker systems, arranged so their enclosures are in contact*. System designers use arrays of multiple enclosures when a single enclosure cannot produce adequate sound pressure levels, when a single enclosure cannot cover the entire listening area, or both. These problems can also be dealt with by distributing single loudspeaker systems around the listening area, but most designers prefer to use arrays whenever possible because it is easier to maintain intelligibility using a sound source that approximates a point source than by using many widely separated sources.

22.2 Array Problems and Partial Solutions: A Condensed History

First-generation portable sound systems designed for music used a very primitive form of array: they simply piled up lots of rectangular full range speaker systems together, with all sources aimed in the same direction, in order to produce the desired *SPL*. This type of array produced substantial interference, because each listener heard the output of several speakers, each at a different distance. The difference in arrival times produced peaks and nulls in the acoustic pressure wave at each location, and these reinforcements and cancellations varied in frequency depending on the distances involved. So although the system produced the desired SPL, the frequency response was very inconsistent across the coverage area. Even where adequate high frequency energy was available, intelligibility was compromised by multiple arrivals at each listening location.

Second-generation systems incorporated compression drivers and horn-loading techniques derived from cinema sound reinforcement and used for large-scale speech-only systems (the original meaning of public address). When two or three of these horns were incorporated in a single enclosure with trapezoidal sides that splayed the horns away from each other, the first arrayable systems were introduced to the marketplace. These products promised to eliminate lobing and dead spots (peaks and nulls) and to drastically reduce comb filtering (interference). They did improve performance over the stack of rectangular enclosures loaded mainly with direct radiating cones. But frequency response across the coverage area remained inconsistent. In addition to the midrange and high frequency variations across the coverage area of the array, low frequency output varied from the front to the rear and side to side. Low frequency energy was focused along the longitudinal axis of the array and close to it, producing a “power alley” that gave the seats with the best views the worst sound, Fig. 22-1.

Figure 22-1. A typical second generation loudspeaker cluster. Even when a single enclosure is designed to resemble a point source, multiple enclosures will always interfere with each other when connected to a coherent audio signal.

22.3 Conventional Array Shortcomings

As we said in the first paragraph, the performance advantages of the array (whether horizontal or vertical) derive from its ability to approximate a perfect acoustical point source. But even the smallest arrays typically including three or more loudspeaker enclosures, each with two or three separate acoustic centers of its own. It's easy to appreciate that getting all those discrete sources to behave like a theoretical point source is difficult in practice. Signal processing solutions attempt to compensate for the difference between theory and reality by sacrificing the coherency of the electronic signal. They apply frequency shading and/or micro-delays to the signals sent to different enclosures, in order to ameliorate the acoustic problems. These approaches are costly, complicated and often meet with limited success.

A rigorous analysis of the acoustical physics can point the way toward a practical, physical solution. First, consider what is probably the most common arrayable system in use today: 60° × 40° horns in enclosures with 15° trapezoidal sides, Fig. 22-2.

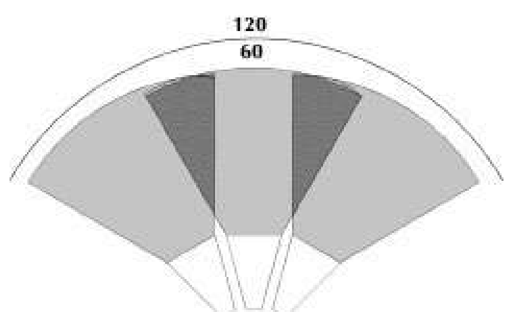


Figure 22-2. A very common array uses three 60° × 40° horns in enclosures with 15° trapezoidal sides; tight-packed, this array produces substantial overlap and interference between adjacent horns.

Tight-packing three of these systems with their 15° sides touching produces a 30° splay between the horns, for a total included angle of 120°. At first glance, this seems like an ideal alignment. But the EASE interference predictions in Fig. 22-3 show the familiar and clearly audible problems with this configuration: significant interference above 1 kHz, with variations of 8–9dB depending on the angle. On axis, there is about 10dB of gain at frequencies below 1 kHz. Where maximum SPL is the main consideration, this type of array will deliver acceptable performance. When the front-of-house mix position can be located on the axis of left and right arrays, they can usually be tweaked to deliver acceptable reproduction in this limited area. Other areas of the house, including the high roller seats up front, will suffer.

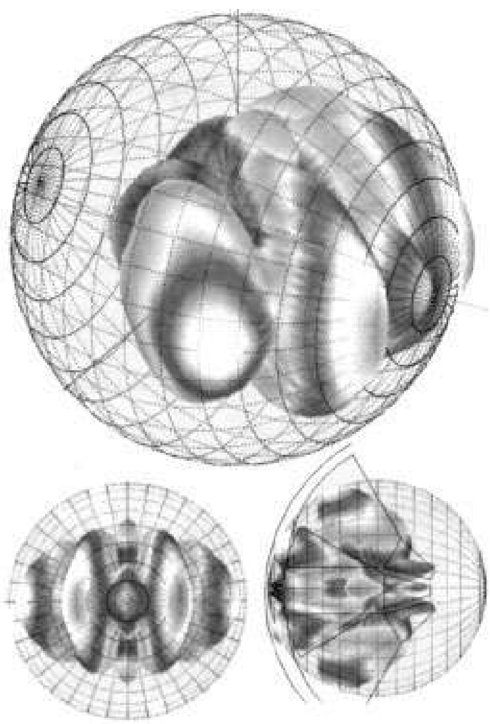


Figure 22-3. The interference patterns shown above were produced by tight-packing three arrayable loudspeakers using 60° × 40° constant directivity horns in enclosures with 15° trapezoidal

sides. While this is an improvement over a pile of direct radiating transducers, it is far from the ideal point source array.

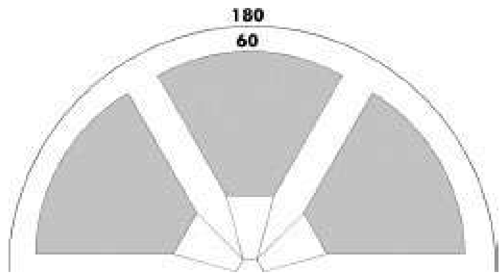


Figure 22-4. Widening the splay between horns reduces interference and widens the coverage angle to 180°, but reduces forward gain. As always, energy is conserved.

The interference patterns displayed in Fig. 22-3 can be reduced by widening the splay between cabinets to 30°, as illustrated in Fig. 22-4. This array will not look as pretty as the first, but it does have much more even response across the coverage area, Fig. 22-5. At 2kHz and 4kHz, the individual horns are clearly discernible in the ALS-1 predictions. Also note that the seams between the horns become deeper with increasing frequency.

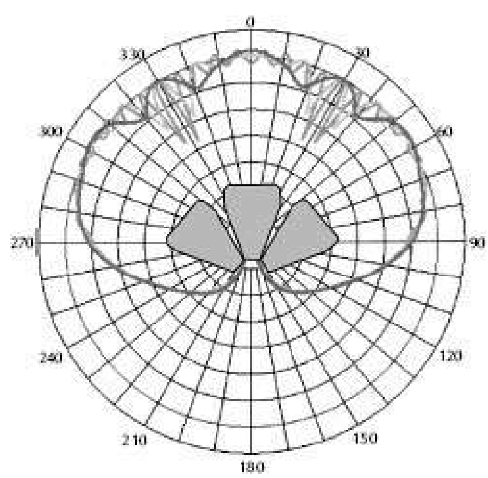


Figure 22-5. ALS-1 interference predictions for a wider splay show reduced interference, but the three horns are clearly apparent

at higher frequencies.

Fig. 22-6 shows why there will always be interference with conventional horn arrays (whether they are enclosed in arrayable cabinets with trapezoidal sides or mounted in free air). As the wavefronts radiate from points of origin that are separated in space, they will always create some interference at the coverage boundaries.

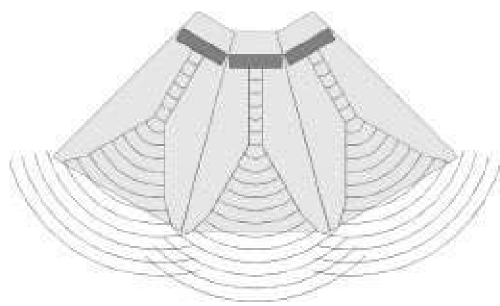


Figure 22-6. The acoustic pressure wave expands as a sphere, and multiple spherical sections will always overlap unless they originate from a common center.

22.4 Conventional Array Shortcoming Analysis

For an array in far field, dependence on angle is

$$SPL(\theta) = 10\log P_0^{\ 2}\ \mathrm{dB} \qquad (22\text{-}1)$$

For a distance to the listening area very much larger than the array dimensions, let the sound pressure *P* be the real part of

$$P(\theta) = A_i(\theta)^{j(\omega\tau - kSi)} \qquad (22\text{-}2)$$

where,
P is the sound pressure,

ω is the angular frequency,

$A_i(\theta)$ is a function of the angle between the array longitudinal axis and the direction of the distant listening point. It gives the ratio of the sound pressure due to the source as a ratio of its on-axis value at the same distance.

For the ith source shown in Fig. 22-7, assuming identical sources, the pressure contribution is given by

$$P_i = A_i(\theta)^{j(\omega\tau - kSi)} \qquad (22\text{-}3)$$

where,

k is $2\pi/\lambda = 2\,\lambda fc$,

λ is the wavelength,

f is the frequency,

c is the speed of sound,

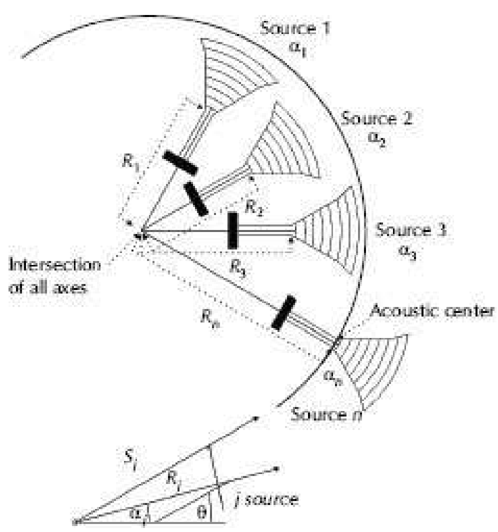


Figure 22-7. For a circular arc array, the additional path length Sj is as shown.

S_i is the distance by which the path length from the *i*th source to the distant point exceeds the distance from the origin to that point.

For an array of n sources, the total pressure P is given by

$$P(\theta) = \sum_{i=1}^{n} A_i q e^{j\omega\tau - kSi}$$
$$= \varepsilon^{j\omega\tau} \sum_{i=1} A(\theta)\varepsilon^{j\omega\tau kSi} \qquad (22\text{-}4)$$

The square of the pressure amplitude is given by

$$P_0^{\,2}(\theta) = \left[\sum_{i=1}^{n} A_i \theta k S_i\right]^2 + [A_i(q\theta)\sin(kS_i)]^2 \qquad (22\text{-}5)$$

where,
$A_i(\theta)$ is $A_i(\theta - \alpha_i)$.

For a circular arc array, the additional path length S_i as shown in Fig. 22-7, for the ith source at radius R and angle α is given by

$$-S_i(\theta) = R_i \cos(\theta - \alpha_i) \qquad (22\text{-}6)$$

Therefore, the smaller R_i is, the smaller the S_i differences, and the less the interference between sources. Ideally, $R = 0$ for all sources. As R approaches 0, the interference will become less audible and frequency response across the array's intended coverage area will become more uniform.

22.5 Coincident Acoustical Centers: A Practical Approach

Clearly, the ideal solution is to collocate all the acoustic points of origin, as shown in Fig. 22-8. We could achieve this by stacking the horns vertically, but this would solve the problem in the horizontal plane by creating a worse situation in the vertical (front to back) direction. Fig. 22-9 shows a more realistic approximation that takes into account the physical constraints of loudspeaker design (the dimensions of the transducers, horns, enclosure walls, etc.). Because the acoustic sources are real physical objects, we cannot reduce R_i to 0. But we can get close enough to make measurable, audible improvements in the performance of the multi-enclosure array.

22.5.1 TRAP Horns: A New Approach

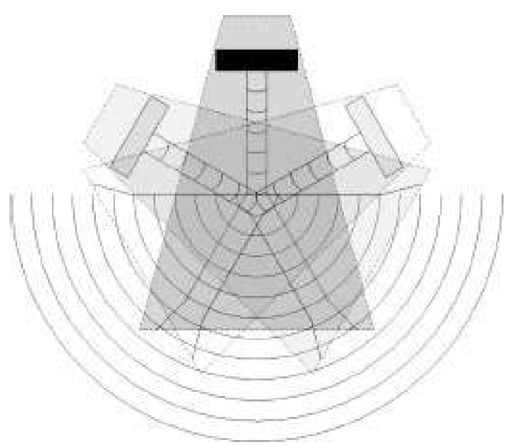


Figure 22-8. The acoustic ideal—colocating the acoustic centers of all horns is not a practical possibility.

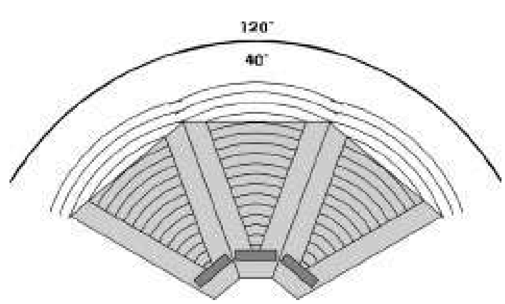


Figure 22-9. Because drivers and enclosures are physical objects, the acoustic centers of TRAP horns are not perfectly coincident, but they are close enough to achieve measurable and audible reductions in interference.

Fig. 22-9 implies that the way to minimize R_i, and the resultant interference, is to move the acoustic centers as far to the rear of the enclosure as possible. We can attempt to minimize the size of the drivers, for instance by using high-output magnetic materials such as neodymium. But the biggest obstacle to coincident acoustic centers is the horn itself. This is because typical constant directivity horns exhibit astigmatism: their apparent points of origin are different in the horizontal and vertical planes. In order to create a wider coverage pattern in the horizontal plane, the apparent apex is moved forward, while the vertical apex is farther to the rear because its coverage pattern is usually narrower. This is certainly the case with the most popular horn patterns in use today: 60° × 40° and 90° × 40°. One approach to a solution, then, is to rotate the horn and use the vertical apex of the horn in the horizontal plane. By doing so, we are effectively moving the acoustic center as far to the rear of the cabinet as possible. This technique when combined with cabinet design that minimizes the space between adjacent drivers in an array, while matching the trapezoidal sides with the opening angle of the horn, creates a system capable of minimal interference in the frequency range where the horn is effective. This forms the basis for what I call the True Array Principle by Renkus Heinz.

Subsequent refinements to the horn flare itself have been awarded U.S. Patent #5,750,943. This Arrayguide topology goes even farther in locating the apparent acoustic origin toward the rear of the enclosure. To repeat, moving the acoustic centers to the rear minimizes *R*, the distance between acoustic points of origin within the array, and the resulting interference between array elements.

Fig. 22-10 shows the ALS-1 predictions for the first generation of TRAP horns. It is clear that interference has almost disappeared.

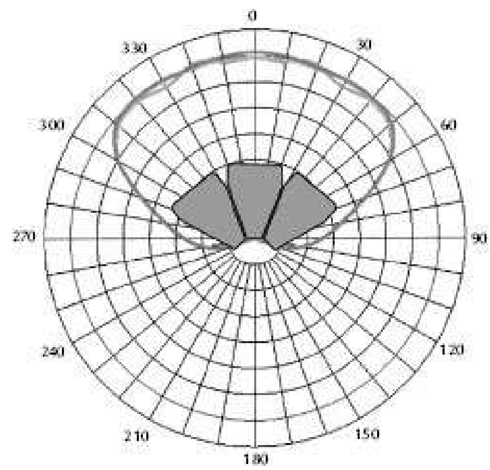


Figure 22-10. TRAP design produces truly arrayable systems with minimal destructive interference in the horns' passband.

Fig. 22-11 shows measured EASE data for a three-wide array of TRAP40 enclosures. Frequency response is consistent in both vertical and horizontal planes within ±4dB. This is an out of the box array, using no frequency shading or micro-delay to improve performance. Measured results don't track the predictions 100% because the actual pattern of the horns varies somewhat with frequency: first generation TRAP horns maintain nominal coverage ±10° from 1–4 kHz.

22.5.2 TRAP Performance

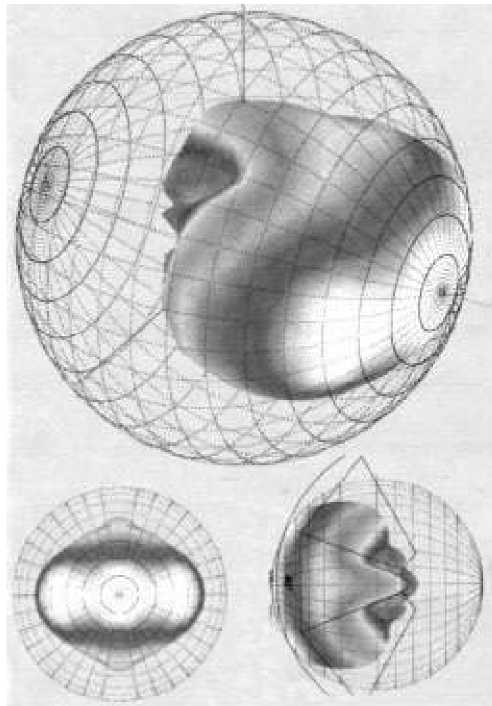


Figure 22-11. The TRAP array produces little measurable interference from a tight-packed three-wide cluster. This is because the three spherical wave-fronts produced by the three horns originate from a common acoustical center. Therefore they behave as a single acoustic unit, without overlap or interference.

Systems based on the True Array Principle can extend pattern bandwidth (the frequency range over which coverage varies less than ±5°) down to the frequency at which mutual coupling between adjacent cabinets ceases. TRAP systems are designed so that the enclosures provide optimum splay angles of 40° between the horns: the trapezoidal sides are therefore steeper than many other designs at 20° per side. The combination of symmetrical horns and steeper sidewall angles maintains coincident acoustic centers for all the elements in the array.

Note that moving the horizontal apex to the same location as the vertical results in a symmetrical 40° × 40° coverage pattern. This in turn requires the use of four enclosures to cover 160° with almost no variation in frequency response in the horizontal (side to side) plane. With 60° × 40° cabinets we could deliver sound to 180° of coverage, albeit with some quite audible variations.

Figure 22-12. TRAP arrays can be quite small; however, the size of the horns will determine the lower frequency limit at which the True Array Principle ceases to operate.

There are other commercially available systems offering similar array performance to that described above. The ARC's system from French loudspeaker manufacturer L-Acoustic uses a type of path length equalizer to force the emerging wavefront to conform to the opening angle of their horn and also puts the acoustic center behind the cabinet. In the case of ARC, the cabinet's trapezoidal side walls also serve as the waveguide for the high frequencies. As the waveguides opening angle matches that of the cabinet, this is certainly an elegant solution to creating minimum interference arrays at the frequency where the horn is effective.

In the KF900 series from EAW, simple phase horns for the mid and high frequencies put the acoustic center as close to the rear of the cabinet as possible, while their opening angles also match the trapezoidal sides of the enclosure. The relatively large size of the KF900 series enclosures and horns brings minimum interference performance to frequencies lower than those based on smaller waveguides. Remember, that this technique for minimum interference arrays, including the True Array Principle, only holds true for those frequencies where the horn is effective.

22.6 Low Frequency Arrays: Beneficial Interference

In the preceding paragraphs, I outlined the parameters necessary to minimize destructive acoustic interference between adjacent cabinets or horns in an array. But these techniques are only beneficial at the frequencies where the horns are effective. Yet these very systems or horns are used at frequencies well below their directivity cutoff and lower, down to frequencies where the woofers piston size offers no directional control at all.

22.6.1 Horizontal Woofer Arrays: Maintaining Wide Dispersion

For our first example, let's look at the additional problems and opportunities we create when arraying small (12 inch woofer, 1 inch compression driver) full range loudspeaker enclosures as in Fig. 22-12. For a full range array module, there are three frequency zones that exhibit different wavelength related behavior. At the lowest frequencies, or longest wavelengths, these modules exhibit only beneficial interference or mutual coupling. Each additional module creates additional on axis acoustic output. The opportunity here, is that less equalization is required to make the array's frequency response flat down to these lower frequencies as compared to a single cabinet.

A potential problem is created when the array becomes too wide however. Four or five element arrays are wide enough as to become quite directional in the forward plane at those lower frequencies (20Hz to roughly 500Hz or more, dependent on the module). Without a signal processing scheme, this array cannot be equalized to have the same frequency response throughout its intended coverage. It will sound boomy in the middle and thin at its coverage extremes. A solution is to taper the length of the array in the horizontal plane in order to maximize horizontal dispersion of the lower frequencies. The entire array can be used for the lowest frequencies as the wavelengths are longest (20Hz up to about 200Hz), but at higher frequencies, as the wavelengths get shorter, the array length must also get shorter to maintain wide dispersion. This is achieved by low passing the outermost woofers of the array, such that only two or three max woofers are used at frequencies higher than this.

The second frequency zone that can be problematic in arrays based on full range modules, occurs at wavelengths where cabinet spacing no longer supports mutual coupling, and the horn has yet to attain its directivity cutoff. This typically applies to a small half octave range where adjacent cabinet spacing approaches a wavelength. Here we observe combinations of destructive and constructive interference at various observation points around the array intended coverage, causing frequency response variations greater than ±6dB. Fortunately there is a signal processing technique that can minimize this effect. By simply notching this frequency range from every other cabinet with a cut equal to the greatest amount a variance (typically 6dB of attenuation), and width equal to the bandwidth of the aberrations (typically half an octave), the frequency response variations throughout the arrays coverage can minimized.

The third frequency zone of wavelength related behavior for arrays based on full range modules, is then at the frequencies above which the horn is effective. Let us assume that the horns depicted in Fig. 22-12 place the acoustic center towards the rear of the cabinets, and that their opening angle also matches that of the trapezoidal sides of the cabinet. Based on these assumptions, the array performance will exhibit minimum interference for frequencies above 1–2kHz which happens to be the effective directivity cutoff of the horn. Each additional module simply adds additional coverage to the array.

22.6.2 Vertical Woofer Arrays

Directivity at Frequencies Where Size Makes Horns Impractical

Beneficial destructive interference sounds like an oxymoron, but there are several commercially available woofer arrays that take advantage of this very technique. By applying the fundamental physics described by Harry Olson, directional woofer arrays are now available that outperform large woofer horns.

When two point sources are superimposed on one another, their outputs simply add up in all directions. As the two point sources are spread apart, the output diminishes along the plane of separation due to phase cancellation. At exactly ½ wavelength, a pure null occurs, and we achieve the classic figure eight, dipole polar pattern. The current commercially available systems take advantage of this phenomena, directivity through off axis attenuation, by placing woofers in a vertical array and spacing them to create this dipolar pattern at frequencies below which horns become too large.

Fig. 22-13 is an example of one such array. Termed *Tri-Pola*r by designer Vance Breshears, it uses the vertical spacing between the three woofers with appropriate signal processing to maintain consistent low frequency pattern control from 400Hz down to below 100Hz. One of the first systems available was developed by Craig Janssen, termed *Tuned Dipolar*, it uses two separate arrays. With drivers, spacing and signal processing appropriate for their respective passbands Tuned Dipolar offer exceptional low frequency pattern control over an extended bandwidth. Even subwoofers are now benefitting from this type of technology. Meyer Sound is achieving cardioid patterns at lowest frequencies from its PSW-6, providing significant attenuation

Figure 22-13. Reference Point Array using four 40° x40° mid-high enclosures and six low frequency modules in Tri-Polar configuration for vertical pattern control, along with appropriate small full range systems for downfill.

22.7 Line Arrays and Digitally Steerable Loudspeaker Column Arrays

For the communication between a source and a listener to be effective, it is important that the listener receive and comprehend the message. In large spaces where people gather, including auditoria, houses of worship, sports venues, transit terminals and classrooms, often the acoustic requirements that enable effective speech are in conflict with the architectural needs of the spaces. When the acoustics of a venue cannot be altered to enable effective speech communication, designing a sound reinforcement system to do so, can be a challenge. Recent advances in efficient amplification and digital signal processing have enabled a new class of loudspeaker; the *digitally steerable column* or *line array* as its often called. The acoustical and architectural benefits of these loudspeakers for sound reinforcement in highly reverberant or reflective environments will be shown.

We will discuss effective communications and define intelligibility and how to measure it both subjectively and objectively. We will look at architecture and acoustics and at reverberation and its effect on intelligibility in large public spaces. Finally we'll look at digitally steerable column arrays, their design considerations, and their performance and benefits when used in large reverberant spaces.

Some of the basic principles involved in voice communications are:

- In voice communications intelligibility is the capability of being understood.
- It assumes the existence of a communication process between a talker and a listener, or between a source and a listener.
- For the conveyance of meaning, the English language is highly dependent upon the effective receipt and comprehension of consonants. This is how we differentiate words based on similar vowels. For example, Zoo, Two, New.
- In terms of frequency response, speech ranges between 100Hz and 8kHz, with maximum energy around 250Hz.
- In speech, the frequency range that conveys the most consonant information is the octave around 2KHz.

22.7.1 What Affects Intelligibility

Major Influences that affect intelligibility are:

- Elocution and pronunciation of the talker. It's hard to understand someone who mumbles under any condition.
- Hearing acuity of listener. An often overlooked influence, those with a hearing loss have trouble understanding what's being said.
- SNR. We've all been places where it was so noisy we couldn't understand what was being said.
- Direct to reverberant ratio. The higher the reverberation level, the more difficult it is to understand what's being said.
- Directivity of the loudspeaker or loudspeakers. Highly directional loudspeakers direct more of the sound onto the audience and less onto the reflective walls and ceilings.
- The number of loudspeakers. Larger numbers of loudspeakers

translate into more acoustic energy being transmitted into the room and higher reverberation levels.

- Reverberation time. The longer the reverberation time, the more likely it will interfere with intelligibility.
- Distance of source to listener, The closer the listener is to the loudspeaker, the less likely reverberation will interfere.

Secondary Influences are:

- Gender of talker.
- Microphone technique.
- Vocabulary and context of speech information.
- Direction of main sound to listener and/or direction of reflections and echoes.
- System fidelity, equalization, and distortion.
- Uniformity of coverage.

22.7.2 Measuring Intelligibility

22.7.2.1 Subjectively

Statistical tests with trained talkers and listeners can be the most reliable metric for determining the intelligibility of a system. To ensure that all speech sounds are represented in a test, Phonemically Balanced (PB) word lists are commonly used. These word list can be a long as 1000 words. Tests using nonsense syllables or logatoms, and Modified Rhyme Tests are also used. These tests are very time consuming and are difficult to set up.

22.7.2.2 Objectively

Articulation Index. Articulation Index or AI was one of the first attempts to quantify intelligibility with measurements. AI is primarily concerned with the affect of noise on speech. The index ranges from 0 to 1 with 0 representing no intelligibility.

%ALcons. %ALcons or the articulation loss of consonants was developed by Peutz in Holland during the 1970s. %ALcons takes both noise and reverberation into account and is based the importance of the octave around 2000Hz in conveying consonant information. %ALcons uses a scale running downwards from 0 where 0 is perfect intelligibility, or 0% articulation loss.

Although Peutz used 2000Hz as the center frequency and 2000Hz is still the European standard, many acousticians in the USA prefer using 1000Hz. As a general rule, %ALcons calculated at 1000Hz show a higher articulation loss than ones calculated at 2000Hz.

STI. STI or Speech Transmission Index considers the source/room/listener as a transmission channel and measures the reduction in modulation depth of a specialized test signal which replicates the burst nature of real speech. The STI scale ranges from 0 to 1, where 1 represents perfect intelligibility. STI is considered the most accurate of the intelligibility measures, Table 22-1.

Table 22-1. STI Scale Ranges

Evaluation	STI	%ALcons
Bad	0.20 to 0.34	24.3 to 57
Poor	0.35 to 0.50	11.3 to 24.2
Fair	0.51 to 0.64	5.1 to 11.2
Good	0.65 to 0.86	1.6 to 5.0
Excellent	0.87 to 1.00	0.0 to 1.5

Copied from *The Audio System Designer Technical Reference* by Peter Mapp and published by Klark Teknik.

22.7.3 Architecture and Room Acoustics

Reverberation

Reverberation is the persistence of sound in a space after the original sound has been removed.

RT is the measure for reverberation, and it is defined as the amount of time required for the average sound energy density in a space to decrease from its original value by 60dB after the original sound has stopped.

The Sabine equation relates RT to the volume of a room with it surface area and the absorption coefficients of the materials applied to the surfaces.

As room volume increases relative to surface area and absorption coefficients, the RT increases.

As surface area and absorption increase relative to room volume, RT decreases. It is this persistence of sound that interferes with our comprehension of consonants and contributes towards degrading intelligibility, Table 22-2.

Table 22-2. Intelligibility Comparison Chart

RT	< 1s	Excellent intelligibility can be achieved.
RT	1 to 1.2 s	Excellent to good intelligibility is possible.
RT	1.2 to 1.5s	Good intelligibility can be achieved.
RT	>1.5 s	Careful system design is required.
RT	>1.7s	Limit for good intelligibility in large spaces.

RT	>2s	Very directional loudspeakers are required, intelligibility can have limitations.
RT	>2.5s	Intelligibility will probably have limitations.
RT	>4s	Highly directional loudspeakers will be required to achieve acceptable intelligibility.

22.7.4 Line Arrays

Figs. 22-14 to 22-16 show the direct sound coverage of various loudspeakers in a sanctuary 100ft × 65ft (30m × 20m. The chancel adds 20ft (6m) to its length. The roof peaks at 52ft (15.8m). The room volume is roughly 250,000ft^3 (7080m^3). The room has plaster walls, wood ceiling, terrazzo floors, and empty wooden pews. This produces a RT of about 3.5s.

Figure 22-14. Flown large format horn array.

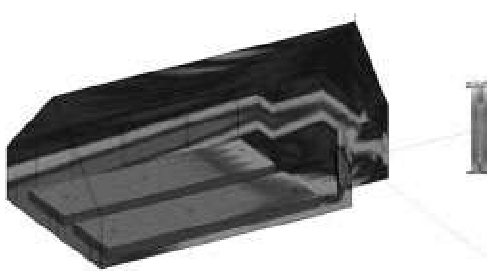


Figure 22-15. Mechanically tilted four meter column array.

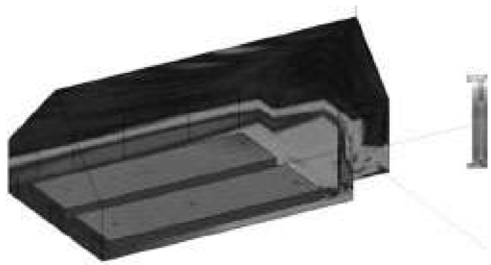


Figure 22-16. Digitally steered column array.

Notice the high *SPL* levels on the walls and ceiling in the flown-horn array simulation. The high frequency beaming of the mechanically tilted column array prevents good coverage of the front of the audience area. The digitally steerable column array covers only the audience area and has very little coverage on the walls and no coverage of the ceiling. Only the steered column array has acceptable (good to fair) intelligibility throughout the audience areas. Digitally steerable column arrays can offer superior coverage and they can provide improved D/R. They can provide improved intelligibility in highly reverberant spaces, plus they blend better with their surrounding architecture and are nearly invisible in use.

22.7.4.1 Digitally Steered Line Arrays

When the room size and volume are fixed and adding absorption to reduce the RT times is not an option, digitally steerable column arrays offer a new solution:

- They have the ability to be much more directional than the largest horns.
- The idea is not new; the concepts for these column arrays were described by Harry Olson in 1957. Only the implementation is new.
- The hardware required to implement these ideas is now available.

- Digital Signal Processing required is now a mature technology, very powerful and relatively inexpensive.
- Compact, highly efficient Class D amplifiers are capable of high-fidelity performance.

Line Arrays are not a new idea. Harry F. Olson did the math and described the directional characteristics of a continuous line source in his classic *Acoustical Engineering*, first published in 1940. Traditional column loudspeakers have always made use of line source directivity.

Simple line arrays (column arrays) are basically a number of drivers stacked closely together in a line, Fig. 22-17. Simple line arrays become increasingly directional in the vertical plane as the frequency increases. The spacing between drivers controls the high frequency limits. The height (length) of the line array determines the low frequency control limit. Fig. 22-18 shows the line source directivity as described by Harry Olson in 1957.

The directivity of a line array is a function of the line length and the wavelength. As the wavelength approaches the line length, the array becomes omnidirectional, Fig. 22-19. Fig. 22-20 shows the vertical dispersion pattern of a typical line array.

22.7.4.2 Controlling High Frequency Beaming

Simple line arrays become increasingly directional as the frequency increase., in fact, at higher frequencies they become too directional. The vertical directivity can be made more consistent by making the array shorter as the frequency increases by using fewer drivers. One amplifier channel and one DSP channel per driver make this possible.

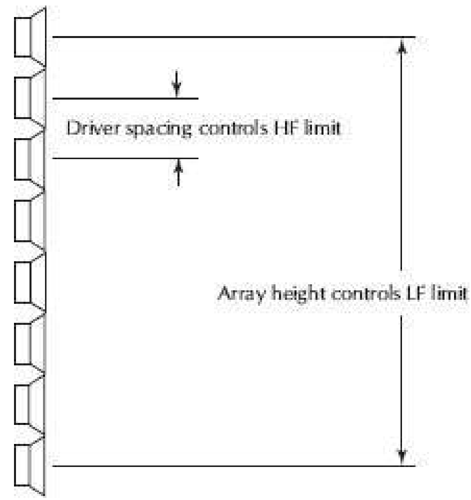


Figure 22-17. Basic line array theory.

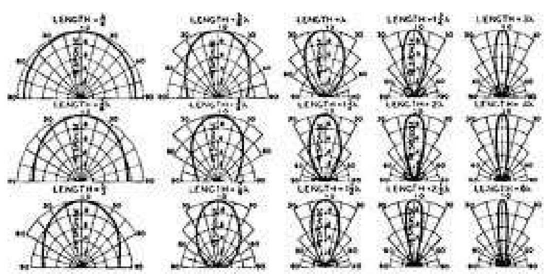


Figure 22-18. Directional characteristics of a line source as a function of the length and the wavelength. The polar graph depicts the sound pressure at a large fixed distance, as a function of angle. The sound pressure for the angle 0° is arbitrarily chosen as unity. The direction corresponding to the angle 0° is perpendicular to the line. The directional characteristics in 3D are surfaces of revolution about the line as an axis. (From *Acoustical Engineering* by Harry Olson.)

22.7.4.3 Beam Steering

The beam can be steered up or down by delaying the signal to adjacent drivers. DSP control also allows us to develop multiple beams from a single line array and individually steer these beams.

DSP control also allows us to move each beams acoustic center up and down the column allowing us to create multiple beams and also steer the beam, Figs. 22-21 and 22-22.

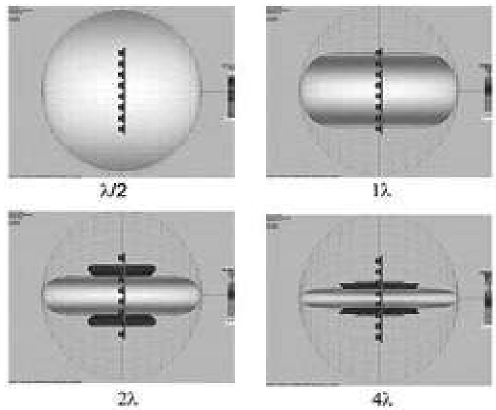


Figure 22-19. Simple line source directivity as a function of line length versus frequency.

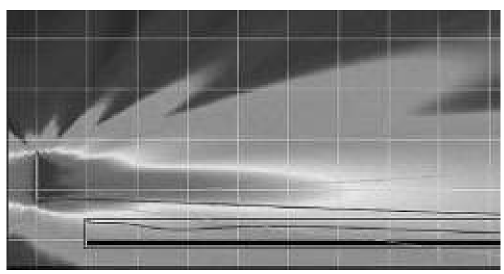


Figure 22-20. Typical line array vertical dispersion display.

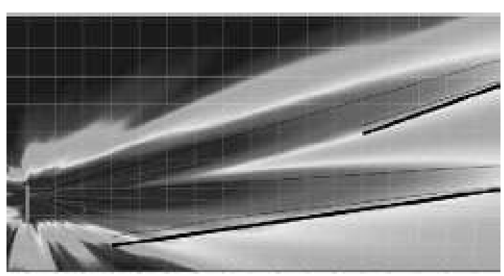


Figure 22-21. Vertical dispersion display showing multiple beam capability.

22.7.5 DSP-Driven Vertical Arrays

22.7.5.1 Acoustical, Electronic, and Mechanical Considerations

Practical examples are taken from the Renkus-Heinz IC Series Iconyx steerable column arrays. Iconyx is a steerable column array that combines very high directivity with accurate reproduction of source material in a compact and architecturally pleasing package, Fig. 22-23.

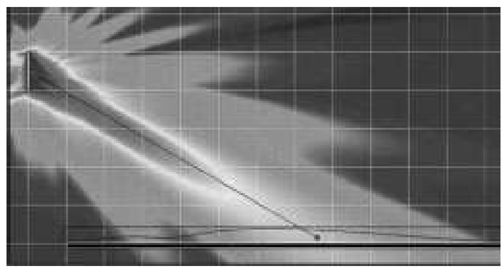


Figure 22-22. Graphic illustration of beam steering.

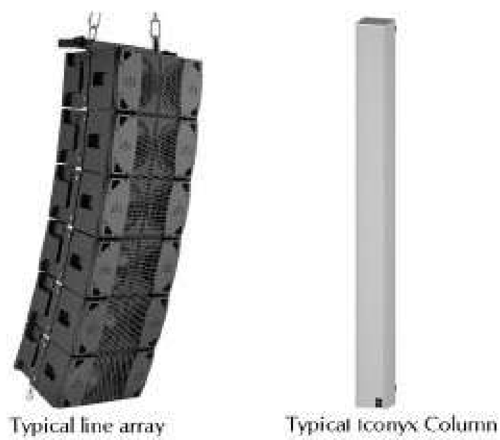


Figure 22-23. Typical line array and a typical Iconyx Column.

Like every loudspeaker system, Iconyx is designed to meet the challenges of a specific range of applications. Many of the critical design parameters are, of course, determined by the nature of these target applications. To understand the decisions that have been made during the design process we must start with the particular problems posed by the intended applications.

The function of individual driver control and DSP is to make more effective use of this phenomenon. No amount of silicon can get around the laws of acoustical physics. The acoustical properties of first-generation column loudspeakers are set by the acoustical characteristics of the transducers and the physical characteristics of the package:

1. The height of the column determines the lowest frequency at which it exerts any control over the vertical dispersion.
2. The inter-driver spacing determines the highest frequency at which the array acts as a line source rather than a collection of separate sources.
3. Horizontal dispersion is fixed and is typically set when the drivers are selected, because column loudspeakers do not have waveguides.
4. Other driver characteristics such as bandwidth, power handling and sensitivity will determine the equivalent performance characteristics of the system.

One unfortunate corollary of these characteristics is that the power response of a conventional column loudspeaker is not smooth. It will deliver much more low-frequency energy into the room and this energy will tend to have a wider vertical dispersion. This can make the critical distance even shorter because the reverberant field contains more low-frequency energy, making it harder for the listener to recognize higher-frequency sounds such as consonants or instrumental attack transients.

22.7.5.1.1 Doublet Source Directivity

Doublet source cancels each other’s output directly above and below, because they are spaced 1/2 wavelength apart in the vertical plane. In the horizontal plane, both sources sum. The overall output looks like Fig. 22-24.

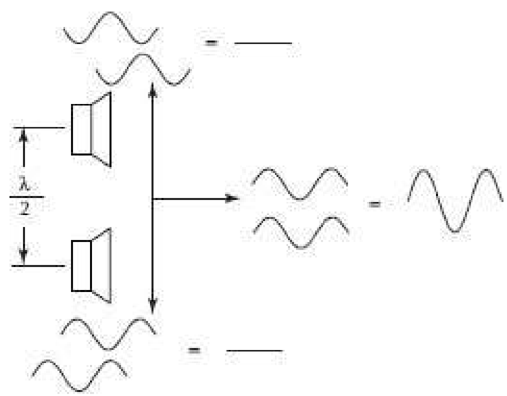


Figure 22-24. Output of a signal whose wavelength is 1/2 of the space between the two loudspeakers.

When two sources are 1/4 wavelength apart or less, they behave almost like a single source. There is very slight narrowing in the vertical plane, Fig. 22-25.

There is significant narrowing in the vertical plane at 1/2 wavelength spacing, because the waveforms cancel each other in the vertical plane, where they are 180° out of phase, Fig. 22-26.

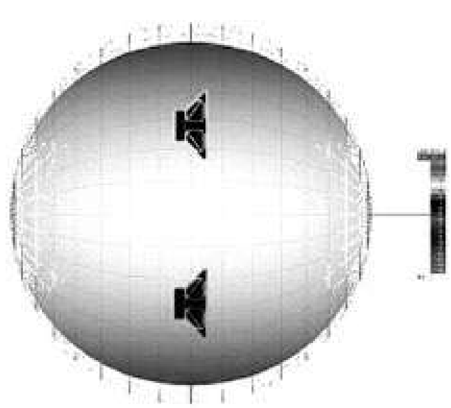


Figure 22-25. A/4 (1/4 wavelength).

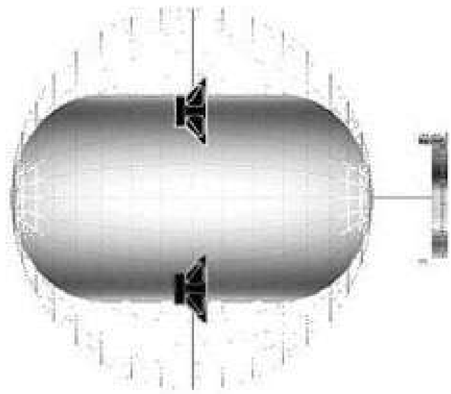


Figure 22-26. λ/2 (1/2 wavelength).

At one wavelength spacing the two sources reinforce each other in both the vertical and horizontal directions. This creates two lobes, one vertical and the other horizontal, Fig. 22-27.

As the ratio of wavelength to inter-driver spacing increases, so do the number of lobes. With fixed drivers as used in line arrays, the ratio increases as frequency increases ($\lambda = c/f$ where f is the frequency and c is the speed of sound), Fig. 22-28.

22.7.5.1.2 Array Height versus Wavelength (λ)

Driver-to-driver spacing sets the highest frequency at which the array operates as a line source. The total height of the array sets the lowest frequency at which it has any vertical directivity.

Figs. 22-29 though Fig. 22-32 show the effect of array height versus wavelength.

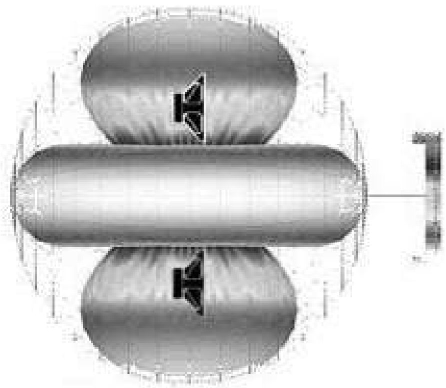


Figure 22-27. λ (1 wavelength).

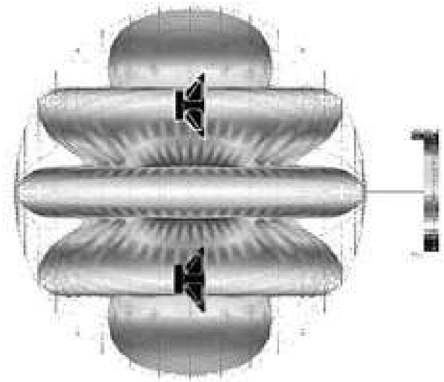


Figure 22-28. Increased wavelength to inter-driver spacing.

At wavelengths of twice the array height, there is no pattern control, the output is that of a single source with very high power handling, Fig. 22-29.

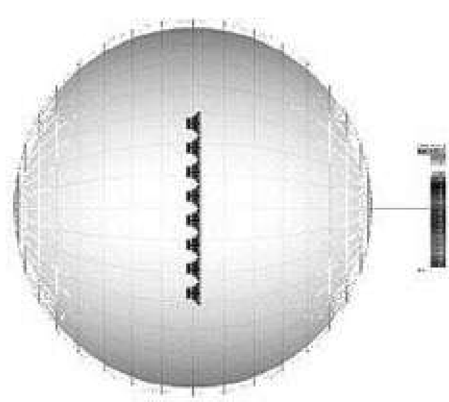


Figure 22-29. Wavelength is twice the loudspeaker height.

As the frequency rises, wavelength approaches the height of the line. At this point there is substantial control in the vertical plane, Fig. 22-30.

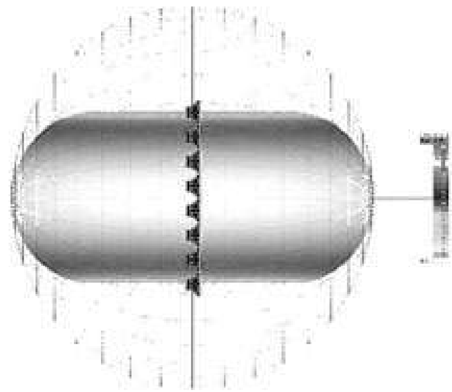


Figure 22-30. Wavelength is the loudspeaker height.

At higher frequencies the vertical beamwidth continues to narrow. Some side lobes appear but the energy radiated in this direction is not significant compared to the front and back lobes, Fig. 22-31.

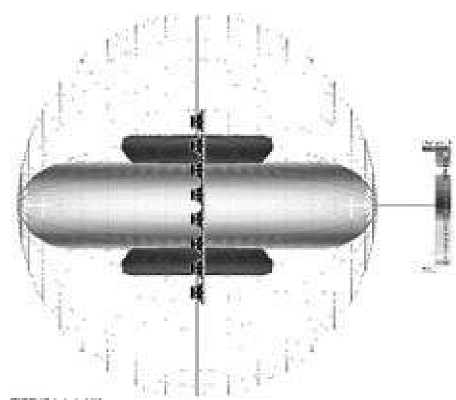


Figure 22-31. Wavelength is one half the loudspeaker height.

Still further vertical narrowing, with side lobes becoming more complex and somewhat greater in energy, Fig. 22-32.

22.7.5.1.3 Inter-Driver Spacing versus Wavelength (λ)

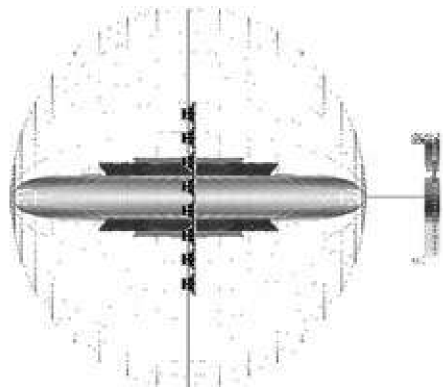


Figure 22-32. Wavelength is ¼ the loudspeaker height.

The distinction between side lobes and grating lobes should to be maintained. Side lobes are adjacent to and radiate in the same direction as the primary lobe. Grating lobes are the strong summations tangential to the primary lobe. Side lobes will be present in any realizable line array, grating lobes form when the inter driver spacing becomes less than 1/2 wavelength. It might also be good to point out that all of the graphics for this section are done using theoretical point sources.

Figs. 22-33 though Fig. 22-36 show the effect of inter-driver spacing versus wavelength.

When the drivers are spaced no more than 1/2 wavelength apart, the array produces a tightly directional beam with minimal side lobes, Fig. 22-33.

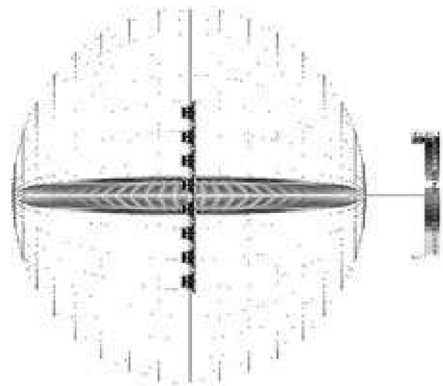


Figure 22-33. Interspacing is 1/2 the wavelength.

As the frequency rises, wavelength approaches the spacing between drivers. At this point, grating lobes become significant in the measurement. They may not be a problem, if most or all of the audience is located outside these vertical lobes, Fig. 22-34.

At still higher frequencies, lobes multiply and it becomes harder to isolate the audience from the lobes or their reflections, Fig. 22-35.

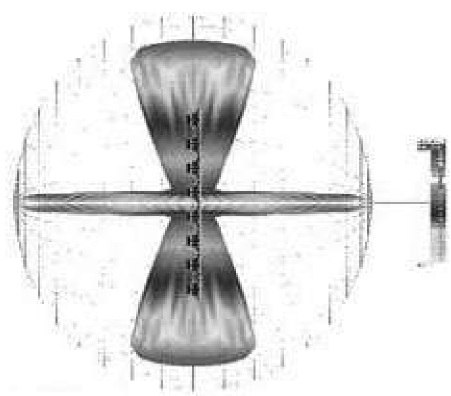


Figure 22-34. Interspacing is one times the wavelength.

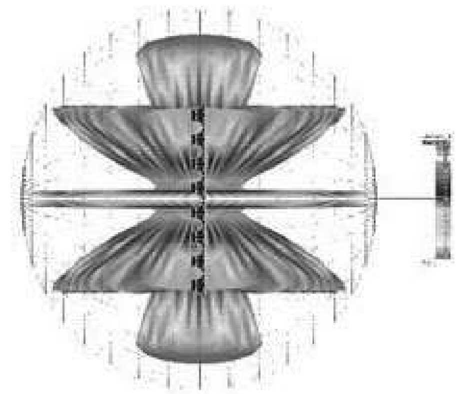


Figure 22-35. Interspacing is two times the wavelength.

As inter-driver spacing approaches four times the wavelength, the array is generating so many grating lobes of such significant energy that its output closely approximates a single point source, Fig. 22-36. We have come full circle to where the array's radiated energy is about the same as it was when array height was 1/2 λ. As shown in Fig. 22-32, this is the high frequency limit of line array directivity.

As real drivers are considerably more directional than point sources at the frequencies where grating lobes are generated, the grating lobes are much lower in level than the primary lobe, Figs. 22-37 and 22-38.

22.7.6 Multichannel DSP Can Control Array Height

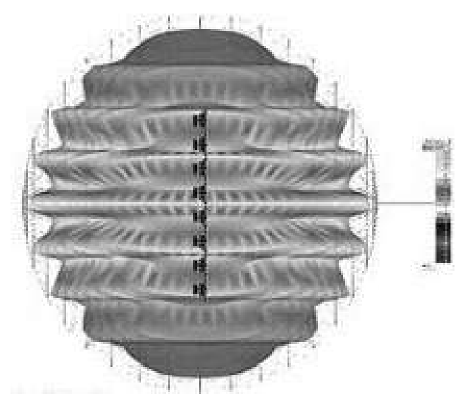


Figure 22-36. Interspacing is four times the wavelength.

The upper limit of a vertical array's pattern control is always set by the inter-driver spacing. The design challenge is to minimize this dimension while optimizing frequency response and maximum output and do it without imposing excessive cost. Line arrays become increasingly directional as frequency increases, in fact, at high frequencies they are too directional to be acoustically useful. However, if we have individual DSP available for each driver, we can use it to make the array acoustically shorter as frequency increases—this will keep the vertical directivity more consistent. The technique is conceptually simple; use low-pass filters to attenuate drive level to the transducers at the top and bottom of the array with steeper filter slopes on the extreme ends and more gradual slopes as we progress to the center. As basic as this technique is, it is practically impossible without devoting one amplifier channel and one DSP channel to each driver in the array.

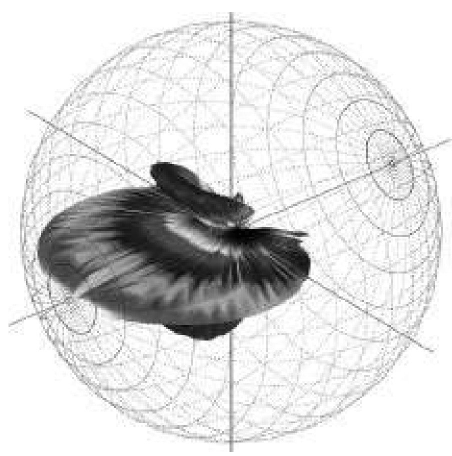


Figure 22-37. 3D view of a second generation Iconyx array at 4000Hz.

A simplified schematic shows how multichannel DSP can shorten the array as frequency increases. For clarity, only half the processing channels are shown and

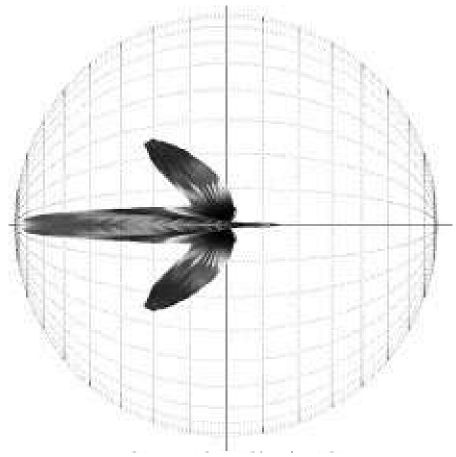


Figure 22-38. Side view of a second generation Iconyx array at 4000 Hz.

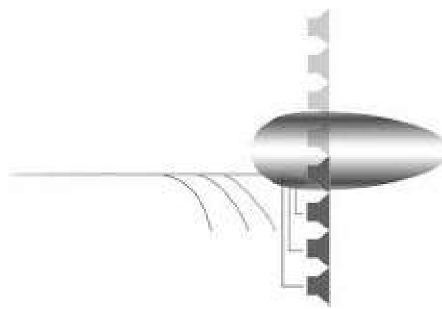


Figure 22-39. Multichannel DSP shortens the loudspeaker length.

22.7.7 Steerable Arrays May Look Like Columns but They Are Not

Simple column loudspeakers provide vertical directivity, but the height of the beam changes with frequency. The overall Q of these loudspeakers is therefore lower than required. Many early designs used small-cone full range transducers, and the poor high-frequency response of these drivers certainly did nothing to enhance their reputation.

22.7.7.1 Beam-Steering: Further Proof that Everything Old is New Again

As Don Davis famously quotes Vern Knudsen, “The ancients keep stealing our ideas.” Here is another illustration from Harry F. Olson’s *Acoustical Engineering*. This one shows how digital delay, applied to a line of individual sound sources, can produce the same effect as tilting the line source. It would be long after 1957 before the cost of this relatively straightforward system became low enough for commercially viable solutions to come to market, Fig. 22-40.

22.7.7.2 DSP-Driven Arrays Solve Both Acoustical and Architectural Problems

22.7.7.2.1 Variable Q

DSP-driven line arrays have variable Q because we can use controlled interference to change the opening angle of the vertical beam. The Renkus Heinz IC Series can produce 5°, 10°, 15° or 20° opening angles if the array is sufficiently tall (an IC24 is the minimum required for a 5° vertical beam). This vertically narrow beam minimizes excitation of the reverberant field because very little energy is reflected off the ceiling and floor.

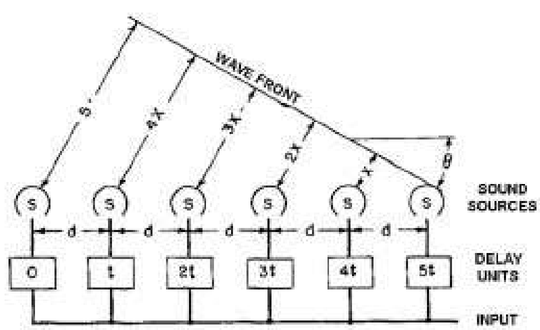


Figure 22-40. A delay system for tilting the directional characteristic of a line sound source. (From *Acoustical Engineering*

by Harry Olson.)

22.7.7.2.2 Consistent Q with Frequency

By controlling each driver individually with DSP and independent amplifier channels, we can use signal processing to keep directivity constant over a wide operating band. This not only minimizes the reverberant energy in the room, but delivers constant power response. The combination of variable Q, which is much higher than that of an unprocessed vertical array, with consistent Q over a relatively wide operating band, is the reason that DSP-driven Iconyx arrays give acoustical results that are so much more useful.

22.7.7.2.3 Ability to Steer the Acoustic Beam Independently of the Enclosure Mounting Angle

Although beam-steering is relatively trivial from a signal-processing point of view, it is important for the architectural component of the solution. A column mounted flush to the wall can be made nearly invisible, but a down-tilted column is an intrusion on the architectural design. Any DSP-driven array can be steered. Iconyx also has the ability to change the acoustic center of the array in the vertical plane which can be very useful at times.

22.7.7.2.4 Design Criteria: Meeting Application Challenges

The previous figures make it clear that any line source, even with very sophisticated DSP, can control only a limited range of frequencies. However, by using full range coaxial drivers as the line source elements could make the overall sound of the system more accurate and natural without seriously compromising the benefits of beam-shaping and steering. In typical program material, most of the energy is within the range of controllable frequencies. Earlier designs radiate only slightly above and below the frequencies that are controllable. Thus much of the program source is sacrificed, without a significant increase in intelligibility.

To maximize the effectiveness of a digitally controlled line source, it's not enough to start with high quality transducers. The Renkus Heinz Iconyx loudspeaker system uses a compact multichannel amplifier with integral DSP capability. The D2 audio module has the required output, full DSP control, and the added advantage of a purely digital signal path option. When PCM data is delivered to the channel via an AES/EBU or CobraNet input, the D2 audio processor/amplifier converts it directly into PWM data that can drive the output stage.

22.7.7.2.5 Horizontal Directivity Is Determined by the Array Elements

Vertical arrays, including Iconyx, can be steered only in the vertical plane. Horizontal coverage is fixed and is determined by the choice of array elements. The transducers used in Iconyx modules have a horizontal dispersion that is consistent over a wide operating band, varying between 140° and 150° from 100Hz to 16kHz.

22.7.7.2.6 Steering is Simple—Just Progressively Delay Drivers

If we tilt an array, we move the drivers in time as well as in space. Consider a line array of drivers that is hinged at the top and tilted downward. Tilting moves the bottom drivers further away from the listener in time as well as in space. We can produce the same acoustical effect by applying progressively longer delays to each driver as we move from top to bottom of the array.

Again, steering is not a new idea. It is different from mechanical aiming–front and rear lobes steer the same direction.

22.7.7.3 BeamWare: The Software That Controls Iconyx Linear Array Systems

A series of low-pass filters can maintain constant beam-width over the widest possible frequency range. The ideas are simple, but for the most basic Iconyx array, the IC16, we must calculate and apply 16 sets of FIR filters, and 16 separate delay times. If we intend to take advantage of constant inter-driver spacing to move the acoustical center of the main lobe above or below the physical center of the array, we must calculate and apply a different set of filters and delays. Theoretical models are necessary, but the behavior of real transducers is more complex than the model. Each of the complex calculations underlying the Iconyx beam-shaping filters were simulated, then verified by measuring actual arrays in our robotic test and measurement facility. Fortunately, the current generation of laptop and desktop CPUs are up to the task. BeamWare takes user input in graphic form (side section of the audience area, location and mounting angle of the physical array) and provides both a simulation of the array output that can be imported into EASE v4.0 or higher, and a set of FIR filters that can be downloaded to the Iconyx system via RS422 serial control. The result is a graphical user interface that delivers precise, predictable and repeatable results in real-world acoustical environments.

22.8 Steerable Column Arrays: One of the Fastest Growing Market Segments in Loudspeaker Design

Despite the complexities involved with the simulation, design and implementation of steerable columns, they now represent the fastest growing segment in loudspeaker design. For good reason too, it turns out they are a great tool for many applications, not just highly reverberant cathedrals, mosques and transportation hubs. With its discrete profile, the steerable column array suits many architecturally sensitive spaces as a sound reinforcement solution. So along with the early adopters of this technology, Duran Audio with their Intellivox arrays, Renkus-Heinz with their Iconyx arrays, Fig. 22-41, and EAW with their DSA, Fig. 22-42, we now have a dozen or more new players in the field, most notably Tannoy with their Q-Flex series, Fig. 22-43, RCF with their VSA arrays, and Meyer Sound with their new CAL arrays, Fig. 22-44.

Figure 22-41. Duran AXYS Target U-16 full range steerable array

Figure 22-42. Renkus-Heinz Iconyx loudspeaker.

Figure 22-43. EAW DSA250c loudspeaker.

Most of these newer designs bring an element of further refinement to this performance category. Many of these later implementations have addressed the compromises in performance at the higher frequencies of operation. By placing drivers ever closer, (the Renkus-Heinz Iconyx arrays now use a co-ax with a triple tweeter array, see Fig. 22-41) and by increasing the granularity of amplification and DSP, these newer arrays reduce the strength of side lobes, push the onset of gradient lobes out beyond the speech range of frequencies, and increase the steering capabilities. In short they have become even better, and more importantly, have proven themselves to be a long term, reliable solution, in problematic acoustic spaces.

Figure 22-44. Tannoy RWGPSW2 loudspeaker.

However, all the steerable columns mentioned above are just that, steerable columns with transducers, amplifiers and enclosures scaled for vocal reinforcement in large, reverberant spaces with low levels of background noise. But why limit this technology to largely speech oriented performance, surely conventional concert and opera halls, as well as contemporary worship spaces, convention centers, ball rooms, in addition to arena, stadium, and festival sites can all benefit from “steered” sound. Fortunately this technology can scale up for these types of applications such as the Renkus-Heinz ICLive steerable array and IC2 (pronounced IC Square) steerable point source loudspeakers.

With the hard part done, scaling beam steering technology up for greater output and fidelity is a relatively simple matter. By leveraging the scientific tools used for creating steering algorithms and simulation methods along with the powerful digital signal processing required to implement the FIR coefficients that result, and applying these to larger, higher output transducers driven by more powerful amplifier channels we can create high output steerable arrays.

In the case of the ICLive, Renkus-Heinz employs five each 6.5 inch woofers and three each 1 inch compression drivers in place of the eight each 4 inch co-axes used for the original IC8. Of note here, is that while real 1 inch compression drivers are used for the frequency range above 1.6kHz, the spacing between each driver is equivalent to that of the 4 inch co-axials used in the IC8. This keeps the onset frequency for gradient lobes relatively high, while the strength of any resultant gradient lobes is kept in check by virtue of the directional waveguides used for the high frequency array, Fig. 22-45. So, in addition to the new transducer arrangement, the amplifiers used for the ICLive have twice the capacity of the original

IC8. This combination of increased driver and amp horsepower enable each ICLive array to be roughly 9dB louder than an IC8 array, while maintaining a discrete columnar form-making it suitable for all but the loudest and largest applications, while its discrete, architecturally pleasing form factor allows it to be installed where it can be most effective, right in plain sight. Although the ICLive, from Renkus-Heinz, was one of the first to appear on the market, the ICLive is joined today by similar designs offered by RCF with their TTL11A and Fohhn with their Focus Modular steerable arrays.

Figure 22-45. Meyer Sound Cal array.

In addition to their ICLive, steerable line array, Renkus-Heinz has also introduced their IC2 (or IC Square) steerable point source array. This is a rather unique solution that further leverages the techniques used for beam steering but in a form factor maximizes power density for extremely high output capabilities. Where the ICLive drivers are arranged in a slender column to maximize array height and vertical low frequency pattern control, the IC2's four each 8 inch woofers are arranged symmetrically about the four each

1 inch compression drivers resulting in a compact, “square” enclosure which offers horizontal pattern control, unique for a product with digital steering capability. Perfectly suitable for use on its own, where its output is equivalent to a loud self-powered point source, the scalable IC2 can be combined in arrays of up to 16 units tall, for applications that require the highest output with precision directivity and steering combined. As of today, only the AXYS Target U-16 from Duran Audio, Fig. 22-46, offers similar levels of full range steerable, scalable performance, so expect more development in this segment, as popularity for this solution grows and becomes more common place.

Figure 22-46. Renkus-Heinz ICLive Dual with waveguide.

22.9 Passive Column Loudspeakers: Most of the Benefits of a Steerable Column Array, in a Simpler

Package

It is difficult to think of an instance where the directional control of a steerable column may not be beneficial. However, real world limitations such as available budget and or accessibility of nearby power, might preclude one from specifying or implementing a steerable column array in all applications. Fortunately the primary benefit of steerable arrays, high directionality at lower frequencies, can be achieved with simpler, all passive designs.

As we learned earlier, it is the height of the array that determines the lowest frequency at which a column maintains directionality. Whether each loudspeaker in the array is powered discretely, as they are in steerable arrays, or wired in series and parallel combinations and ultimately powered by a single amplifier channel, as they would in a passive array, if the two arrays have the same effective height they will offer the same directionality at lower frequencies. So a passive array can be designed to achieve good vertical control at lower frequencies, thereby minimizing its impact on reverberation in the venue in which it is installed. The main limitation here is that the passive column loudspeaker requires mechanical aiming, by pointing it down towards the audience, away from the wall. Architecturally this is not considered as elegant as a steerable array mounted directly up against a wall.

Despite being driven by a single amplifier channel, some passive columns use internal passive networks to provide uniformity of coverage across a wide frequency range. Some of the best examples of these types of columns include JBL's CBT or Constant Beamwidth Transducer arrays, Fig. 22-47, which are based on research done by Don B. Keele Jr. on applying sonar array concepts to the challenges of generating frequency invariant dispersion in passive column loudspeakers. Starting with an array of drivers on a baffle of fixed curvature, an amplitude shading scheme is applied to the transducers from the center outwards, where the drive to the outer drivers is reduced sequentially according to a Legandre function, Fig. 22-48.

Another successful passive column implementation is the Community Professional Loudspeaker's Entasys series of column loudspeakers. Designer Bruce Howze, started with a multi-way approach with driver spacing and driver type appropriate for the two or three pass bands chosen, Fig. 22-49. Unique to the Entasys solution is a continuous passive HF array that to a limited degree, can be curved (to increase or decrease vertical coverage) and aimed by the installer to direct sound at the audience, no software or computers are involved, Fig. 22-50.

While the audience coverage from a digitally steered array will be superior to that of a passive column, for rooms with lower reverberation times, shorter throw distances or smaller budgets, these new generation passive column arrays offer exceptional performance and sound quality.

Figure 22-47. JBL CBT-70J loudspeaker array.

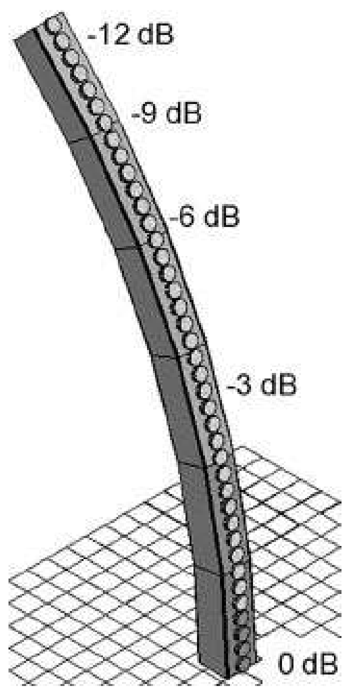


Figure 22-48. JBL Ground-Plane CBT Array with Legandre shading

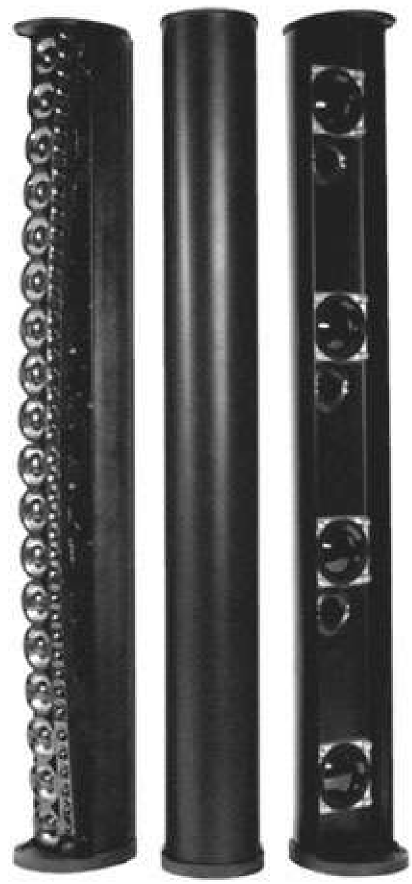


Figure 22-49. Community Entasys loudspeaker system.

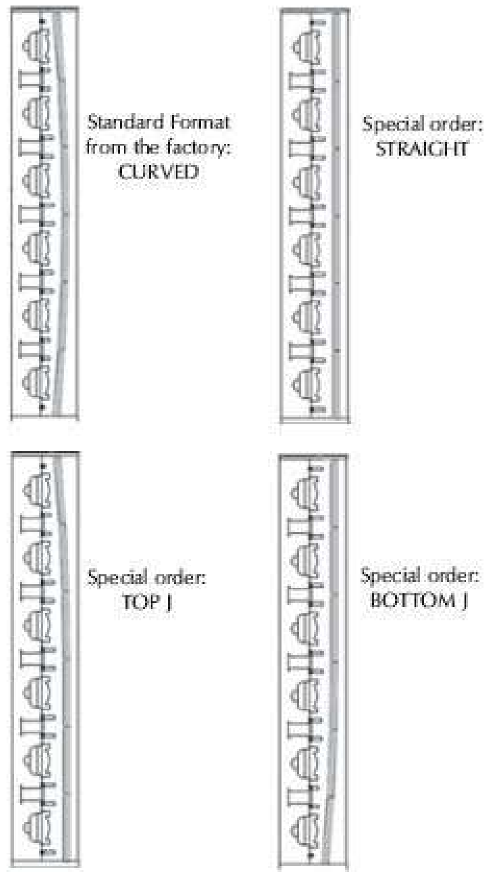


Figure 22-50. Community Entasys loudspeaker system curvature configurations.

Part 5

Electronic Audio Circuits and Equipment

Chapter 23

Circuit Protection and Power Supplies

by Ballou Glen

23.1 Watts versus Volt-Amperes

Watts and volt-amperes are units of measurement of electrical power. Watts is the measurement of "real power" and volt-amperes is the measurement of "apparent power."

Real power for dc circuits is calculated by using

$$W = V_{dc} \times I_{dc} \tag{23-1}$$

Real power for ac circuits is also voltage times current, however since the voltage and the current are not necessarily in phase, we need the instantaneous voltage with time, $v(t)$, and the instantaneous current with time, $i(t)$, to obtain instantaneous power with time, $p(t)$. To obtain the average power in watts over a period of time requires the equation

$$\begin{aligned} P_{ave} &= \frac{1}{T}\int_0^T (v(t) \times i(t))dt \\ &= \frac{1}{T}\int_0^T p(t)dt \end{aligned} \tag{23-2}$$

Apparent power in volt-amperes for dc circuits is the same as real power, i.e. $VA = W$.

Apparent power in volt-amperes for ac circuits is

$$VA = V_{rms} \times I_{rms} \quad (23\text{-}3)$$

Power factor, *PF*, is the ratio of watts to volt-amperes. In a dc circuit, the power factor is always 1. The power factor for ac circuits is always between 0 and 1 and is found with

$$PF = \frac{W}{VA} \quad (23\text{-}4)$$

If the load is purely resistive, the power factor would be 1, however, if the load is purely capacitive of inductive, the power factor would be 0. The circuit would draw 0 W but could draw substantial current resulting in substantial VA.

A wattmeter or a voltmeter and ammeter can be used to measure dc watts. A multimeter capable of measuring ac volts and ac amperes can be used to measure ac volt-amperes.

23.2 Circuit Protection

All electronic circuits require protection against transients, lightning, electrostatic discharge, and surges. Circuit protection is usually supplied in front of or inside of power supplies. This can be accomplished in several ways including lightning rods, fuses, circuit breakers, surge protection devices, and GFCIs.

23.2.1 Power Distribution Safety

Electricity kills! No matter how confident we are we must always be careful around electricity. Fibrillation is a nasty and relatively slow death so it is important that Defibrillators are accessible when working around electricity. Table 23-1 displays the small amounts of current that is required to hurt or kill a person.

Table 23-1. Physiological Effects of Shock Current on Humans (from Amundson)

Shock Current in mArms	Circuit Resistance at 120Vac	Physiological Effects
0.5–7mA	240,000Ω down to 17,000Ω	**Threshold of Perception:** Large enough to excite skin nerve endings for a tingling sensation. Average thresholds are 1.1mA for men and 0.7mA for women.
1–6mA	120,000Ω down to 20,000Ω	**Reaction Current:** Sometimes called the Surprise current. Usually an involuntary reaction causing the person to pull away from the contact.
6–22mA	20,000Ω down to 5400Ω	**Let-Go Current:** This is the threshold where the person can voluntarily withdraw from the shock current source. Nerves and muscles are vigorously stimulated, eventually resulting in pain and fatigue. Average let-go thresholds are 16mA for men and 10.5mA for women. Seek medical attention.
15mA and above	8000Ω and below	**Muscular Inhibition:** Respiratory paralysis, pain and fatigue through strong involuntary contractions of muscles and stimulation of nerves. Asphyxiation may occur if current is not interrupted.
60mA–5A	2000Ω down to 24Ω	**Ventricular Fibrillation:** Shock current large enough to desynchronize the normal electrical activity in the heart muscle. Effective pumping action ceases, even after shock cessation. Defibrillation (single pulse shock) is needed or death occurs.
1A and above	120Ω and below	**Myocardial Contraction:** The entire heart muscle contracts. Burns and tissue damage via heating may occur with prolonged exposure. Muscle detachment from bones

	possible. Heart may automatically restart after shock cessation.

Ground-Fault Interrupters

Ground-fault circuit interrupters (GFCIs) are sometimes called earth leakage or residual-current circuit breakers. GFCIs sense leakage current to earth ground from the hot or neutral leg and interrupt the circuit automatically within 25ms if the current exceeds 4 to 6mA. These values are determined to be the maximum safe levels before a human heart goes into ventricular fibrillation. GFCIs do not work when current passes from one line to the other line through a person, for instance. They do not work as a circuit breaker.

One type of GFCI is the core-balance protection device, Fig. 23-1. The hot and neutral power conductors pass through a toroidal (differential) current transformer. When everything is operating properly, the vector sum of the currents is zero. When the currents in the two legs are not equal, the toroidal transformer detects it, amplifies it, and trips an electromagnetic relay. The circuit can also be tested by depressing a test button which unbalances the circuit.

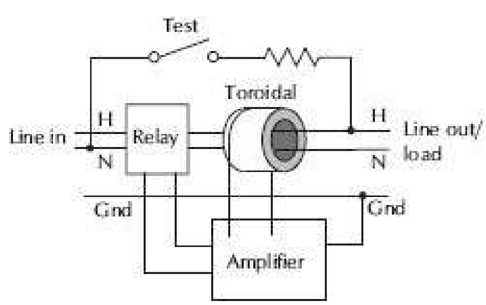


Figure 23-1. Typical ground-fault circuit interrupter.

23.2.2 Voltage Transients

Voltage Transients are defined as short duration surges of electrical energy and are the result of the sudden release of energy previously stored, or induced heavy inductive loads, or lightning. In electronic circuits, this energy can be released in a predictable manner via controlled switching actions, or randomly induced into a circuit from external sources.

Repeatable transients are frequently caused by the operation of motors, generators, or the switching of reactive circuit components. Random transients are often caused by lightning and electrostatic discharge (ESD) which generally occur unpredictably. The key characteristics of several transients are shown in Table 23-2.

The exponential rise time of lightning is in the range 1.2µs to 10µs (essentially 10% to 90%) and the duration is in the range of 50 µs to 1000 µs (50% of peak values).

Table 23-2. Various Transient Sources and Voltage, Current Rise Time, and Duration Magnitude

	Voltage	Current	Rise Time	Duration
Lightning	25kV	20kA	10 us	1 ms
Switching	600V	500A	50us	500ms
EMP	1kV	10A	20 ns	1 ms
ESD	15kV	30A	<1ns	100ns

ESD has a much shorter duration event. The rise time has been characterized at less than 1.0ns and the overall duration is approximately 100ns.

Microprocessors have structures and conductive paths which are unable to handle high currents from ESD transients as they operate at very low voltages, so voltage disturbances must be controlled to prevent device interruption and latent or catastrophic failures.

Table 23-3 shows the vulnerability of various component technologies in electronic circuits today.

Table 23-3. Range of Device Vulnerability

Device Type	**Vulnerability**-V
VMOS	30–1800
MOSFET	100–200
GaAs FET	100–300
JFET	140–7000
EPROM	100
CMOS	250–3000
Schottky Diodes	300–2000
Bipolar Transistors	380–7000
SCR	650–1000

23.2.3 Mechanical Protection

Mechanical protection, such as lightning rods, won't necessarily save equipment. A lightning strike can deliver hundreds of thousands of amperes of current, only a portion of which is handled by mechanical means. After passing through lightning rods and primary protection systems, stray currents in a building may be 3000A and 6000V. While such surges last for typically a few microseconds, the period is long enough to damage circuit boards.

Fuses, PTCs, and Circuit Breakers

Fuses and circuit breakers are used to prevent fires from overheated equipment. UL, CSA, etc. requires them.[1]

Fuses. A fuse is a piece of meltable metal that melts or vaporizes when the current running through it exceeds the fuse's rating. The fuse may be a fast blow fuse which, as its name implies, responds quickly to an overload, or a slow blow fuse which takes longer to melt when the overload is moderate allowing for start-up surges and temporary device overloads.

Fuses normally do not respond to currents less than 1.5 times their rated value, but can vary according to the standards (IEC or UL) with which the fuse is rated. IEC rated fuses carry 100% of their rated current continuously. UL rated fuses carry 75% of their rated current continuously. Standards for fuse-links are given in Table 23-4.

Table 23-4. Standards for Fuse-links

IEC 60127		**Miniature fuses (general title)**
IEC 60127-1	Part 1:	Definitions for miniature fuses and general requirements for miniature fuse-links
IEC 60127-2	Part 2:	Cartridge fuse-links
IEC 60127-3	Part 3:	Sub-miniature fuse-links
IEC 60127-4	Part 4:	Universal modular fuse-links
IEC 60127-5	Part 5:	Guidelines for quality assessment for miniature fuse-links
NF C 93435		Cartridge fuses with improved characteristics
UL 248-1		Low-voltage fuses: General requirements
UL 248-14		Low-voltage fuses: Supplemental fuses
CSA/C22.2 No.	248.1	Low-voltage fuses: General requirements
CSA/C22.2 No.	248.14	Low-voltage fuses: Supplemental fuses

Above the rated current fuses respond to I^2t, or the time it takes for a fuse to blow. It is related to the square of the current, which is a function of the power the fuse dissipates and is expressed as ampere square seconds (A^2s). Fuses dissipate power, albeit very little in comparison with the circuit they are protecting and have a breaking current rating which is the maximum short circuit current that the fuses are guaranteed to break. Fuses listed in accordance with UL/CSA/ANCE 248 are required to have an interrupting rating of 10,000A at 125 V. Fuses have a voltage rating that insures they can break the current in the circuit without creating a continuous arc. Common voltages are 32, 63, 125, 250, and 600 V. When specifying a fuse, check its rated current, the rating system, the delay (slow or fast), the breaking current, the breaking voltage, and the "I^2t" rating. Fuse sizes began with the early "Automobile Glass" fuses, thus the term "AG". Other fuses are made with the outer case made of ceramic or other materials so their designations are "AB". Table 23-5 gives the dimensions of popular fuses.

Table 23-5. Dimensions of Popular Fuses

Size	Diameter–inches		Length–inches	
1AG	1/4	0.250	5/8	0.625
2AG		0.177		0.588
3AG	1/4	0.250	1¼	1.25
4AG	9/32	0.281	1¼	1.25
5AG	13/32	0.406	1½	1.5
7AG	1/4	0.250	7/8	0.875
8AG	1/4	0.250	1	1.0

Fig. 23-2 shows the average time current curves for Littelfuse 3AG Slo-Blo and Fast-Acting fuses.

PTCs. PTCs (positive temperature coefficient) are typically used in a wide variety of telecom, computer, consumer electronics, battery and medical electronics applications where overcurrent events are common and automatic resettability desired.

Polymeric PTC devices react to heat generated by the excessive current flow in a circuit. A PTC limits current flow as it rises in temperature, changing from a low to a high resistance state which is called “tripping.” Fig. 23-3 shows the typical response of a PTC to temperature. The time-current curve for a PTC is about the same as a Slo-Blo fuse.

Polymer PTCs are made of high density polyethylene mixed with graphite. During an overcurrent event, a Polymer PTC will heat and expand, which in turn causes the conducting particles to break contact and stop the current. The general procedure for resetting the device after an overload has occurred is to remove power and allow the device to cool down.

Littelfuse PTCs come with the following forms and features:

1. Surface Mount Devices:
   - A full range of compact footprints.
   - Low hold current.
   - Very fast trip time.
   - Low resistance.

2. Radial Leaded Series:
   - Protection devices up to 600Vdc.
   - Very high hold current.
   - Low trip-to-hold current ratio.,
   - Low resistance.

3. Battery Strap Devices:
   - A narrow low profile design.
   - A weldable band Nickel terminal.
   - Low resistance—for extended battery run time.

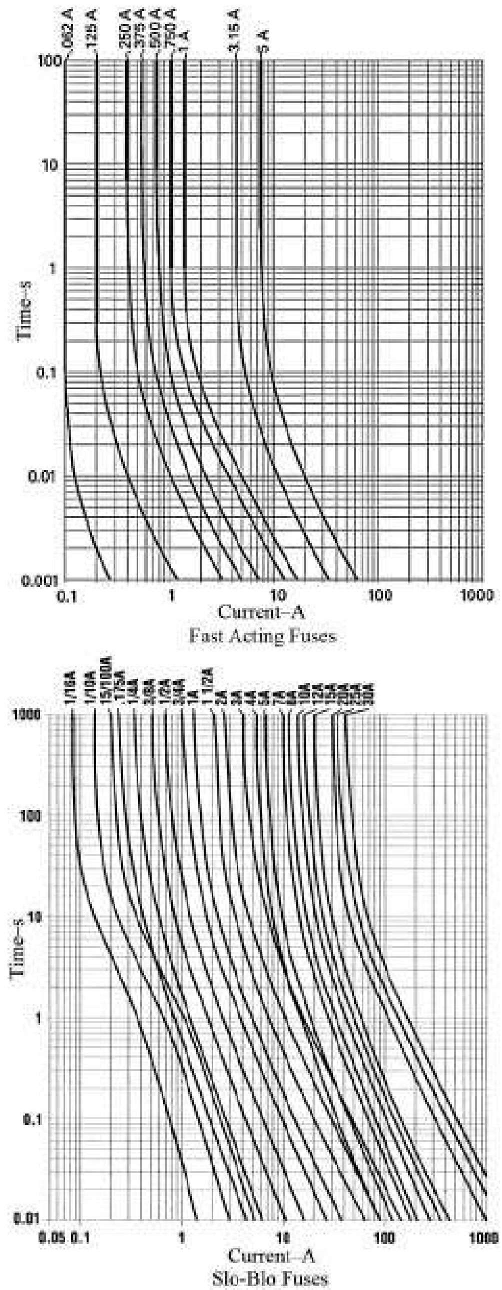


Figure 23-2. Average time current curves for Fast Acting and Slo-Blo fuses. Courtesy Littelfuse, Inc.

The most obvious difference between fuses and PTCs is that PTCs are automatically resettable whereas fuses need to be replaced after they are tripped. A fuse will completely stop the flow of current in an overcurrent event, PTCs continue to enable the equipment to function, except in extreme cases.

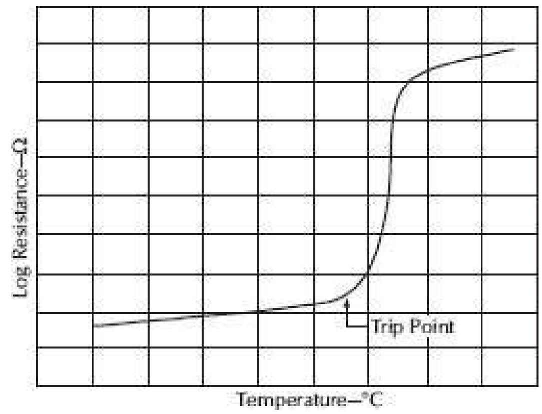


Figure 23-3. Typical response of a PTC resistance to temperature. Courtesy Littelfuse, Inc.

Because they reset automatically, PTCs are useful where overcurrent events are expected to occur often, and constant system uptime, and/or user transparency are at a premium. They are also often chosen where remote locations make fuse replacement difficult.

Circuit breakers. Circuit breakers operate very much like fuses. They have continuous current ratings and trip ratings and are available in thermal trip and magnetic trip types or a combination of the two. Typically, they require to be reset manually.

Thermal circuit breakers for equipment, CBEs, simulate the electrothermal behavior of the protected components (conductors in wiring, motors, transformers, etc.) by a thermo-bimetal. The mechanical element simulates the heating effect of the current, transforming electric energy into a motion (deflection) and triggers a mechanism to interrupt the current causing the effects. Using the heat created by the current instead of the magnitude of the current itself offers a great advantage because heat determines the admissible stress of the insulation and the admissible duration of the various overload conditions encountered in practical applications. Bimetals can handle frequencies from dc to 400Hz, without necessitating any change in ratings or characteristics.

Thermally operated CBEs, allow the surplus energy required for start-up or high-torque operation of motors, and they allow high inrush spikes which occur in switching power supplies, transformers, tungsten filament lamps, etc., avoiding nuisance tripping due to such transients. Thermally operated CBEs are temperature sensitive which is an advantage because the component to be protected is almost always temperature sensitive.

Their delay time can be influenced in several ways and may be achieved by using a different method of heating the bimetal. The most widely used method is the direct heating of a bimetal strip by the internal losses produced by the current passing through the bimetal). When the losses are insufficient to produce enough heat to cause sufficient deflection, a heater winding is wrapped around the bimetal strip to obtain the required heat. Since the heat has to pass through an insulation before it reaches the bimetal, a time lag will occur and a delayed action will result.

The typical tripping zone of thermal CBEs changes with ambient temperature in a similar way as PVC insulated wire does.

Thermal magnetic CBEs have two releases to achieve automatic interruption of an overcurrent, a thermo-bimetal for overload current, and an electro magnet for short circuit or extremely high current.

The operating characteristic is essentially composed of two zones, linked by a third zone where either one or the other mode of tripping will be effective.

The electro magnet should be dimensioned so that it will not trip during transients likely to occur in the intended application. This determines the level of the current below which instantaneous tripping should not occur. The upper level, indicating the current above which instantaneous tripping must occur, is of interest in considerations concerning the selective action of two protective devices.

In the short circuit range of overcurrents (above 8–12 times the rated current), the faster interruption obtainable with the magnetic release is an advantage. It can save the heater windings of indirectly heated bimetals from overheating and it can improve the breaking capacity of the CBE. CBEs primarily intended for overload protection are usually capable of interrupting, without back-up assistance, currents up to 100 to 300A without destroying the CBE. The performance at higher fault levels usually relies on back-up assistance by fuses or breakers.

23.2.4 Electronic Protection Methods

For circuit protection, the first line of resistance is resistance. Resistance limits current flow and, in conjunction with capacitance, slows the rate of increase of voltage, giving other protection measures in the circuit time to act. Resistance has its drawbacks. If the current through the resistance is significant, power is wasted and power waste generates heat. Resistance also causes voltage drops which can adversely affect circuit behavior.

23.2.4.1 Inrush Current Protection

Thermistors. Using negative temperature coefficient thermistors

(NTC) is simple way to control high inrush current. Inrush current is affected by:

1. Energy of the inrush current.
2. Minimum resistance required by the NTC thermistor at normal operating temperature.
3. Steady state current.
4. Ambient temperature.

The energy of the inrush current caused by the input capacitors is:

$$E = \frac{C \times V_{peak}^{2}}{2} \qquad (23\text{-}5)$$

where,
E is the energy of the inrush current in J,
C is the capacitance in F,

$$V_{peak} = (V_{input} + 10\% V_{input})\sqrt{2}\,.$$

The minimum NTC thermistor resistance at 25°C to control inrush current is:

$$R_{min} = \frac{V_{peak}}{I_{limit}} \qquad (23\text{-}6)$$

where,
R_{min} is the minimum resistance required,
V_{peak} is the resultant peak voltage,
I_{limit} is the maximum desired inrush current.

23.2.4.2 Overvoltage Protection

The increasing use of integrated circuits has resulted in a requirement to reduce the lightning overvoltage hazard. Integrated circuits, especially large-scale integrated circuits, have very low insulation strength so they are vulnerable to lightning overvoltage transients.[2]

Lightning surges frequently get into electronic systems through ac power supplies, therefore it is good to install the surge protective devices (SPDs) on the supply lines to protect the electronic equipment. According to IEC standards, it should include common and differential protective modes.

A traditional SPD circuit with the both protective modes is shown in Fig. 23-4. Metal oxide varistors M_1 through M_6 and gas discharge tubes G_1 and G_2 are used in the two stages. In the first stage, M_1 provides overvoltage limiting for differential mode, and M_2-G_1 and M_3-G_1 provides for common mode. M_4, M_5-G_2 and M_6-G_2 do the same in the second stage. L_1 and L_2 are decoupling inductances which are used to coordinate the protective characteristic between the two stages. This circuit complies with the IEC standards, and is widely used on single-phase ac supply lines.

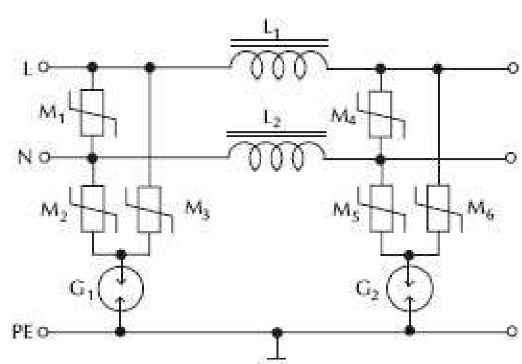


Figure 23-4. Traditional SPD circuit.

23.2.4.3 Power Protection

Varistors offer surge handling capabilities for applications involving high currents and high voltages. Polymer-based electrostatic discharge (ESD) suppressors are used on sensor lines and input/output of measurement equipment as they offer low capacitance, and are unlikely to distort the sensor signal that is being monitored by the equipment. Silicon-based devices or arrays offer speed and can clamp down almost instantaneously on surges. They cannot handle large currents and voltages like other suppression devices, such as varistors.

For input ac power protection (surge protection) ac rated TVS diodes and Metal Oxide Varistors (MOV) are used. These are relatively high capacitance devices, but in 50 or 60Hz circuits it is not significant. They are rated for working voltage, breakdown voltage, and surge current. It is important to look at the peak voltage in the circuit and not the rms voltage. Table 23-6 shows the size, surge current, and energy ratings for various MOVs.

23.2.4.4 Electrostatic Discharge (ESD) Protection

One of the most common electrostatic discharge effects can be seen in environments with low humidity when you rub your shoes on the carpet and touch for a metal door knob. The spark that jumps to the door knob from the voltage stored in your body’s capacitance is ESD and it can be thousands of volts. Rotating machinery can produce ESD, and it can be more severe because generally capacitance is higher and resistance is lower. To protect circuits from ESD, an input resistor in conjunction with a capacitor, is used to slow the rate of rise of the voltage.[2]

Table 23-6. Surge Current and Energy Ratings of Various Size MOVs at Typical Voltage Ratings

Nominal Diameter	Voltage Rating (Vac)	Maximum Surge Current (8 × 20μs, 1 pulse)	Maximum Energy (2mS)
14mm	130	6,000A	50J
20mm	130	1,0000A	100J
32mm	130	25,000A	200J
34mm	130	30,000A	270J
40mm	130	30,000A	270J
14mm	275	6,000A	110J
20mm	275	10,000A	190J
32mm	275	25,000A	360J
34mm	275	40,000A	400J
40mm	275	40,000A	400J

Zener diodes can also be used where the rate of rise of the voltage transient is not too high. For inductive kickback from a dc powered inductor or coil, a high speed diode of the proper speed, current, and voltage rating is effective. For applications where the coil or inductor is de-energized infrequently, the most important diode rating is the surge current rating. Transient voltage suppression (TVS) diodes (tranzorbs), are a type of zener diode that respond in nanoseconds to overvoltage. They are rated for operating voltage, breakdown voltage, and surge current.

23.2.4.5 Uninterruptible Power Supply

The Uninterruptible power supply (UPS) is probably the most effective way to process and filter the ac lines feeding electronic power supplies and circuits. Table 23-7 shows the effects of various power problems and their effect on off-line, line-interactive, and on-line UPSs.

Table 23-7. Effects of Various Power Problems on Three Types of UPSs

Power	Description	Effect	Will the UPS solve?
Temporary Interruption	Planned or accidental total loss of utility power in a localized area of the community. Seconds to minutes.	Equipment shutdown, loss of work and data, file and hard disk and operating system (OS) corruption, loss of fiber optic, Ti and ISDN connections.	Off-line–Yes Line-interactive–Yes On-line–Yes
Long-Term Interruption	Planned or accidental total loss of utility power in a localized area of the community. Minutes to hours.	Equipment shutdown, loss of work and data, file and hard disk and OS corruption. Loss of fiber optic. Ti and ISDN connections.	Off-line–No Line-interactive–95% No On-line–Yes
Momentary Interruption	Very short planned or accidental power loss. Milliseconds to seconds.	Computer hangs, computer and network equipment reboots or hangs, loss of work and data, file and hard disk and OS corruption.	Off-line–Maybe Line-interactive– Maybe On-line–Yes
Sag or Under-Voltage	A decrease in utility voltage. Sags–Milliseconds to a few seconds. Under-voltage–Longer than a few seconds.	Shrinking display screens, equipment hang or reset, equipment power supply damage. Computer hangs, computer and network equipment reboots or hangs, toss of work and data, file and hard disk and OS corruption.	Off-line–No Line-interactive–Yes On-line–Yes
Swell or Over-Voltage	An increase in utility voltage. Swell–Milliseconds to a few seconds. Over-voltage–Longer than a few seconds.	Permanent equipment damage. Computer hangs, computer and network equipment reboots or hangs, loss of work and data, file and hard disk and OS corruption.	Off-line–No Line-interactive–Yes On-line–Yes
Transient, Impulse, or Spike	A sudden change in voltage up to several hundred to thousands of volts. Microseconds.	Network Errors, Burned or damaged equipment and circuitry. Computer hangs, computer and network equipment reboots or hangs, loss of work and data, file and hard disk and OS corruption.	Off-line–Yes Line-interactive–Yes On-line–Yes, Higher level of protection
Notch	A disturbance of opposite polarity from the waveform. Microseconds.	Slow LAN due to excessive errors, audible noise in telephone and audio equipment.	Off-line–No Line-interactive–No On-line–Yes
Noise	An unwanted electrical signal of high frequency from other equipment. Sporadic.	Slow LAN due to excessive errors, audible noise in telephone and audio equipment Equipment hangs.	Off-line–No Line-interactive–No On-line–Yes
Harmonic Distortion	An alteration of the pure sine wave (sine wave distortion), due to non-linear loads such as computer switching power supplies.	Causes motors, transformers and wiring to overheat, lowers operating efficiency of office equipment.	Off-line–No Line-interactive–No On-line–Yes

Electrical power supply interference can come in a range of forms, such as voltage dips and surges, harmonics, or voltage spikes. These disruptions can cause serious harm to sensitive electrical equipment. To reduce the risk of power supply distortion, UPS systems are often incorporated in electrical networks. A UPS system can be a helpful tool for ensuring proper power supply performance.

UPS systems can provide backup power to keep systems operational in case the power supply is cut off where power fluctuations or outages frequently occur. Under conditions involving short fluctuations or voltage disruption, the UPS can maintain constant power to keep loads running, and if electrical failure occurs, it activates reserve power to keep systems operating until they can be safely shut down. UPS systems can also reduce the risk posed by harmonic disruptions and line transients. An effective UPS includes several of the following features:

1. Regulated output voltage with low harmonic distortion unaffected by input voltage or load changes.
2. An input current with reduced harmonic distortion.
3. A low degree of electromagnetic interference and acoustic noise.
4. Minimal transition times between normal and backup operations.
5. High levels of reliability and efficiency.
6. Relatively low cost, weight, and size.

Most UPS systems do not provide all of these features simultaneously, but it is usually possible to find a UPS with the characteristics suited to an applications needs. There are several types of UPSs as defined in the following:

Standby UPS Systems. Standby UPS, also known as off-line or line-preferred UPS, typically consists of an ac/dc and a dc/ac inverter, a battery, a static switch, a low pass filter to reduce switching frequency from the output voltage, and a surge suppressor. Without a power failure, power is fed through surge and noise suppression circuitry to the load. An ac/dc inverter/battery charger keeps an internal battery charged and ready for use. During a blackout, brownout or overvoltage, the inverter converts battery power into a simulated sine wave output. When power returns, the UPS switches back to ac power and the battery is recharged. The standby UPS system operates with the ac input as a primary power source, and the battery and inverter as backup sources in case of primary power failure. To save power, the inverter normally remains on standby, and activates when the power fails. Sensing of a low voltage situation and switching to battery power happens so quickly that equipment continues to operate. A standby UPS system is highly efficient, has a small footprint, and a low cost, makes it a common option for personal computing.

Standby-Ferro UPS. The standby-Ferro UPS has a saturating transformer with multiple power connections. The primary power flows from the ac input, moving through the transformer and on to the output. During a power failure, the transfer switch activates the inverter to supply the output load. The specialized Ferro transformer can provide some degree of voltage regulation and control over the output waveform. Standby-Ferro systems are useful for their reliability and line filtering characteristics, however, it carries the risk of voltage distortion and overheating.

Line Interactive UPS. In a line-interactive UPS system power is fed through surge and noise suppression circuitry to built-in line conditioning circuitry which regulates high or low voltages back to normal levels and sends clean power to the load, without using battery power. The line conditioning circuitry is usually a tap-changing transformer allowing it to provide voltage regulation so the UPS does not switch to battery power prematurely. A battery charger keeps an internal battery charged and ready for use. During a blackout, an inverter switches on and converts battery power into a simulated sine wave output. When power returns, the inverter switches off and the battery is recharged. Switching happens within a few milliseconds, so equipment is almost always unaffected. Because the inverter is continuously connected to the output, the UPS provides additional filtering and reduces switching transients. The line interactive design UPS has high efficiency and reliability, small size, and low cost.

Double Conversion UPS. A double conversion UPS provides very good protection from outages and power quality problems. Instead of switching from main power to backup (battery) power as needed, it converts all ac power to dc. Some of the dc power charges the battery, and the rest is converted back to ac to power the connected equipment. This essentially prevents any spikes, transients, etc. from reaching the equipment, yielding high protection level. This system is expensive and also has decreased operating efficiency because of the conversion of ac to dc and then back to ac during normal operation. The units operate at higher temperatures, requiring more cooling.

Delta Conversion UPS. Delta conversion UPS was introduced to alleviate some of the disadvantages presented by double conversion systems. Like the double conversion UPS, the delta conversion UPS inverter continuously supplies load voltage, plus it delivers power to the inverter output. Under power failure or electrical distortions, the UPS acts similarly to a double conversion unit, but has better efficient energy performance by converting power from input to output rather than cycling between power and battery sources.

On-Line UPS Systems. In an on-line UPS system power is first broken down and then perfectly reconstructed by the inverter, which is “on-line” 100% of the time. There is no transfer switching time. This system completely eliminates incoming surge and line noise, adjusts high or low voltages, and produces perfect sine wave power.

Flywheel UPS. An alternative energy storage approach is the flywheel which is a rotating wheel that stores energy in the form of motion (angular momentum). When an outage occurs, the flywheel’s energy of motion is converted back to electrical energy to supply the equipment. The flywheel slows during this process as more energy is removed. Some flywheels are heavy and rotate slowly. Some are lighter and run at much higher speeds. Flywheels generally do not store as much energy as a battery, however they do offer some advantages. http://Processor.com (“UPS Flywheel Technology”) notes that the flywheel offers “superior performance without the high cost of ownership and the environmental impacts that lead batteries present.” Furthermore, its “rapid recharging and broad operating temperature range allow it to be used where batteries cannot operate. The footprint of flywheels is also much smaller and lighter than a battery’s footprint.”

23.3 Direct Current Power Supplies

Direct current (dc) power supplies come in various configurations including simple unregulated, complex regulated, and batteries.

23.3.1 Power Supply Terminology

Power Supply. A device that supplies electrical power to another unit. Power supplies obtain their prime power from the ac power line or from special power systems such as motor generators, inverters, and converters.

Rectifier. A device that passes current in only one direction. The rectifier consists of a *positive anode* and a *negative cathode*. When a positive voltage is applied to the anode of the rectifier, that voltage minus the voltage across the rectifier will appear on the cathode and current will flow. When a negative voltage is applied to the anode with respect to the cathode, the rectifier is turned off and only the rectifier leakage current will flow.

Forward Resistance. The resistance of an individual cell measured at a specified forward voltage drop or current.

Forward Voltage Drop. The internal voltage drop of a rectifier resulting from the current flow through the cell in the forward direction. The forward voltage drop is usually between 0.3Vdc and 1.25Vdc.

Reverse Resistance. The resistance of the rectifier measured at a specified reverse voltage or current. Reverse resistance is in megohms (MΩ).

Reverse Current. The current flow in the reverse direction, usually in microamperes (μA).

Maximum Peak Current. The highest instantaneous anode current a rectifier can safely carry recurrently in the direction of the normal current flow.

The value of the peak current is determined by the constants of the filter sections. With a choke filter input, the peak current is less than the load current. With a large capacitor filter input, the peak current may be many times the load current. The current is measured with a peak-indicating meter or oscilloscope.

Maximum Peak Inverse Voltage. The maximum instantaneous voltage that the rectifier can withstand in the direction opposite to which it is designed to pass current. Referring to Fig. 23-5, when anode A of a full-wave rectifier is positive, current flows from A to C, but not from B to C because B is negative. At the instant anode A is positive, the cathodes C of A and B are positive with respect to anode B. The voltage between the positive cathode and the negative anode B is inversely related to the voltage causing the current flow. The peak value of this voltage is limited by the resistance and nature of the path between the anode B and the cathode C. The maximum value of voltage between these points, at which there is no danger of breakdown, is termed *maximum peak inverse voltage*.

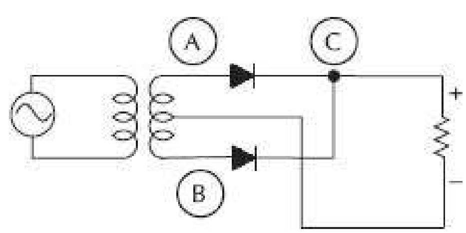


Figure 23-5. Peak inverse voltage analysis.

The relationship between peak inverse voltage, rms value of ac input voltage and dc output voltage depends largely on the individual characteristics of the rectifier circuit. Line surges, or any other transient or waveform distortion, may raise the actual peak voltage to a value higher than that calculated for a sine-wave voltage. The actual inverse voltage (and not the calculated value) should be such as not to exceed the rated maximum peak inverse voltage for a given rectifier. A peak-reading meter or oscilloscope is useful in determining the actual peak inverse voltage.

The peak inverse voltage is approximately 1.4 times the rms value of the anode voltage for single-phase, full-wave circuits with a sine-wave input and no capacitance at the input of the filter section. For a single half-wave circuit, with a capacitor input to the filter section, the peak inverse voltage may reach 2.8 times the rms value of the anode voltage.

Ripple Voltage. The alternating component (ac) riding on the dc output voltage of a rectifier-type power supply. The frequency of the ripple voltage will depend on the line frequency and the configuration of the rectifier. The effectiveness of the filter system is a function of the load current and the values of the filter components.

The ripple factor is the measure of quality of a power supply. It is the ratio of the rms value of the ac component of the output voltage to the dc component of the output voltage or

$$ripple\ factor = \frac{V_{rms}}{V_{dc}} \qquad (23\text{-}7)$$

where,
V_{rms} is the alternating current voltage at the output terminals,
V_{dc} is the direct current output voltage at the output terminals.

Internal Output Impedance. The impedance presented to the equipment receiving the power supply voltage. In the operation of many devices, it is necessary that the internal power supply impedance be as near to zero as possible.

Static Line Regulation. The variation in output voltage as the input voltage is varied slowly from rated minimum to rated maximum with the load current held at the nominal value.

Dynamic Load Regulation. The variation in output when the load change is sudden. The power supply may be unable to respond instantaneously, and an additional momentary excursion in the output voltage may result, subsiding afterward to the static load regulation level. The positive and negative excursion limits are superimposed on the static line and load regulation region. The positive and negative components are not necessarily equal or symmetrical. The most stringent rating is for a change from no load to full load or vice versa.

Dynamic Line Regulation. The momentary additional excursion of output voltage as a result of a rapid change in input voltage.

Thermal Regulation. Variations in the output voltage over the rated operating temperature range due to ambient temperature variations influencing various components of the power supply. This is also known as *thermal drift*.

23.3.2 Simple dc Power Supplies

The simplest type of dc power supply is a rectifier in series with the load. As more rectifiers are installed into the circuit, along with filters, the power supply becomes more sophisticated. The rectifier in series with the load supply will always remain simple and have poor regulation and transient response. Table 23-8 shows various power supplies and their characteristics. To determine the value of the parameter in the left column, multiply the factor shown in any of the center columns by the value in the right column.

23.3.2.1 One-Half Wave Supplies

A one-half wave unit can be connected directly off the ac mains, Fig. 23-6A, or off the mains through a transformer, Fig. 23-6B. Since a rectifier only passes a current when the anode is more positive than the cathode, the output waveform will be one-half of a sine wave, Fig. 23-6C. The dc voltage output will be 0.45 of the ac voltage input, and the rectifier current will be the full dc current; the peak inverse voltage (*piv*) across the rectifier will be 1.414Vac, and the ripple will be 121%. In the transformerless power supply, the 115Vac power line is connected directly to the rectifier system. This type of power supply is dangerous to both operating personnel and to grounded equipment. Also, power supplies of this type will cause hum problems that can only be solved by the use of an isolating transformer between the line and power supply.

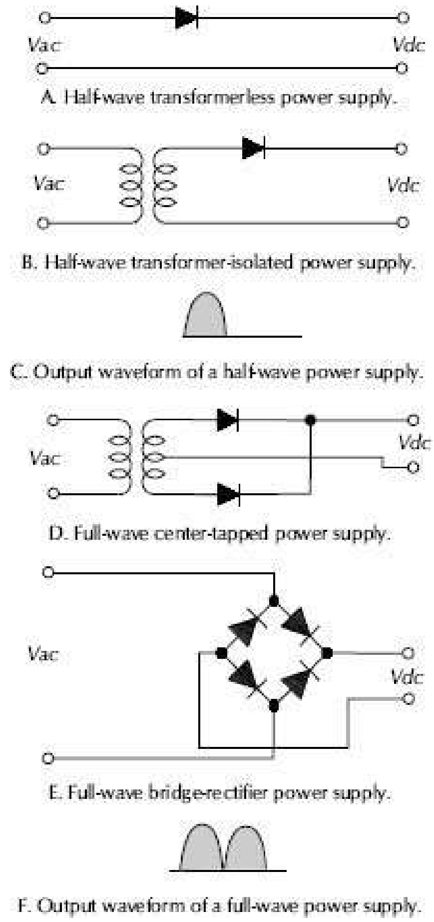


Figure 23-6. Two one-half wave and two full-wave power supplies.

23.3.2.2 Full-Wave Supplies

The full-wave supply is normally used in electronic circuits because it is simple and has a good ripple factor and voltage output. A full-wave supply is always used with a transformer. Full-wave supplies may be either a single-phase center tap design or a full-wave bridge. In either case, both the positive and negative cycles are rectified and mixed to produce the dc output.

The center tap configuration, Fig. 23-6D, uses two rectifiers and a center-tapped transformer. The Vdc is approximately equal to Vac where Vac is from one side of the transformer to the center tap.

Because the output is from each half wave, ripple is 48% of the output voltage and at twice the input frequency. Each rectifier carries one-half of the load current. The piv/rectifier is 2.828Vac.

A full-wave bridge rectifier supplies full-wave rectification without a center tap on the transformer. The bridge rectifier is not a true single-ended circuit, since it has no terminal common to both the input and output circuits.

A full-wave bridge rectifier consists of four rectifier elements, as shown in Fig. 23-6E. This circuit is the most familiar and is the type most commonly employed in the electronics industry.

With the full-wave bridge circuit, the dc output voltage is equal to 0.9 of the rms value of the ac input voltage.

Full-wave bridge rectifier circuits may be grounded by three methods shown in Fig. 23-7. In Fig. 23-7A and B, either the input (ac source) or output (dc load) may be grounded, but not both simultaneously. If an isolation transformer is used between the ac source and the input to the rectifier, as shown in Fig. 23-7C, both ac and dc sides may be grounded permanently. A method of grounding a bridge rectifier is shown in Fig. 23-7D where the center tap of an isolation transformer is grounded.

When designing rectifier circuits, dc load current, dc load voltage, peak inverse voltage, maximum ambient temperature, cooling requirements, and overload current must be analyzed. For example, assume a full-wave rectifier using silicon rectifiers is to be designed as in Fig. 23-7D and the dc load voltage Vdc under load is 25 V at 1 A.

Using Table 23-8, determine the current per rectifier using the equation

$$\begin{aligned} I_{rect} &= 0.5 \times I_{dc} \\ &= 0.5 \times 1 \\ &= 0.5\,\text{A} \end{aligned} \tag{23-8}$$

where,

I_{rect} is the current per rectifier,

0.5 is a constant from Table 23-8,

I_{dc} is the rectified ac current, which is the dc current.

This is the current each rectifier must carry. Next, the ac voltage required from the transformer is determined by the equation

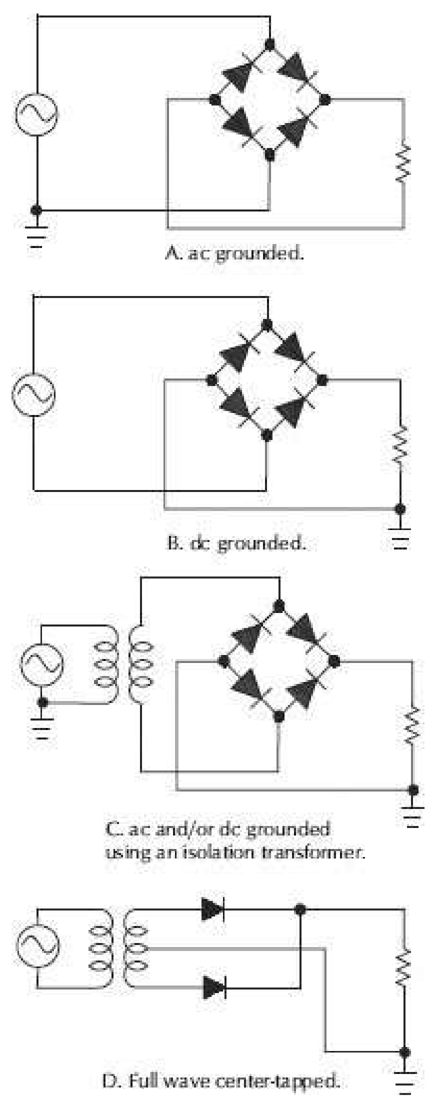


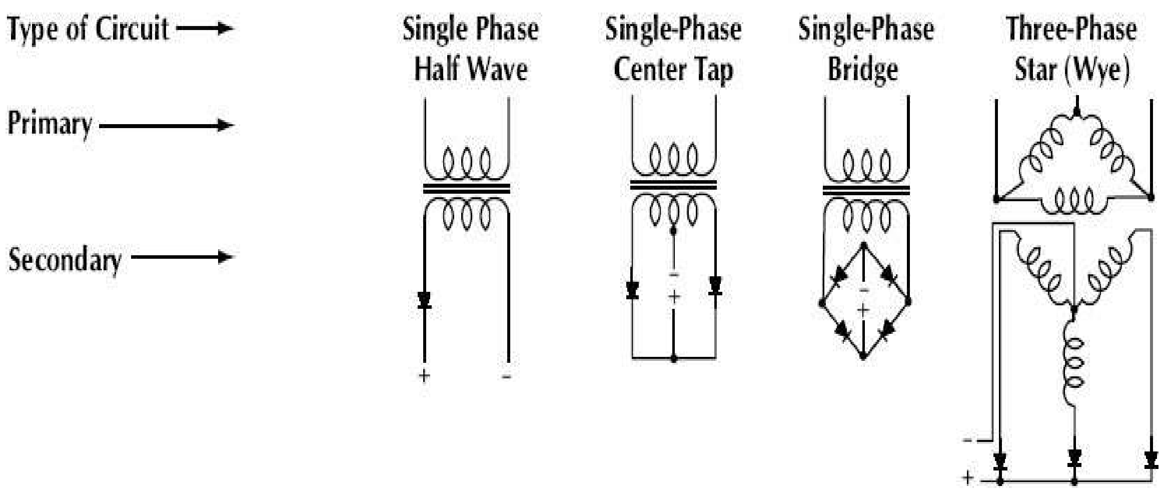
Figure 23-7. Methods of grounding a power supply.

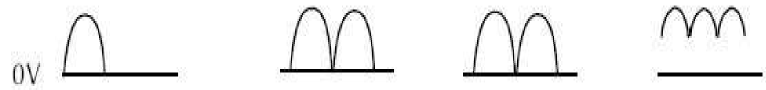
$$
\begin{aligned}
V_{ac} &= 1.11 \times V_{dc} \\
&= 1.11 \times 25 \qquad (23\text{-}9) \\
&= 27.75 \text{ Vrms}
\end{aligned}
$$

where,
V_{ac} is the transformer voltage,
1.11 is a constant from Table 23-8.

This is the voltage as measured from each side of the transformer center tap; the total voltage across the secondary is 55.50Vrms.

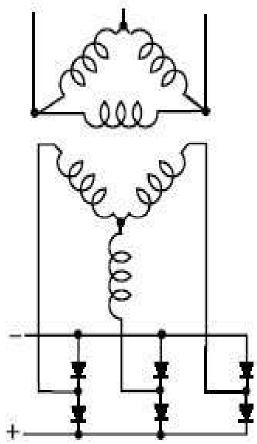
Table 23-8. Rectifier Circuit Chart



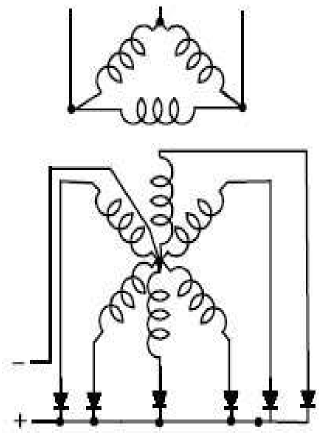


Number of rectifier elements	=	1	2	4	3
Rms dc volts output	=	1.67	1.11	1.11	1.02
Peak dc volts output	=	3.14	1.57	1.57	1.21
Peak reverse volts per rectifier element	=	3.14	3.14	1.57	2.09
	=	1.41	2.82	1.41	2.45
	=	1.41	1.41	1.41	1.41
Average dc output current	=	1.00	1.00	1.00	1.00
Average dc output current per rectifier element	=	1.00	0.500	0.500	0.333
Rms current per rectifier element: Resistive load	=	1.57	0.785	0.785	0.587
Inductive load	=	-	0.707	0.707	0.578
Peak current per rectifier element: Resistive load	=	3.14	1.57	1.57	1.21
Inductive load	=	-	1.00	1.00	1.00
Ratio of peak to average current per element: Resistive load	=	3.14	3.14	3.14	3.63
Inductive load	=	-	2.00	2.00	3.00
% Ripple (rms of ripple/average output voltage)	=	121%	48%	48%	18.3%
Ripple frequency	=	1	2	2	3
		Resistive Load	Inductive Load or Large Choke Input Filter		
Transformer secondary rms volts per leg	=	2.22	1.11 (to center tap)	1.11 (total)	0.855 (to neutral)
Transformer secondary rms volts line-to-line	=	2.22	2.22	1.11	1.48
Secondary line current	=	1.57	0.707	1.00	0.578
Transformer secondary volt-amperes	=	3.49	1.57	1.11	1.48
Transformer primary rms amperes per leg	=	1.21	1.00	1.00	0.471
Transformer primary volt-amperes	=	2.69	1.11	1.11	1.21
Average of primary and secondary volt-amperes	=	3.09	1.34	1.11	1.35
Primary line current	=	1.21	1.00	1.00	0.817
Line power factor	=	-	0.900	0.900	0.826

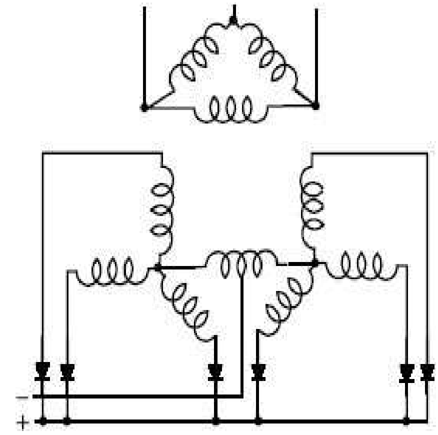
Three-Phase
Bridge



Six-Phase Stat
(Three Phase Diametric)



Three-Phase Double Wye with
Interphase Transformer



6	6	6	
1.00	1.00	1.00	× average dc voltage output
1.05	1.05	1.05	× average dc voltage output
1.05	2.09	2.42	× average dc voltage output
2.45	2.83	2.83	× secondary rms volts per transformer leg
1.41	1.41	1.41 (diametric)	× secondary rms volts line-to-line
1.00	1.00	1.00	× average dc output current
0.333	0.167	0.167	× average dc output current
0.579	0.409	0.293	× average dc output current
0.578	0.408	0.289	× average dc output current
1.05	1.05	0.525	× average dc output current
1.00	1.00	0.500	× average dc output current
3.15	6.30	3.15	
3.00	6.00	3.00	
4.2%	4.2%	4.2%	
6	6	6	× line frequency, *f*
Inductive Load or Large Choke Input Filter			
0.428 (to neutral)	0.740 (to neutral)	0.855 (to neutral)	× average dc voltage output
0.740	1.48 (max)	1.71 (max-no load)	× average dc voltage output
0.816	0.408	0.289	× average dc output current
1.05	1.81	1.48	× dc watts output
0.816	0.577	0.408	× average dc output current
1.05	.28	1.05	× dc watts output
1.05	1.55	1.26	× dc watts output
1.41	0.817	0.707	× (average load current × secondary leg voltage)/primary line V
0.955	0.955	0.955	

The peak inverse voltage is

$$\begin{aligned} piv &= 2.82 \times V_{ac} \\ &= 2.82 \times 27.75 \qquad (23\text{-}10) \\ &= 78.4 \text{ Vrms} \end{aligned}$$

where,
V_{ac} is the secondary ac voltage per leg,
2.82 is found in Table 23-8.

If a rectifier with the required *piv* rating is not available, two or more may be connected in series to obtain the desired *piv* rating. Unequal values of *piv* ratings may be used, provided the lowest rating is greater than half of the total piv rating needed.

Parallel operation of rectifiers can be used obtain higher current ratings. However, because of a possible imbalance between the units due to the forward voltage drop and effective series resistance, one unit may carry more current than the other and could conceivably fail. To prevent this, small equal value resistors must be connected in series with each individual rectifier to balance the load currents, as shown in Fig. 23-8.

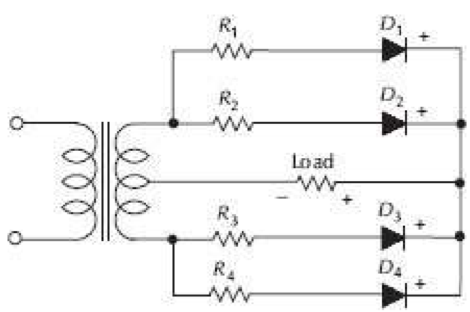


Figure 23-8. Small resistors connected in series with each rectifier to balance the current through each unit of parallel-connected rectifiers.

An interesting approach to a full wave bridge is replacing the rectifiers with low loss n-channel MOSFETs and a controller, Fig. 23-9.

Linear Technology's LT4320 ideal-diode bridge controller reads the incoming ac voltage waveform and smoothly turns on the appropriate pair of MOSFETs for each half cycle. The gate drive for the MOSFETs comes from an internal charge pump which includes the charging capacitors.[3]

Replacing bridge rectifier diodes with MOSFETs eliminates the silicon diode's 0.6V forward voltage drop and associated thermal effects. MOSFET forward voltage losses are much lower, often by a factor of 10.

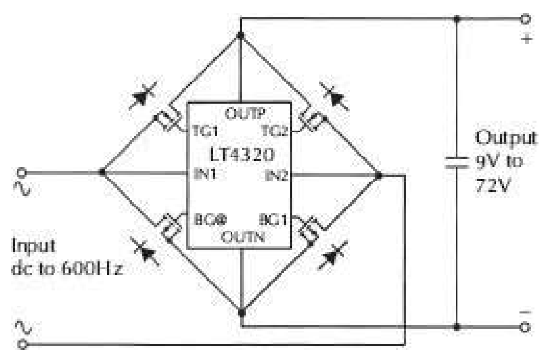


Figure 23-9. Full wave bridge power supply using MOSFETs to replace bridge rectifiers. Courtesy Linear Technology Corporation.

The LT4320 operating temperatures range from – 40°C to 85°C. Packaging options comprise an eight-pin, 3 by 3mm dual flat no-lead (DFN) package and a 12-lead mini small-outline package (MSOP) with enhanced high-voltage pin spacing.

23.3.3 Three-Phase Power Supplies

Three-phase power supplies are common in the industry but are seldom used to power audio circuits directly. They are used as the input power to power an entire system—for instance, a portable high-power outdoor rock system. The characteristics of three-phase supplies are shown in Table 23-8.

23.3.4 Resistance Voltage Dividers

A *resistance voltage divider* is shown in Fig. 23-10. In this system of voltage division, the resistors are connected in series with the particular load they feed. The resistors are calculated by means of Ohm's law

$$R = \frac{V}{I} \tag{23-11}$$

The wattage is computed by

$$\begin{aligned} P &= \frac{V^2}{R} \\ &= I^2 R \end{aligned} \tag{23-12}$$

Generally, when a series-resistance voltage divider is used, a separate bleeder resistor is also used to secure better regulation. Each section should have a separate bypass capacitor of 10μF or more to ground. The bypass capacitors stabilize and improve the filtering and decouple the various levels. This is particularly true for the series-type voltage divider.

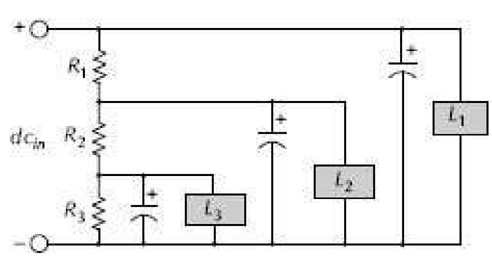


Figure 23-10. Shunt-type voltage-divider system.

There are two common types of voltage dividers, the *shunt* and the *series* types. The shunt type shown in Fig. 23-10 is designed to supply three different voltages to external devices. The upper circuit supplies load L_1, the second circuit supplies L_2, and the third circuit supplies L_3. All circuits are common to ground.

The total current required is the total current of the three external circuits, or $I_{L1} + I_{L2} + I_{L3}$, plus an additional current called the *bleeder current*. The bleeder current flows only through the resistors and not through the external circuits. It is generally 10% of the total current.

Resistor R_1 is calculated first, because only bleeder current flows through this resistor,

$$R_1 = \frac{V}{I} \qquad (23\text{-}13)$$

where,
V is the L_1 voltage also across R_1,
I is the bleeder current.

The voltage at the top of R_2 is the L_2 voltage to ground. Subtracting the voltage drop across R_1 results in a voltage across R_2. The current through R_2 is the current of load L_1 plus the bleeder current

$$R_2 = \frac{V_{L_2} - V_{L_1}}{I_{R_1} + I_{L_1}} \qquad (23\text{-}14)$$

Resistor R_3 has the current of loads L_1 and L_2 plus the bleeder current flowing through it or

$$R_3 = \frac{V_{L_3} - V_{L_2}}{I_{R_1} + I_{L_1} + I_{L_2}} \qquad (23\text{-}15)$$

The current of load L_3 does not flow through any part of the voltage-divider system; therefore, it requires no further considerations.

23.4 Filters

A *power-supply filter* is a series of resistors, capacitors, and/or inductors connected either passively or actively to reduce the ac or ripple component of the dc power supply.

23.4.1 Capacitor Filters

A *capacitor filter* employs a capacitor at its input, as shown in Fig. 23-11A. Power supplies with an input capacitor filter have a higher output voltage than one without a capacitor because the peak value of the rectifier output voltage appears across the input filter. As the rectified ac pulses from the rectifier are applied across capacitor C, the voltage across the capacitor rises nearly as fast as the pulse. As the rectifier output drops, the voltage across the capacitor does not fall to zero but gradually diminishes until another pulse from the rectifier is applied to it. It again charges to the peak voltage. The capacitor may be considered a storage tank, storing up energy to the load between pulses. In a half-wave circuit, this action occurs 60 times per second, and in a full-wave circuit, it occurs 120 times per second.

For a single-phase circuit with a sine-wave input and without a filter, the peak inverse voltage at the rectifier is 1.414 times the rms value of the voltage applied to the rectifier. With a capacitor input to the filter, the peak inverse voltage may reach 2.8 times the rms value of the applied voltage. This data may be obtained by referring to Table 23-8.

When a dc voltmeter is connected across the unfiltered output of a rectifier, it will read the average voltage. Because of the inertia of the meter pointer movement, the meter cannot respond to the rapidly changing pulses of the half-wave rectified current but acts as a mechanical integrator. The pointer will be displaced an amount proportional to the time average of the applied voltage waveform.

The average voltage (V_{av}), as read by the dc voltmeter, is

$$V_{av} = \frac{V_p}{\pi} \tag{23-16}$$

where,
V_p is the peak voltage,
π is 3.1416….

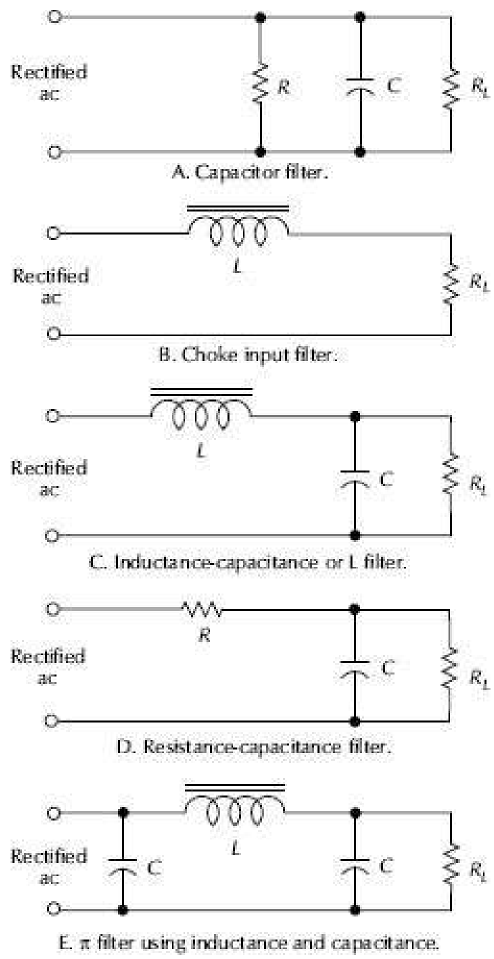


Figure 23-11. Capacitive, inductive, and π filters.

The ripple factor is

$$\gamma = \frac{I_{dc}}{4\pi\sqrt{3}fCV_{dc}} \qquad (23\text{-}17)$$

where,
I_{dc} is the output dc current,
f is the ripple frequency,
C is the filter capacitor in F,
R_L is the load resistance in Ω.

Capacitor filters operate best with large filter capacitors and high-resistance loads. As the load resistance is lowered, the ripple increases and regulation decreases.

Filtering efficiency is reduced, and the internal leakage is increased when the capacitor's power factor increases. Electrolytic capacitors should be removed when their power factor reaches an excessive value. In an ideal capacitor, the current would lead the voltage by 90°. Capacitors are never ideal because a small amount of leakage current always exists through the dielectric. Also, a certain amount of power is dissipated by the dielectric, the leads, and their connections. All this adds up to power loss. This power loss is termed *phase difference* and is expressed in terms of power factor (PF). The smaller the power factor value, the more effective the capacitor. Since most service capacitor analyzers indicate these losses directly in terms of power factor, capacitors with large power factors may be readily identified. Generally speaking, when an electrolytic capacitor reaches a power factor of 15%, it should be replaced. The filtering efficiency for different values of power factor can be read directly from Table 23-9.

Table 23-9. Filtering Efficiency versus % Power Factor

Filtering Efficiency	% PF	Filtering Efficiency	% PF
100	0.000	35	0.935
90	0.436	30	0.955
80	0.600	25	0.968
70	0.715	20	0.980
60	0.800	15	0.989
50	0.857	10	0.995
45	0.895	5	0.999
40	0.915		

23.4.2 Inductive Filters

An *inductive filter* employs a choke rather than a capacitor at the input of the filter, as shown in Fig. 23-11B. Although the output voltage from this type of filter is lower, the voltage regulation is better.

A choke filter operates best with maximum current flow. It has no effect on a circuit when no current is flowing. The critical inductance is the inductance required to assure that current flows to the load at all times. An inductor filter depends on the property of an inductor to oppose any change of current.

To assure that current flows continuously, the $\sqrt{2}I_{rms}$ peak current of the ac component of the current must not exceed the direct current $I_{dc} = I_{dc}/R_L$. Therefore,

$$X_L \geq \frac{\sqrt{2}}{3R_L} \qquad (23\text{-}18)$$

and

$$L_C = \frac{R_L}{3 \times 2\pi f} \qquad (23\text{-}19)$$

where,
L_C is the critical inductance,
R_L is the load resistance.

Filter chokes should be selected for the lowest possible dc resistance commensurate with the value of inductance.

The ripple factor (γ) for an inductive filter is

$$\gamma = \frac{R_L + R_C}{3\sqrt{2} \times 2\pi f} \qquad (23\text{-}20)$$

where,

R_L is the load resistance in Ω,

R_C is the choke resistance in Ω,

f is the ripple frequency.

23.4.3 Combination Filters

Combination filters use a combination of resistors, capacitors, and inductors to improve the filtering. The simplest is a resistor-capacitor filter and the more complicated is a series of inductance-capacitor (*LC*) circuits.

23.4.3.1 Inductance-Capacitance Filters (LC)

Inductance-capacitance filters, sometimes called *L filters*, use an inductor as an input filter and a capacitor as the second stage of the filter, Fig. 23-11C. LC filters operate well under varying load conditions.

The inductive reactance of the choke in an LC filter section tends to oppose any change in the current flowing through the winding, creating a smoothing action on the pulsating current of the rectifier. The capacitor stores and releases electrical energy, also smoothing out the ripple voltage, resulting in a fairly smooth output current.

The ripple factor for an LC filter is

$$\gamma = \frac{\sqrt{2}X_C}{3X_L} = \frac{\sqrt{2}}{3 \times 2\pi fC \times 2\pi fL} = \frac{0.01}{f^2 CL} \qquad (23\text{-}21)$$

where,
X_C is the capacitance reactance in Ω,
X_L is the inductive reactance in Ω,
f is the frequency of ripple,
C is the capacitance in F,
L is the inductance in H.

When multiple LC filters are connected together, the ripple factor is

$$\gamma = \frac{\frac{\sqrt{2}}{3}}{(16\pi^2 f^2 LC)^n} = \frac{0.47}{(157.9 f^2 LC)^n} \quad (23\text{-}22)$$

where,
L is the inductance in H,
f is the ripple frequency,
C is the capacitance in F,
n is the number of sections.

23.4.3.2 Resistance-Capacitance Filters

Resistance-capacitance filters, Fig. 23-11D, employ a resistor and capacitor rather than an inductor and capacitor. The advantages of such a filter are its low cost, light weight, and the reduction of magnetic fields. The disadvantage of such a filter is that the series resistance induces a voltage drop that varies with current and could be detrimental to the circuit operation. An RC filter system is generally used only where the current demands are low. RC filters are not as efficient as the LC type, and they may require two or more sections to provide sufficient filtering.

23.4.3.3 π Filters

A π (pi) *filter* has a capacitor input followed by an LC section filter, Fig. 23-11E. π filters have a smooth output and poor regulation. They are often used where the transformer voltage is not high enough and low ripple is required. By using the input capacitor, the dc voltage is boosted to the peak voltage. The ripple factor for a π filter is

$$\gamma = \sqrt{2}\,\frac{X_{C1}X_{C2}}{R_L X_{L1}} \qquad (23\text{-}23)$$

where,
X_{C1} is the capacitive reactance of the first capacitor,
X_{C2} is the capacitive reactance of the second capacitor,
R_L is the load resistance,
X_{L1} is the inductive reactance of the choke.

When the choke is replaced with a resistor, the ripple factor becomes

$$\gamma = \sqrt{2}\,\frac{X_{C1}X_{C2}}{R_L R} \qquad (23\text{-}24)$$

where,
R is the filter resistor.

23.5 Regulated Power Supplies

A *regulated power supply* holds the output constant with variations in load, current, or input voltage. Regulated supplies may be simple shunt or series regulators with 1–3% regulation or high gain supplies with 0.001% regulation and 0.001% ripple.

Power supplies may be connected in parallel, but to protect the supplies, diodes are connected in the positive lead of each power supply. When the diode is in its normal conducting mode, it must be capable of withstanding the short-circuit current of its regulator. The piv rating of the diode must be equal to or greater than the maximum open-circuit potential of the highest-rated power supply.

All regulated supplies have a reference element and a control element. The amount of electronics between the two elements determines the quality and regulation of the supply, Fig. 23-12.

The reference element is the unit that forms the foundation of all voltage regulators. The output of the regulated power supply is equal to or a multiple of the reference. Any variation in the reference voltage will cause the output voltage to vary; therefore, the reference voltage must be maintained as stable as possible.

The control element regulates the output voltage keeping it constant. The regulator type is named after the control element—namely, series, shunt, or switching, Fig. 23-12A, B, C. The control element is an electronic variable resistor that drops voltage either in series with the load or across the load. Control element configurations are shown in Fig. 23-13.

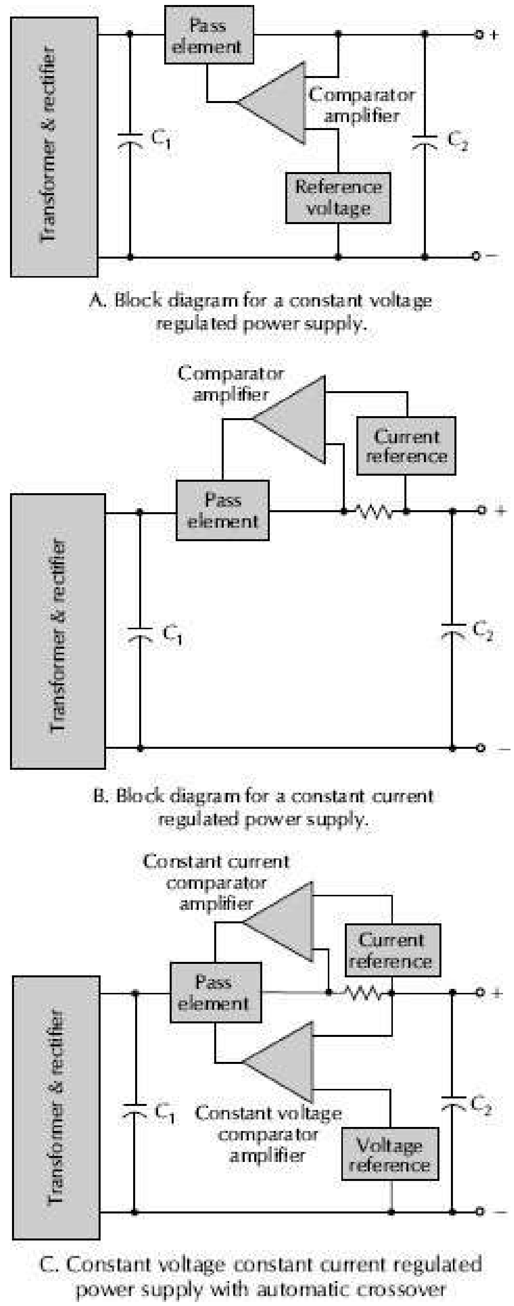


Figure 23-12. Regulated power supplies.

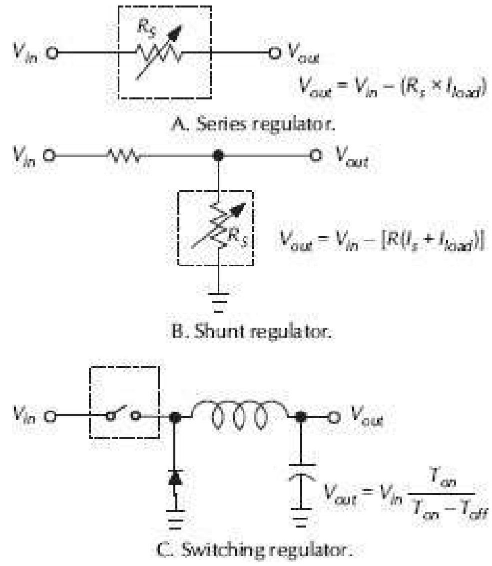


Figure 23-13. Examples of various control elements.

All regulated supplies draw standby current, which is the current drawn by the power supply with no output load. The input voltage to regulated supplies is filtered dc. The smoother the input voltage is, the smoother the output will be. The capacitor C_1 shown in Fig. 23-12 is used to smooth the output and reduce ripple of the dc source.

The comparison amplifier constantly monitors the output, reducing ripple because the reference voltage is smooth dc and the output ripple voltage appears to the comparator like a varying load. The regulator or pass transistor attempts to follow it, reducing ripple.

A *constant-voltage regulated power supply* is designed to keep its output voltage constant, regardless of the changes in load current, line voltage, or temperature. For a change in the load resistance, the output voltage remains constant to a first approximation, while the output current changes by whatever amount is necessary to accomplish this, Fig. 23-12A. Its impedance curve is shown in Fig. 23-14A.

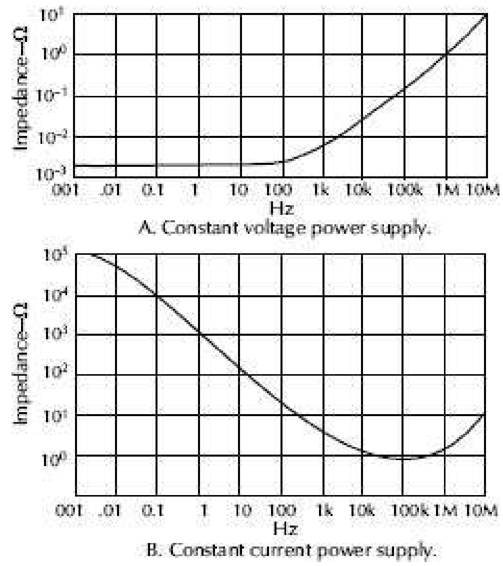


Figure 23-14. Typical internal impedance characteristics for a regulated power supply.

An ideal constant-voltage power supply has zero impedance. For well-designed voltage-regulated power supplies, the internal output impedance ranges from 0.001 to 3 Ω for frequencies from dc to 1MHz. The actual impedance is a function of the load and the type of equipment being fed by the supply. A practical constant-voltage power supply has a very low impedance at the lower frequencies, and the impedance rises with frequency.

A *constant-current regulated power supply* is designed to keep its output current constant, regardless of the changes in load impedance, line voltage, or temperature. For a change in the load resistance, the output current remains constant to a first approximation, although the output voltage changes by whatever amount is necessary to accomplish this, Fig. 23-12B. Its impedance characteristics are given in Fig. 23-14B.

A perfect constant-current supply would have infinite impedance at all frequencies. However, these ideals are not achieved. The constant-current supply has a rather high impedance at the lower frequencies which decreases at the higher frequencies.

A *constant-voltage, constant-current regulated power supply*, Fig. 23-12C, acts as a constant-voltage source for comparatively large values of load resistance and as a constant-current source for comparatively small values of load resistance. An automatic crossover (or transition) between the two modes of operation occurs at the critical or crossover value of load resistance (R_C)

$$R_C = \frac{V_s}{I_s} \qquad (23\text{-}25)$$

where,
V_s is the voltage-control setting,
I_s is the current-control setting.

23.5.1 Simple Regulated Supplies

A simple supply consists of only the control element and the reference element. The *solid-state zener diode* has almost replaced the gaseous tube reference element because it is smaller and has better regulation, wide voltage range, and wide power range. Referring to the basic design in Fig. 23-15A, the zener diode is connected in series with the limiting resistor R_1 and in parallel with the output. As a rule, the zener diode current I_Z is chosen for a value of 10% of the load current I_L. The value of the series resistance R_1 can be calculated using the equation

$$R_1 = \frac{V_s - V_{L_1 out}}{I_L + I_Z} \qquad (23\text{-}26)$$

where,

V_s is the voltage source,
V_{out} is the output voltage,
I_L is the load current,
I_Z is the zener current, (normally 10% of I_L.)

The power dissipated in R_1 is I^2R. The dissipation is only for a condition where the load current remains constant at its design current. If the load current is completely removed, the current through the diode increases to the design load current plus the design zener current.

Zener diodes can be connected in series across the output of a dc supply, provided the power-handling capabilities and the current-operating ranges are similar, Fig. 23-15B.

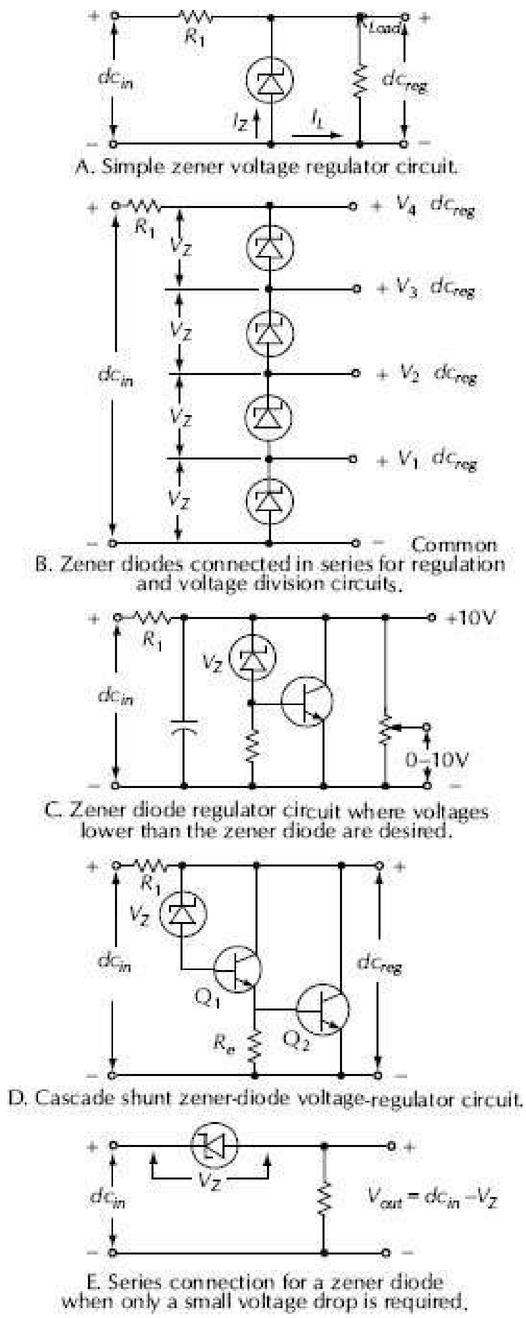
Fig. 23-15C shows a regulated 10V supply with an adjustable 0 to 10V output. The adjustable output is not regulated because there is series resistance between the output and the input from the potentiometer.

A cascade shunt regulator is given in Fig. 23-15D. The zener diode controls the base potential of transistor Q_1, which functions as an emitter follower and circuit amplifier. This circuit is used where large current variations are encountered.

If only a small voltage drop is required, i.e., 5 to 6V, the configuration in Fig. 23-15E might be employed. In this instance, the entire load current plus the current through R_1 must flow through the diode, and it could be easily damaged.

23.5.2 Phase-Controlled Regulated Power Supplies

In the *phase-controlled* supply, the pass element is switched on and off at line frequency and controls the output voltage by a varying pulse width. This is most often accomplished by using an SCR as the pass element. By delaying the firing point of the SCR_1 in each cycle, the output voltage can be varied, Fig. 23-16. SCR_1 is fired by applying a voltage to the gate. The voltage is obtained by C_1 charging through R_2 and the ballast lamp. When the gate firing voltage is reached across C_1, SCR_1 fires. Once the SCR_1 is on, it remains on until its anode voltage goes to zero, which is during the second half of the cycle. When SCR_1 is on, C_1 discharges and remains discharged until the phase of the line voltage returns to zero. The rate that the C_1 charges is controlled by Q_1. When Q_1 is turned on, much of the C_1 charging current is shunted around C_1, requiring a longer time to charge C_1, thus delaying the firing of SCR_1. As the line voltage increases, the resistance of VDR_1 and VDR_2 decreases, turning Q_1 on more and thus slowing the charging rate of C_1. Since the output is a series of pulses with a high rise time of the leading edge, a filter is required on the output to smooth the dc.



A. Simple zener voltage regulator circuit.

B. Zener diodes connected in series for regulation and voltage division circuits.

C. Zener diode regulator circuit where voltages lower than the zener diode are desired.

D. Cascade shunt zener-diode voltage-regulator circuit.

E. Series connection for a zener diode when only a small voltage drop is required.

Figure 23-15. Various regulator circuits using zener diodes.

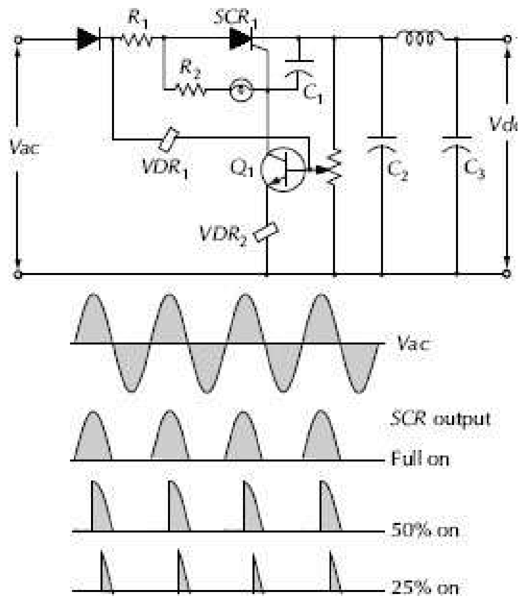


Figure 23-16. Phase controlled regulated supply.

23.5.3 Complex Power Supplies

Complex supplies include a pass element, a sampling element, and a comparator element, and they may include a preregulator, current limiting, undervoltage and overvoltage protection, and remote sensing.

Pass Elements. A transistor or group of transistors connected in parallel and placed in series with the output of a regulated power supply to control the flow of the output current. A *pass element* is another name for *control element*.

Reference Elements. The unit that forms the foundation of all voltage regulators. The output of the regulated power supply is equal to or a multiple of the reference. Any variation in the reference voltage will cause the output voltage to vary; therefore, the reference voltage must be maintained as stable as possible.

Sampling Elements. The device that monitors the output voltage and translates it into a level comparable to the reference voltage. The variations in the sampling voltage versus the reference voltage is the error voltage that ultimately controls the regulator output.

Preregulator. Monitors the voltage across the series regulator and adjusts the input V_{in} to maintain the regulator voltage at approximately 3V. This regulator voltage is held relatively constant regardless of input or output conditions. This reduces the power dissipated and the number of transistors in the series regulator, Fig. 23-17.

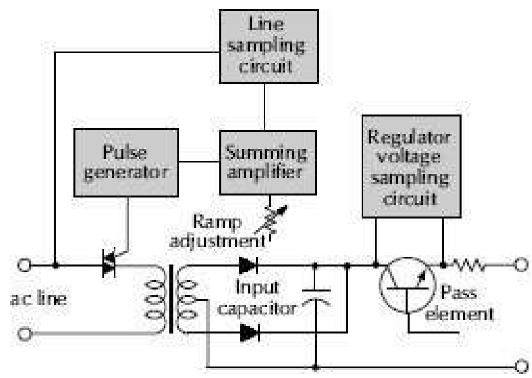


Figure 23-17. Diagram of a preregulated power supply.

Comparator Elements. Compares the feedback voltage from the sampling element with the reference and provides gain for the detected error level. This signal controls the control circuit.

Current Limiting. A method used to protect the pass transistor by limiting the current within the safe operating range. The simplest current-limiting device is a resistor in series with the load which unfortunately affects regulation by the IR drop across the resistor.

To overcome this, constant current limiting is used where the voltage drop across the series resistor is sampled. The output voltage remains constant up to a predetermined current at which time the voltage decreases to limit the output current.

A third current limiting is foldback current limiting in which the load current actually decreases as the load continues to increase beyond I_{max}. This is usually only used in high current supplies.

The conventional current-limiting power supply of Fig. 23-18A is protected from instantaneous short circuits but long duration shorts can overheat Q_2, leading to its eventual failure. In Fig. 23-17B this circuit is modified to produce foldback by adding two voltage feedback resistors, R_3 and R_4. The control transistor Q_1 emitter voltage depends on the power supply output voltage as sampled by the R_3, R_4 voltage divider. If R_1 senses a current overload, the drop across it decreases the output voltage and lowers the emitter voltage of Q_1. Then Q_1 turns on at reduced current through R_1, which limits current flow through Q_2, as shown in the current-foldback characteristic of Fig. 23-18B. The foldback ratio can be adjusted by changing R_3, R_4, or R_1, or all three.

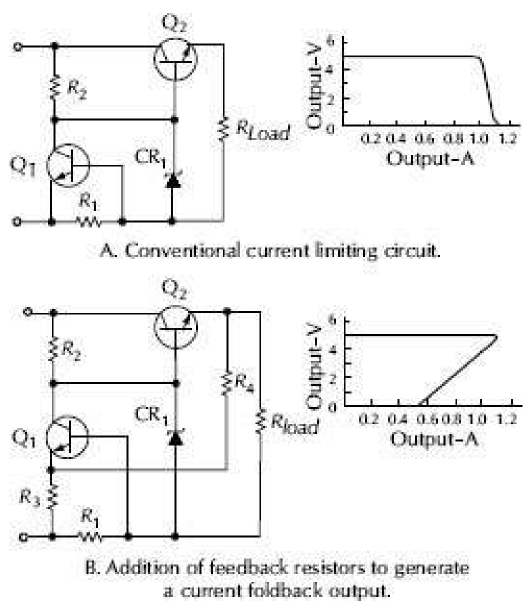


Figure 23-18. Current limiting circuits.

Overvoltage Protection. Protects the load from overvoltage. This may be accomplished internally or as an add-on to the power supply. The circuit monitors the output voltage of a power supply and controls a crowbar circuit where a silicon controlled rectifier (SCR) connected across the output terminals of the supply unit creates a short circuit across the output terminals when a preset voltage is reached.

Remote Sensing. Most medium to high powered ac to dc power supplies and some dc to dc power converters include remote sense connection points. These points, + and – sense are used to regulate the supplies output voltage at the load.[4]

The output cables that connect the power supplies output to its load have resistance so as the current flow increases the voltage drop across the cables also increases. The use of remote sense wires connected to the load will compensate for these unwanted voltage drop, Fig. 19-19.

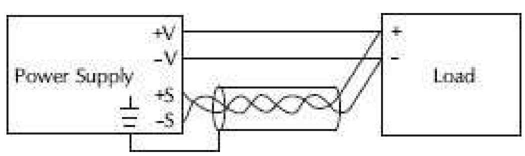


Figure 23-19. Power supply with remote sense lines.

To reduce excessive voltage drop the remote sense function automatically increases the output voltage at the output terminals of the supply to compensate for the unwanted voltage drop in the output cables when a heavy load appears. Most remote sensing circuits are capable of compensating for a 0.25 to 0.75V drop across the output cables. If the voltage drop exceeds the compensating range of the remote sense circuits, the voltage at the load will no longer be regulated which can be remedied by either reducing the length of the output cables or increasing the size of the output cable.

The remote sense leads carry very little current so light gauge wire can be used. Steps should be taken to ensure that the remote sense wires do not pick up radiated noise by either twisting the wires or by shielding them. Fig. 23-20 is a simplified schematic of the power supplies remote sense circuit. It is important to observe the correct polarities, i.e. the plus sense wires should connect to the load near the plus load connection and the minus sense wire should connect the load at the minus loads connection. If the remote sense wires are cross connected current will flow in a sense lines and burnout the internal sense resistors in the supply.

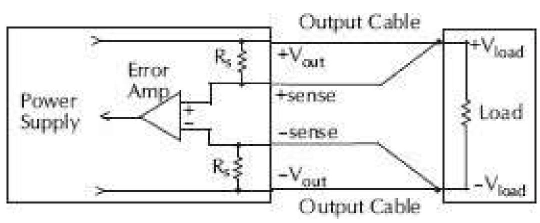


Figure 23-20. Schematic of a remote sense circuit with external output and sense wires.

It is important not to put a switch or fuse in series with one or both of the output wires. If the remote sense lines remain connected to the load with the load lines open, the current will flow in the sense lines destroying the internal resistors in the power supply. If fuses or switches are required, they should be at the load after the connection of the sense wires. Most power supplies are shipped from the factory with the local sense jumpers installed on the power supply.

The wire size and voltage drop for regulated power supplies can be determined by Ohm's Law or with the use of the nomograph in

Tables 18-2, 18-3, and 23-19. Since regulated power supplies are designed to control the output at the power supply output terminals, the conductors used for the supply line must be considered as a part of the power supply load.

Table 23-10. Minimum Recommended dc AWG for 90°C Cabling for Protected Outputs (Courtesy TDK-Lambda)

Total Power module Current Rating (A)	Wire and Lug Gauge (AWG) using 90°C Wire (NEC Table 310.16)
5	18
10	16
15	16
20	14
30	12
40	10
50	8
75	6
100	2
125	2
150	(1) 1AWG or (2) 6 AWG
175	(2) 4 AWG
200	(2) 2 AWG
225	(2) 2 AWG
250	(2) 2 AWG
300	(2) 1 AWG

23.5.4 Switching Regulators

In a switching regulator, the pass transistor operates in an on-off mode, increasing efficiency and reducing heat. The simple switching regulator shown in Fig. 23-21 incorporates a pulse generator circuit that pulses on the pass transistor as the output voltage decreases. As the output voltage increases, the comparator circuit reduces the pulse generator, reducing the on-time of the pass transistor thus reducing the average output voltage. Since the output voltage is a series of pulses, a filter is required to smooth the dc output. An inductance-capacitance filter is commonly used.

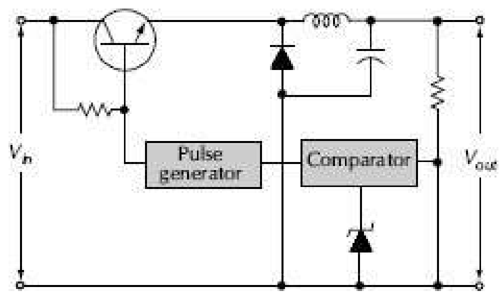


Figure 23-21. Basic switching regulator.

Switching regulators normally operate at 20kHz or higher and have the following advantages:

- Switching regulators are on-off devices, so they avoid the higher power dissipation associated with the rheostat like action of a series regulator. Transistors dissipate very little power when either saturated (on) or nonconducting (off); most of the power losses occur elsewhere in the supply. Efficiency to 95% is typical for switching supplies, as compared to 30 to 45% for linear supplies. Less wasted power means switching supplies run cooler, cost less to operate, and have smaller regulator heat sinks.
- Size and weight reductions for switching supplies are achieved because of their high switching rate. Typically, a switching supply is less than one-third the size and weight of a comparable series-regulated supply.
- Switching supplies can operate under low ac input voltage (brownout) conditions and sustain a relatively long carryover (or

holdup) of its output if input power is lost momentarily because energy is stored in the input filter capacitors. In a switching supply, the input ac voltage is rectified directly, and the filter capacitor charges to the ac voltage peaks. The ac voltage input of the standard linear supply is stepped down through a power transformer and then rectified, resulting in a lower voltage across its filter capacitor. Since the energy stored in a capacitor is proportional to CV^2 and V is higher in switching supplies, their storage capability (and thus their holdup time) is better.

Switching disadvantages include the following:

- A switching supply transient recovery time (dynamic load regulation) is slower than that of a series-regulated supply. In a linear supply, recovery time is limited by the speed of the semiconductors used in the series regulator and control circuitry. In a switching supply, recovery is limited mainly by the inductance in the output filter.
- Electromagnetic interference (emi) is a natural byproduct of the on-off switching. This interference can be transferred to the load (resulting in higher output ripple and noise), it can be transferred back into the ac line, and it can be radiated into the surrounding atmosphere.

Switching regulators can step down (buck), step up (boost), or step down/up and invert (buck boost) the voltage that powers them, Fig. 23-22. The external transistor switches are often included within the switching regulator, usually when the device is specified for moderate-load currents.

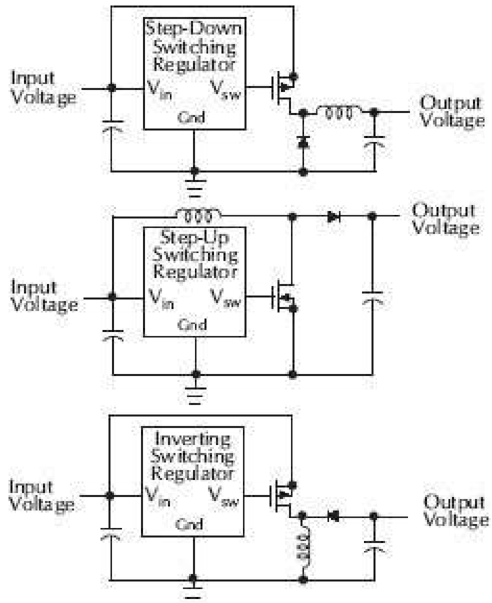


Figure 23-22. Step-down, step-up, and inverting switching regulators. Courtesy Maxim Integrated Products.

23.6 Synchronous Rectification Low-Voltage Power Supplies

Synchronous rectifiers can improve switching-power-supply efficiency, particularly in low-voltage low-power applications compared to Schottky-diode types of supplies.[5]

A synchronous rectifier is an electronic switch that improves power-conversion efficiency by placing a low resistance conduction path across the diode rectifier in a switch-mode regulator. MOSFETs or bipolar transistors and other semiconductor switches can be used.

The forward-voltage drop across a switch-mode rectifier is in series with the output voltage, so losses in the rectifier determine efficiency.

Even at 3.3 V, rectifier loss is significant. For step-down regulators with a 3.3 V output and a 12 V input, the 0.4 V forward voltage of a Schottky diode represents a typical efficiency penalty of about 12%. The losses are less at lower input voltages because the rectifier has a lower duty cycle and thus a shorter conduction time. However, the Schottky rectifier's forward drop is usually the dominant loss mechanism.

For an input voltage of 7.2 V and an output of 3.3 V, a synchronous rectifier improves on the Schottky diode rectifier's efficiency by around 4%. As output voltage decreases, the synchronous rectifier provides even larger gains in efficiency, Fig. 23-23.

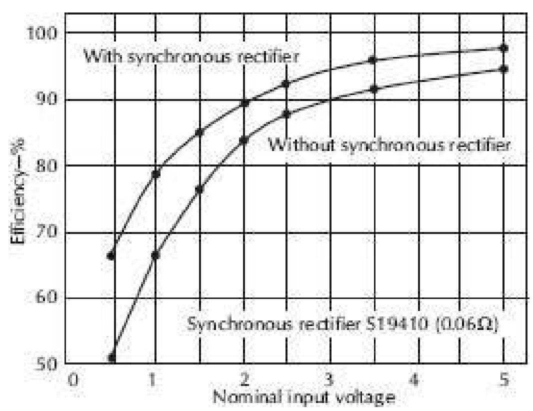


Figure 23-23. Data based on a high-performance buck switch-mode regulator and powered from a standard 7.2V notebook-computer battery shows that the synchronous rectifier has little effect on efficiency at 5V, but offers significant improvements at 3.3V and below. Courtesy Maxim Integrated Products.

23.6.1 Diode versus Synchronous Rectifiers

In the absence of a parallel synchronous rectifier, the drop across the rectifier diode in a switching regulator, Fig. 23-24A causes an efficiency loss that worsens as the output voltage falls. The Schottky diode simple buck converter clamps the switching node, the inductor's swinging terminal, as the inductor discharges.

In the synchronous-rectifier version of Fig. 23-24B, a large N-channel MOSFET switch replaces the diode and forms a half-bridge configuration that clamps the switching node to – 0.1V or less. The diode in Fig. 23-24A clamps the node to −0.35 V. Intuitively, losses in either type of rectifier increase with reduced output voltage. At V_{IN} to V_{OUT}, the rectifier voltage drop is in series with the load voltage for about half the switching period. As the output voltage falls, power lost in the rectifier becomes a greater fraction of the load power.

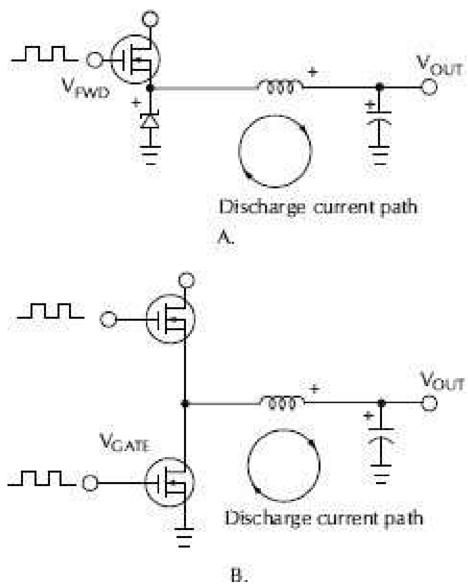


Figure 23-24. A synchronous rectifier replaces the Schottky diode in A with a low $R_{DS(ON)}$ MOSFET in B. The lower-resistance conduction path improves efficiency for the 5V to 3.3 V 3 A converter by 3–4%. Courtesy Maxim Integrated Products.

The basic trade-off between using diode or MOSFET rectifiers is whether the power needed to drive the MOSFET gate cancels the efficiency gained from a reduced forward-voltage drop. The synchronous rectifier's efficiency gain depends strongly on load current, battery voltage, output voltage, switching frequency, and other application parameters. Higher battery voltage and lighter load current enhance the value of a synchronous rectifier. The duty factor, which equals 1 – D, where D equals $t_{on}/(t_{on} + t_{off})$, for the main switch, increases with the battery voltage. Also, the forward drop decreases with the load current.

When comparing diode and synchronous rectifiers, note that the synchronous rectifier MOSFET doesn't always replace the usual Schottky diode. To prevent switching overlap of the high-side and low-side MOSFETs that might cause destructive cross-conduction currents, most switching regulators include a dead-time delay. The synchronous rectifier MOSFET contains an integral, parasitic body diode that can act as a clamp and catches the negative inductor voltage swing during this dead time. This diode is lossy, is slow to turn off, and can cause a 1–2% efficiency drop.

To squeeze the last percent of efficiency out of a power supply, a Schottky diode can be placed in parallel with the synchronous rectifier MOSFET. This diode conducts only during the dead time. A Schottky diode in parallel with the silicon body diode turns on at a lower voltage, ensuring that the body diode never conducts. Generally, a Schottky diode used in this way can be smaller and cheaper than the type the simple buck circuit requires, because the average diode current is low. (Schottky diodes usually have peak current ratings much greater than their dc current ratings.) It's important to note that conduction losses during the dead time can become significant at high switching frequencies. For example, in a 300kHz converter with a 100ns dead time, the extra power dissipated is equal to

$$I_{LOAD} \times V_{FWD} \times td \times f = 6\,\text{mW} \qquad (23\text{-}27)$$

where,
f is the switching frequency,
td is the dead time.

For a 2.5 V, 1W supply, this represents an efficiency loss of about 0.5%.

A logic-control input can shift the synchronous-rectifier operation from the complementary-drive option to the off-at-zero option, Fig. 23-25. When low, "SKIP" allows normal operation. The circuit employs pulse-width modulation (PWM) for heavy loads and automatically switches to a low-quiescent-current pulse-skipping mode for light loads. When high, "SKIP" forces the IC to a low-noise fixed-frequency PWM mode, regardless of the load. Applying a high level to "SKIP" disables the IC's zero-crossing detector, allowing the inductor current to reverse direction, which suppresses the parasitic resonant LC tank circuit.

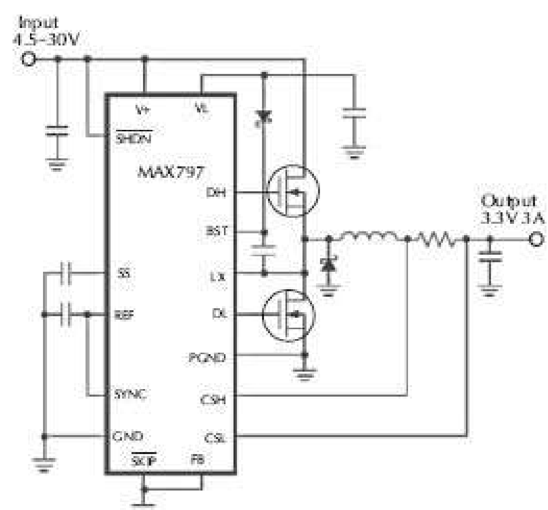


Figure 23-25. A N-channel buck regulator which has a low-noise logic-control input that adjusts the synchronous rectifier's timing

on the fly. Courtesy Maxim Integrated Products.

An issue related to a synchronous rectifier's gate-drive timing is the cross regulation of multiple outputs obtained using flyback windings. Placing an extra winding or a coupled inductor on a buck regulator's inductor core can provide an auxiliary output voltage for the cost of a diode, a capacitor, and a little wire, Fig. 23-26.

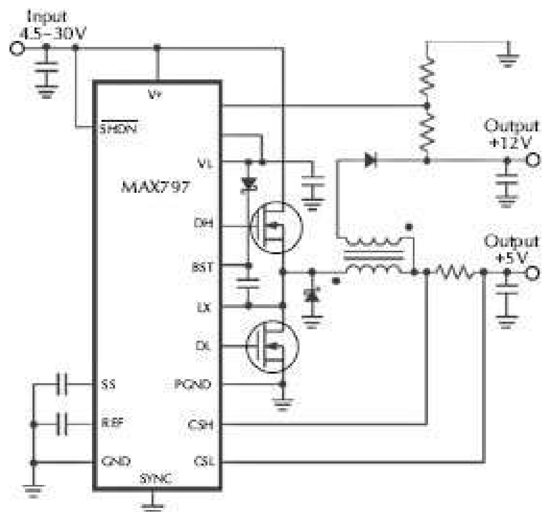


Figure 23-26. A feedback input for the secondary winding (SECFB) greatly improves the cross regulation for multiple outputs under conditions of light primary loading or low I/O differential voltage. Courtesy Maxim Integrated Products.

Normally, the coupled-inductor flyback circuit in Fig. 23-26 stores energy in the core when the high-side switch is on and discharges some of it through the secondary winding to an auxiliary 15V output when the synchronous rectifier's low-side switch is on. During discharge, the voltage across the primary is equal to V_{OUT} + V_{SAT}, where V_{OUT} is the main output and V_{SAT} is the synchronous rectifier's saturation voltage. Therefore, the secondary output voltage equals the primary output times the turns ratio.

Unfortunately, if the synchronous rectifier turns off at zero current and the primary load is light or nonexistent, the 15 V output sags to ground because the core stores no energy at this time. If the synchronous rectifier remains on, the primary current can reverse and let the transformer operate in the forward mode, providing a theoretically infinite output-current capability that prevents the 15 V output from sagging. Unfortunately, quiescent supply current suffers a great deal.

However, the circuit in Fig. 23-26 achieves excellent cross regulation with no penalty in quiescent supply current. A second, extra feedback loop senses the 15 V output. If this output is in regulation, the synchronous rectifier turns off at zero current as usual. If the output drops below 13 V, the synchronous rectifier remains on for an extra microsecond after the primary current reaches zero, so the 15 V output can deliver hundreds of milliamps even with no load on the main 5 V output. This scheme also provides a better 15 V load capability at low values of $V_{IN} - V_{OUT}$, which becomes important if the input voltage drops.

23.6.2 Secondary-Side Synchronous Rectifiers

Multiple synchronous rectifiers on the secondary windings can replace the usual high-voltage rectifier diodes in multiple-output nonisolated applications, Fig. 23-27. This substitution can dramatically improve load regulation on the auxiliary outputs and often eliminates the need for linear regulators, which are otherwise added to increase the output accuracy. The MOSFET must be selected with a breakdown rating high enough to withstand the flyback voltage, which can be much higher than the input voltage. Tying the gates of the secondary-side MOSFETs directly to the gate of the main synchronous MOSFET (the DL terminal) provides the necessary gate drive.

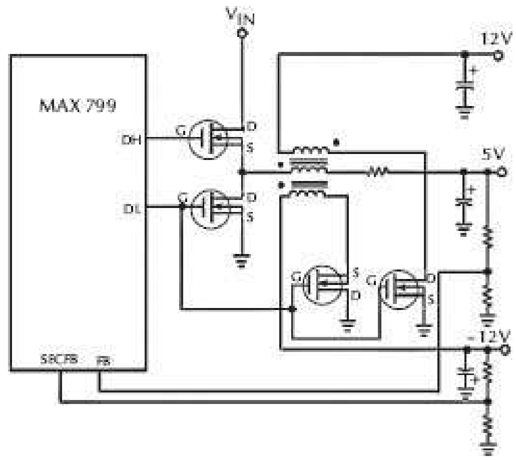


Figure 23-27. Coupled-inductor secondary outputs can benefit from synchronous rectification. To accommodate negative auxiliary outputs, swap the secondary-side MOSFET's drain and source terminals. (For clarity, this simplified schematic omits most of the ancillary components needed to make the switching regulator work.) Courtesy Maxim Integrated Products.

Another neat trick enables a synchronous rectifier to provide gate drive for the high-side switching MOSFET. Tapping the external switching node to generate a gate-drive signal higher than the supply voltage enables the use of N-channel MOSFETs for both switches in a synchronous-rectifier buck converter. Compared to P-channel types, N-channel MOSFETs have many advantages because their superior carrier mobility confers a near 2:1 improvement in gate capacitance and on-resistance.

A flying-capacitor boost circuit provides the high-side gate drive, Fig. 23-28. The flying capacitor is in parallel with the high-side MOSFET's gate-source terminals. The circuit alternatively charges this capacitor from an external 5 V supply through the diode and places the capacitor in parallel with the high-side MOSFET's gate-source terminals. The charged capacitor then acts as supply voltage for the internal gate-drive inverter, which is comparable to several 74HC04 sections in parallel. Biased by the switching node, the inverter's negative rail rides on the power-switching waveform at the LX terminal.

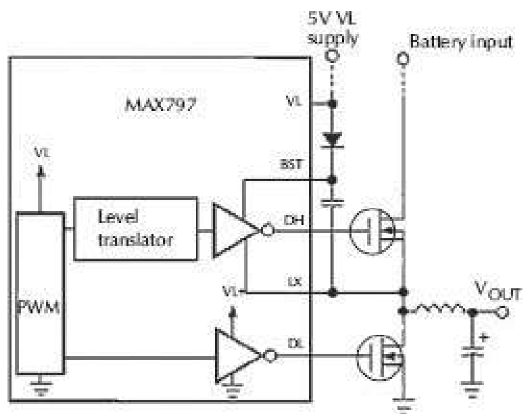


Figure 23-28. Driven by the switching node (the left end of the inductor), the capacitor between BST and LX provides an elevated supply rail for the upper gate-drive inverter. Courtesy Maxim Integrated Products.

A flying capacitor then acts as supply voltage for the internal gate-drive inverter, which is comparable to several 74HC04 sections in parallel. Biased by the switching node, the inverter's negative rail rides on the power-switching waveform at the LX terminal.

The synchronous rectifier is indispensable to the Fig. 23-28 gate-drive boost supply. Without this low-side switch, the circuit may not start at initial power-up. When power is first applied, the low-side MOSFET forces the switching node to 0V and charges the boost capacitor to 5 V.

Synchronous rectifiers can be incorporated in the boost and inverting topologies. The boost regulator in Fig. 23-29 employs an internal pnp synchronous rectifier in the active rectifier block. Boost topologies require the rectifier in series with V_{OUT}, so the IC connects the pnp collector to the output and the emitter to the switching node. The rectifier control block's fast-comparator detects whether the rectifier is forward or reverse-biased and drives the pnp transistor on or off accordingly. When the transistor is on, an adaptive base-current control circuit keeps the transistor on the edge of saturation. This condition minimizes the efficiency loss due to base current and maintains high switching speed by minimizing the delay due to stored base charge.

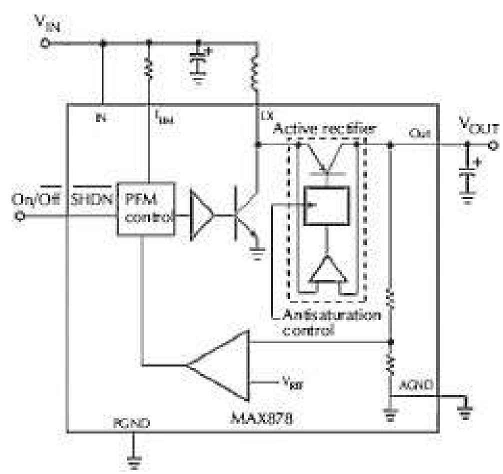


Figure 23-29. The internal synchronous rectifier in this boost regulator, the active rectifier, replaces the Schottky rectifier often used at that location. Courtesy Maxim Integrated Products.

The pnp synchronous rectifier can provide both step-up and step-down action. For ordinary boost regulators, the input voltage range is limited by an input-to-output path through the inductor and the diode. This unwanted path is inherent in the simple boost topology. If V_{IN} exceeds V_{OUT}, the conduction path through the rectifier can drag the output upward, possibly damaging the load with overvoltage.

The pnp-rectifier circuit in Fig. 23-29 operates in switch mode, even when V_{IN} exceeds V_{OUT}, with the active rectifier acting as the switch. This action is more like a regulating charge pump than to a buck regulator, because the buck mode of operation requires a second switch on the high side.

Inverting-topology regulators that generate negative voltages, sometimes called buck-boost regulators, are useful applications for synchronous rectification. Like the boost topology, the inverting topology connects the synchronous rectifier in series with the output rather than to ground, Fig. 23-30. In this example, the synchronous switch is an N-channel MOSFET with its source tied to the negative output and its drain tied to the switching node.

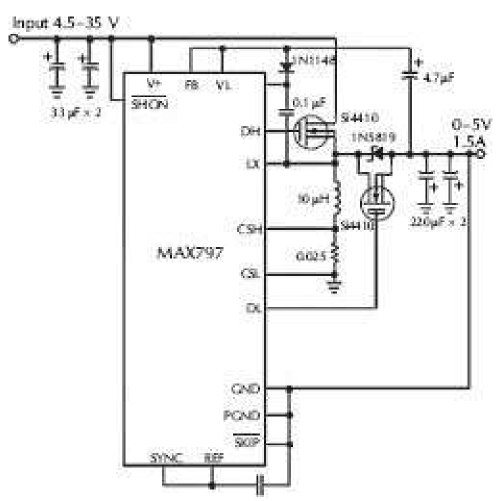


Figure 23-30. The inverting topology requires that the synchronous switch be in series with the output. Courtesy Maxim Integrated Products.

The circuit makes the 300kHz buck regulator perform as an inverting-topology switcher by connecting the IC's GND pin to the negative output voltage instead of circuit ground. This switching regulator's efficiency of about 88% exceeds that of comparable asynchronous-rectifier supplies by 4%.

23.6.3 Autoranging Power Supplies

Autoranging power supplies, sometimes simply called autorangers, are designed to provide a greater range of operation than a typical supply. The power supply is an autoranger if the maximum volts times maximum amps is greater than maximum watts.[6]

Fig. 23-31 shows the output characteristic of a typical nonautoranging power supply. This is called a rectangular output supply. The power supply can operate anywhere within these limits of voltage and current. As long as the required voltage is less than the rated voltage, and the required current is less than the rated current, the power supply will operate properly.

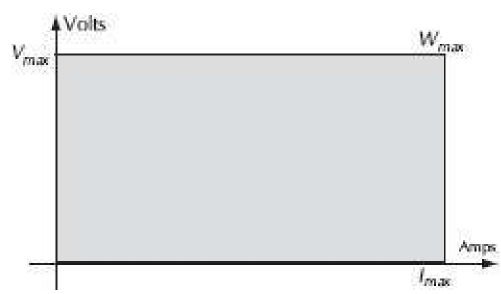


Figure 23-31. A rectangular output power supply can operate anywhere within the limits of V_{max} and I_{max} but achieves W_{max} only at $V = V_{max}$ and $I = I_{max}$.

The issue is the size (i.e., power rating) of the supply. While the supply is rated for a certain maximum power, you can only draw that maximum power when operating at maximum rated voltage and maximum rated current.

Fig. 23-32 shows the output characteristic of an autoranging supply. The diagram's key feature is the curve, which is the locus of points where voltage times current equals maximum power. At the ends of the curve, you will still have a limit on voltage and a limit on current. A figure of merit on autorangers is the ratio of voltage at the endpoints of the curve.

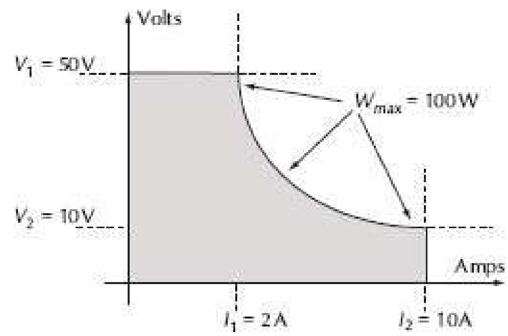


Figure 23-32. In an autoranging power supply, V_1 represents the maximum voltage and I_1 represents the corresponding current such that $I_1 = W_{max}/V_1$. I_2 represents the maximum current, and V_2 represents the corresponding voltage such that $V_2 = W_{max}/I_2$. Along the curved portion, the available voltage and current is limited by W_{max}. The characteristic shown is for a 50V, 10A, 100W 5:1 autoranging power supply.

For the output characteristic shown in Fig. 23-32, the maximum voltage (V_1) is 50 V, and the maximum current is 2A, therefore this is a 100W supply. The maximum current available is 10A at a maximum voltage of 10 V (V_2). The ratio of V_1/V_2 is 5, so this is a 5:1 autoranger. As this ratio gets larger, the power supply becomes more flexible and can operate over a wider range. Typical autorangers will be 2:1 to 5:1.

Autorangers provide greater application flexibility, but they suffer from issues with accuracy. Given the wide voltage and current range over which they operate, the built-in measurement system also needs to operate over this wide range. In the example, the measurement system must be sized for 50 V and 10A even though this is only a 100W supply.

Autorangers tend to be 20% to 30% more expensive than their rectangular equivalents. First, they need to have an additional control loop to keep the output power within the power envelope of the supply. Second, the components must be rated for the highest voltage and the highest current, as these maximum conditions will appear at some point, although never at the same time. Third, achieving the best possible measurement accuracy over this wider operating range will mean a more costly measurement system.

Autorangers operate over a wider range, and can replace several rectangular output supplies of the same rating. However, another approach is just to use a larger rectangular output supply. A 500W rectangular output supply provides the same operating points as the 100W 5:1 autoranger.

While there is something appealing about the flexibility and efficiency of the 100W autoranger, a larger rectangular output supply may be a better choice. To decide which is better, compare the cost and the physical size of the two alternatives. The 100W autoranger may seem less costly and smaller than a 500W rectangular power supply, but given today's modern switching power supply designs, the 500W rectangular output supply may be similar to the 100W 5:1 autoranger.

When a power supply output requires a limited range of voltage adjustment, the extra cost for wide operating range of the autoranger is wasted. Conversely, if the power supply requires a wide range of operation, an autoranger could be useful as it gives more flexibility with a smaller power supply, saving cost and space.

23.7 Converters

A *converter* changes low-voltage dc to high-voltage dc. Basically, a dc-to-dc converter consists of a dc source of potential (often a battery) applied to a pair of switching transistors. The transistors convert the applied dc voltage to a high-frequency ac voltage. The ac voltage is then transformed to a high voltage that is rectified to dc again and filtered in the conventional manner. Power supplies of this nature are often used for a source of high voltage, where the usual ac line voltage is not available.

23.8 Inverters

An *inverter* converts direct current to alternating current. Inverters are used in applications where the primary source of power is direct current. Because direct current cannot be transformed, it is converted to alternating current so that alternating current output from the inverter may be applied to a transformer to supply the desired voltage.

An inverter operates much like the switching circuit and transformer section of a converter. In Fig. 23-33, R_1 and R_2 assure that the oscillator (switch) will start. T_1 is a saturable base-drive transformer that determines the drive current to turn on Q_1 or Q_2. T_2 is a non-saturable transformer; therefore, collector current through Q_1 and Q_2 is dependent upon load. Base resistors R_b are current limiting resistors. By adding a rectifier and filter section, this inverter can be changed to a converter.

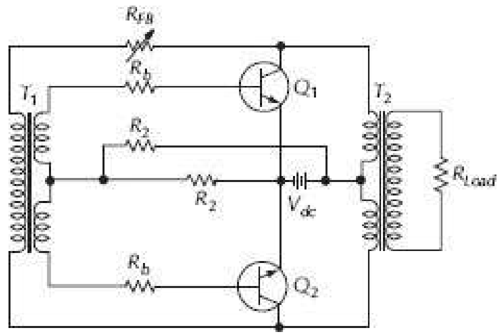


Figure 23-33. Two-transistor, two-transformer, push-pull inverter that uses a resistive voltage-divider network to provide starting bias.

23.9 Ultracapacitor Uninterrupted Power Supply

Ultracapacitors (UC) also known as supercapacitors (SC) or electric double layer capacitors (EDLC) do not degrade in time as batteries do so reliability is high. Also see Chapter 14.1.18 Supercapacitors.

UCs have a capacity one million times that of a standard electrolytic capacitors. This is accomplished by depositing carbon on an aluminum substrate which can be etched to give it a dramatically increased surface area and low impedance. A typical 30mm × 50mm UC can have up to 400 F of capacitance with a working voltage of 2.7 V.

The energy stored in a capacitor is $CV^2/2$. To have the same energy density as current Lithium Ion batteries, the working voltage for a given size would have to be increased to 5 V. This would yield 4 times the energy density of current UC technology and make UCs the ultimate energy storage device. It is just a matter of time before technology reveals materials capable of having dielectrics in the 5 V or higher range and making batteries as we know them today obsolete.

UCs are often paired with batteries where they can produce the short high current pulses that degrade batteries. UCs work well with the following battery types:

- Lithium Thionyl Chloride
- Lithium Manganese Dioxide
- Lithium Iodine
- Zinc Air
- Zinc/Silver Oxide
- Poly (Carbon Monofloride-Lithium

Charging. Ultracapacitors have extremely low internal impedance so power limited charging must be used to avoid overloading the charger. UCs are also sensitive to voltage. They must be operated below their rating to avoid destruction. Designers tend to charge them as close as possible to their maximum voltage to extract the maximum energy from them.

Discharging: Unlike batteries ultracapacitors store their energy over their entire voltage range which makes the design of the boost complex due to the wide input voltage range and overall efficiency.

UPS. An uninterrupted power supply (UPS) for medical computers manufactured by Ram Technologies utilizes ultracapacitor technology. The model 8000 Ultra UPS module contains the charge and discharge circuitry to ensure high-efficiency energy transfer, Fig. 23-34. The proprietary patent pending module is designed to directly interface with RAM Technologies line of ATX/SFX medical-grade power supplies. The unit can be modified by Ram Technologies to operate with other sensitive and/or life-threatening devices. The module may be expanded by adding additional ultracapacitor modules. The base module contains 8000J of energy;

expansion modules also contain 8000J of energy. Any number of additional modules can be added to increase load capabilities. Fig. 23-33 shows a typical installation in a computer.

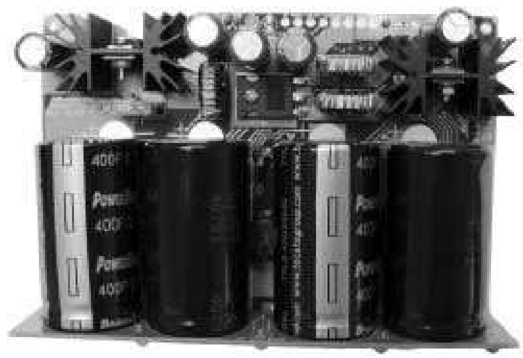


Figure 23-34. Ultracapacitor UPS power supply module. Courtesy Ram Technologies LLC.

Figure 23-35. Typical installation of a ultracapacitor UPS module in a computer. Courtesy Ram Technologies LLC.

The module's input voltage is +12Vdc and has an efficiency of >90%. Charge time is 2 minutes for each 8kJ module.

The maximum output current is 30A at 12Vdc. Run time is

$$Run\ time = \frac{number\ of\ modules \times 133}{dc\ load} \qquad (23\text{-}28)$$

where,
Run time is in min,
dc load is in W.

23.10 Batteries

The end requirements determine the best type of battery to use. Table 23-11 shows the characteristics of various types.[7]

23.10.1 Lead-Acid Batteries

The lead-acid storage battery was invented by Gaston Planté in 1860 and is one of the most widely used forms of battery power. The principal drawback to this type of battery has been the liquid electrolyte and the fumes given off when charging and discharging. Today the sealed lead-acid battery may take its place with other rechargeable batteries such as the nickel-cadmium battery. Since small amounts of gas may be generated in any battery during the charge or discharge cycle, lead-acid batteries are vented so that the gas but not the electrolyte escapes.

Lead-acid cells are normally 2.1V and are easily connected in series to produce 6 V and 12 V automotive types, 24 V aircraft types and 36 V types for golf carts, etc. Lead-acid batteries, because of their availability, high Ah ratings, and ability to be connected in series, work well powering sound systems in the field.

The type and amount of charge determines the condition of the cell. If a lead-acid battery is overcharged, excessive water consumption and hydrogen evolution results, while constant undercharging results in a battery with less and less capacity.

The recharge factor (RF) is defined as the charge Ah divided by the previous discharge Ah. The RF must always be greater than 1 to bring the battery back to capacity. The actual RF is between 1.04 and 1.20, with sealed lead acid batteries requiring less than the standard vented type. Fig. 23-36A shows the state of charge (SOC) achieved versus the RF for a lead-acid battery. Fig. 23-36B shows the SOC versus the RF after a number of cycles. Note that the battery rapidly loses its capacity if it is not overcharged, i.e., more is put in than is taken out.

Table 23-11. Battery types and typical specifications

	Lead Acid	NiMH	NiCd	Li-Ion	Li-Ion Polymer	Reusable Alkaline
Energy Density (Wh/kg)	30-50	60-120	45-80	110-160	100-130	80 (initial)
Cycle Life (to 80% of initial capacity)	200 to 300	300 to 500	1500	500 to 1000	300 to 500	50 (to 50%)
Fast Charge Time (hours)	8-16	2-4	1 typical	2-4	2-4	2-3
Self-discharge/Month (room temperature)	5%	30%	20%	10%	~10%	0.3%
Cell Voltage (nominal)	2V	1.25V	1.25V	3V	3.6V	1.5V
Operating Temperature (discharge only)	−20° to 60°C	−20° to 60°C	−40° to 60°C	−20° to 60°C	0° to 60°C	0° to 60°C
Maintenance Requirement	3 to 6 months	60 to 90 days	30 to 60 days	not required	not required	not required
Typical Battery Cost (Reference only)	$25/(6V)	$60/(7.2V)	$50/(7.2V)	$100/(7.2V)	$100/(7.2V)	$5/(9V)
Cost per Cycle	$0.10	$0.12	$0.04	$0.14	$0.29	$0.10-$0.50

The use of a trickle charger with a storage battery shortens the life of the battery because of overcharging. Trickle chargers should only be used when it is impractical to charge a battery by other means. A practical approach to the problem is to adjust the charging voltage to a value between 2.15 V and 2.17 V per cell.

A better, but more elaborate, method is to check the specific gravity of the cells over a period of several months and adjust the charging voltage to a value where the specific gravity is maintained at 1.250. Compensation must be made for temperature changes when reading the specific gravity. Four gravity points are added to the reading for every 10°F (5°C) the electrolyte is above a temperature of 80°F (27°C).

The freezing point of a battery electrolyte depends on the specific gravity of the electrolyte, Table 23-12. Lead acid batteries freeze when in a discharged state, so it is imperative that they be kept fully charged when in subfreezing temperatures. If a storage battery is left in a discharged condition for any length of time, the plates may be damaged due to sulfation.

23.10.2 Lead-Dioxide Batteries

A *lead-dioxide battery* is a gelled electrolyte, maintenance-free type that exhibits high capacity and long life when properly applied and charged. To prevent electrolyte movement in the battery, the electrolyte in sealed batteries is immobilized by the use of a gelling agent that stores the electrolyte in highly porous separators. With this construction, the loss of water is minimized.

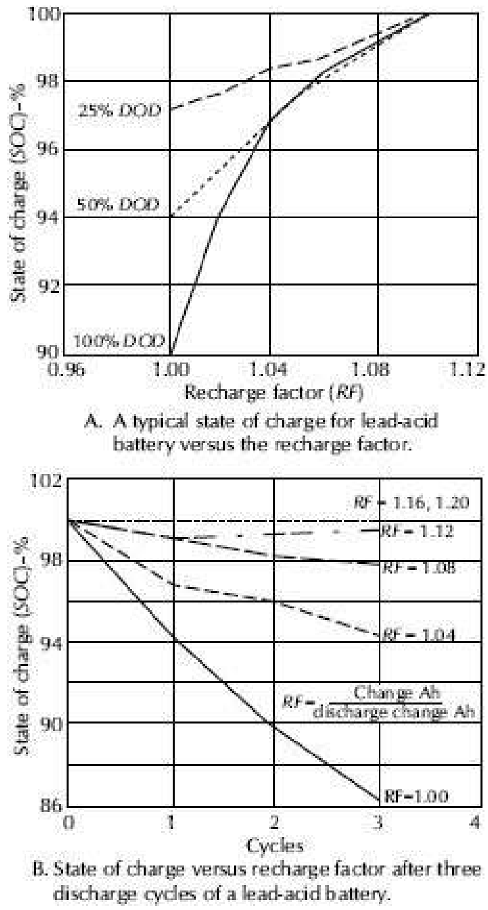


Figure 23-36. Lead acid battery.

Table 23-12. Effect of Specific Gravity on Freezing Point of a Battery

Specific Gravity	Freezing Point	Specific Gravity	Freezing Point
1.275	−85°F (−65°C)	1.175	+4°F (−16°C)
1.250	−62°F (−52°C)	1.150	+5°F (−15°C)
1.225	−35°F (−37°C)	1.125	+13°F (−11°C)
1.220	−16°F (−27°C)	1.100	+19°F (−7°C)

The terminal voltage of each cell is approximately 2.12 V. The cell voltage is higher for a battery that has just been taken off charge, but in all instances it should adjust to about 2.12 V after a period of time.

Gel/Cells comes as a type A or type B. The type A Gel/Cell is conservatively designed for 4 to 6 years of continuous charging in standby power applications. During this period over 100 normal discharge/charge cycles can be expected. Even more are obtained if only minor discharges are experienced. The *end of life* is actually determined by when the equipment will no longer perform its required function. Since the battery may still have 40% to 60% of its initial capacity, the service life may be much longer.

The type A Gel/Cell has near its full nominal capacity upon shipment from the factory. Type A cells are used for alarm systems, memory standby, etc. where they are normally in a standby mode.

The type B Gel/Cell is designed to provide 3 to 5 years of service in standby power applications or 300–500 normal discharge-charge cycles in portable power applications.

As the battery is discharged, the terminal voltage will slowly decrease. For instance, when the rated capacity of the battery is removed over a 20h period, the terminal voltage would decrease to 1.75 V per cell. Batteries are rated at a 20h current rate at room temperature. This means a 2.6Ah battery would put out 0.13A for 20h. This does not mean, however, that it will put out 2.6A for 1h (it would put out about 1.7 A for 1h).

Lead-dioxide batteries can be charged by the constant-current or constant-voltage method. The constant-current method is used when charger cost is the primary consideration. The battery is forced to receive a constant amount of current regardless of its needs. While charger component economy is achieved, it is sometimes done at the expense of recharge time or service life if the current is not properly set. Trickle charging current ranges from 0.5 to 2.0mA per rated Ah capacity of the battery.

When charging with the constant-voltage method, a voltage of 2.25 to 2.30 V per cell should be used. To maintain the battery at 100°F (38°C), a voltage of 2.2V/cell is required, while at 30°F (0°C), 2.4V/cell is required.

23.10.3 Absorbed Glass Mat Batteries

Absorbed glass mat batteries (AGM) are sealed batteries that can be operated in any position. AGM was developed to provide increased safety, efficiency, and durability. In AGM batteries the acid is absorbed into a very fine glass mat that is not free to slosh around. The plates are kept only moist with electrolyte, so gas recombination is more efficient (99%). The AGM material has an extremely low electrical resistance so the battery delivers high power and efficiency. AGM batteries offer exceptional life cycles.

The plates in an AGM battery may be flat like wet cell lead-acid batteries, or they may be wound in a tight spiral. Their construction also allows for the lead in their plates to be purer as they no longer need to support their own weight. AGM batteries have a pressure relief valve that activates when the battery is recharged at voltage greater than 2.30V/cell. In cylindrical AGM batteries, the plates are thin and wound into spirals so they are sometimes referred to as spiral wound.

AGM batteries have several advantages over both gelled and flooded, at about the same cost as gelled:

- All the electrolyte (acid) is contained in the glass mats so they cannot spill or leak, even if broken. Since there is no liquid to freeze and expand, they are practically immune from freezing damage.

- Most all AGM batteries are recombinant—i.e., the oxygen and hydrogen recombine inside the battery. Using the gas phase transfer of oxygen to the negative plates to recombine them back into water while charging prevents the loss of water through electrolysis. The recombining is typically 99+% efficient.
- AGM batteries have a self-discharge of 1 to 3% per month.
- AGM batteries do not have any liquid to spill, and even under severe overcharge conditions, hydrogen emission is far below the 4% max, specified for aircraft and enclosed spaces.
- The plates in AGM's are tightly packed and rigidly mounted so they withstand shock and vibration.

23.10.4 A Comparison of the Three Types of Deep Cycle Batteries

Safety. Batteries can be dangerous. They store a tremendous amount of energy, create explosive gas during charge and discharge, and contain dangerous chemicals. Both gel and AGM batteries are sealed batteries that use recombinant gas technology. AGM is more efficient and completes its gas recombination near the plates. Gel recombinant gas batteries should incorporate automatic temperature-compensated voltage regulators to prevent explosions associated with their overcharging. Flooded batteries will spew acid, will definitely spill and leak if tipped over, and they generate dangerous and noxious explosive gases. AGM batteries are best at protecting both equipment and passengers.

Longevity. All batteries die. The number of cycles it takes to kill them is a function of the type and quality of the battery. When cycled between 25% and 50% depth of discharge (recommended deep cycle use), AGM batteries will normally outlast the other two types.

Durability. Some battery designs are simply more durable than others are. They are more forgiving in abusive conditions—i.e., they are less susceptible to vibration and shock damage, over-charging, and deeper discharge damage. Gel acid batteries are the most likely to suffer irreversible damage from overcharging. Flooded acid batteries are the most likely to suffer from internal shorting and vibration damage. AGM batteries are usually more durable and can withstand severe vibration, shocks, and fast charging.

Efficiency. Internal resistance of a battery denotes its overall charge/discharge efficiency and its ability to deliver high cranking currents without significant drops in voltage and is a measure of how well it has been designed and manufactured. Internal resistance in NiCad batteries is approximately 40%—i.e., you need to charge a NiCad battery 140% of its rated capacity to have it fully charged. The flooded wet battery's internal resistance can be as high as 26%, which is the charging current lost to gassing, or breaking up of water. Gel acid batteries are better at approximately 16% internal resistance and require roughly 116% of rated capacity to be fully charged. AGM batteries have an internal resistance of 2%, allowing them to be charged faster and deliver higher power.

23.10.5 LeClanche (Zinc-Carbon) Batteries

LeClanche batteries consist of a carbon anode, zinc cathode, and electrolyte solution of ammonium chloride, zinc chloride, and mercury chloride in water (called a *mix*). The nominal voltage is 1.5V. This type of cell is quite inefficient at heavy loads, and its capacity depends considerably on the duty cycle. Less power is available when it is used without a rest period. Maximum power is produced when it is given frequent rest periods, since the voltage drops continuously under load. Shelf life is limited by the drying out of the electrolyte. A typical discharge curve is given in Fig. 23-37.

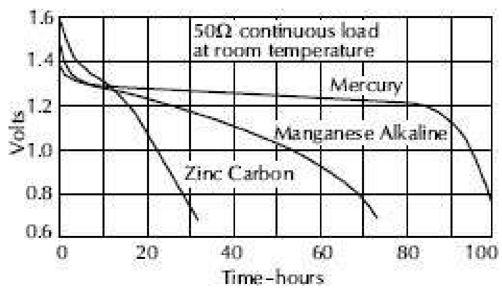


Figure 23-37. Typical discharge curves for three different types of penlight cells discharged continuously into a 50Ω load.

Zinc-carbon cells may be recharged for a limited number of cycles. The following information is extracted from the National Bureau of Standards Circular 965:

> The cell voltage for recharge must not be less than 1 V and should be recharged within a short time after removing from service. The ampere hours of charge should be within 120 to 180% of the discharge rate. The charging rate is to be low enough to distribute the recharge over 12 to 16h. Cells must be put into service soon after recharging as the shelf life is poor.

23.10.6 Nickel-Cadmium Batteries

For optimum performance, many battery-operated items require a relative constant voltage supply. In most applications, nickel-cadmium cells hold an almost constant voltage throughout most of the discharge period, and the voltage level varies only slightly with different discharge rates, Fig. 23-38. Nominal discharge voltage is 1.25V at room temperature.

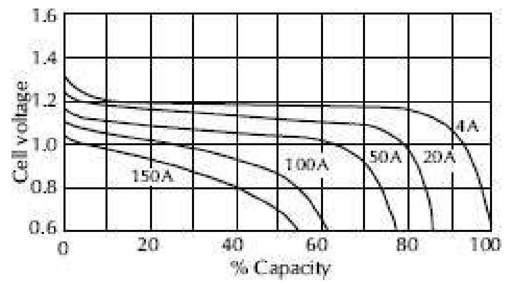


Figure 23-38. Discharge characteristic for Sonotone type 20L420 nickel-cadmium battery, rated 25Ah.

Nickel-cadmium cells are especially suited to high discharge or pulse currents because of their low internal resistance and maintenance of discharge voltage. They are also capable of recharge at high rates under controlled conditions. Many cells can be rapidly charged in 3 to 5 h without special controls, and all can be recharged at a 14h rate.

Nickel-cadmium cells are designed to operate with a wide temperature range and can be discharged from −40°F to +140°F (−40°C to +60°C).

These cells can be continuously overcharged at recommended rates and temperature. This will not noticeably affect life unless the charge rate exceeds design limitations of the cell.

The cell construction eliminates the need to add water or electrolyte, and, under certain conditions, the cell will operate on overcharge for an indefinite period. A typical discharge curve for a cell, rated at 25 Ah and weighing approximately 2lb, is shown in Fig. 23-36.

The charge retention varies from 75% for 1 month to as low as 15% for 5 months. Storage at high temperatures will reduce high retention. Cells should be charged prior to use to restore full capacity. Nickel cadmium eventually fails due to permanent or reversible cell failure. A reversible failure is usually due to shallow charge and discharge cycles and the battery appears to have lost capacity. This is often called the *memory effect*. This problem can be removed by deep discharge and a full recharge. A loss of capacity can also come from extended overcharging. If this should occur, full capacity can be restored by a discharge followed by a full recharge.

The capacity of a nickel-cadmium cell is the total amount of electrical energy that can be obtained from a fully charged cell. The capacity of a cell is expressed in ampere-hours (Ah) or milliampere-hours (mAh), which are a current-time product. The capacity value is dependent on the discharge current, the temperature of the cell during discharge, the final cutoff voltage, and the cell's general history.

The nominal capacity of the nickel-cadmium cell is that which will be obtained from a fully charged cell discharged at 68°F (20°C) for 5h to a 1.0 V cut off. This is called the C/5 rate.

Discharges at the 20, 15, 10, and 1h rates are called C/20, C/15, C/10, and C, respectively. Higher rates are designated as 2C, 3C, etc.

When three or more cells are connected in series for higher voltages, the possibility exists that during discharge, one of the cells, which may be slightly lower in capacity than the others, will be driven to a zero potential and then into reverse. At *discharge rates* (C) in the vicinity of C/10, cells can be driven into reverse without permanently damaging the cell. Prolonged, frequent, or deep reversals should be avoided since they shorten cell life or cause it to vent. Cell voltage should never be allowed to go below −0.2 V.

Nickel-cadmium batteries may be charged using either a constant-current or constant-voltage charger. There are four major factors that determine the charge rates, which can be used on nickel-cadmium batteries. They are charge acceptance, voltage, cell pressure, and cell temperature.

No charge control is required for *charge rates* up to C/3. This allows the use of the least expensive charger design. When charging rates equal or exceed 1.0 C, the charging current must be regulated to prevent overcharge.

In Table 23-13, the notation that includes the letter C is used to describe current rates in terms of a fraction of the capacity rating of the battery. A comparison of cells from different manufacturers requires rationalization to a common standard for capacity rating at the same discharge rate.

Table 23-13. Charging Rates for a Nickel-Cadmium Battery

Method of Charging		Charge Rate		
Name	Nickname	Current Rate	Fraction	Hour Rates
Standby	Trickle	0.01C	C/100	100h
		0.02C	C/50	50h
		0.03C	C/30	30h
		0.04C	C/25	25h
Slow	Overnight	0.05C	C/20	20h
		0.1C	C/10	10h
Quick	Rapid	0.2C	C/5	5h
		0.25C	C/4	4h
		0.3C	C/3	3h
Fast		C.0	C	1h
		2C.0	2C	30 min
		3C.0	3C	20 min
		4C.0	4C	15 min
		10C.0	10C	6 min

In general, discharge times will be shorter than those for C rates greater than 1 and longer than those for C rates less than 1. The charge input must always be more than discharged output. For example, to ensure full recharge of a completely discharged battery, the constant-current charge time at the 10h rate must be longer than 10 hours due to charge acceptance characteristics.

23.10.7 Nickel-Metal-Hydride Batteries

Nickel-metal hydride (NiMH) offers 40% higher energy density than NiCd batteries. NiMH is less durable than NiCd under heavy loads and service life is reduced under high temperatures.

When charging, NiMH batteries require a more complex charge algorithm, generate some heat, and require a longer charge time than a NiCd battery.

The advantages of NiMH are simple storage and transportation, high capacity and are less prone to memory.

Their disadvantages are high self discharge, limited service life (200 to 300 cycles, three year shelf life, and heavy loads reduce the batteries cycle life.

23.10.8 Alkaline-Manganese Batteries

The *alkaline-manganese battery* is gaining considerable importance in the electronic field since it is a primary battery and is rechargeable.

The polarity of this cell is reversed from the conventional zinc-carbon cell, in which the can is negative. However, because of packaging, the outward appearance is similar to the zinc-carbon cell, with the same terminal arrangement. Although this cell has an open-circuit voltage of approximately 1.5 V, it discharges at a lower voltage than the zinc-carbon cell. Also, the discharge voltage decreases steadily but more slowly. Alkaline-manganese batteries have 50–100% more capacity than their zinc counterparts. Zinc-carbon cells yield most of their energy above 1.25 V and are virtually exhausted at 1 V, while the alkaline cell yields most of its energy below 1.25V with a considerable portion released at less than 1 V.

If the discharge rate is limited to 40% of the nominal capacity of the cell and recharge is carried out over a period of 10 to 20h, alkaline-manganese cells can be cycled 50–150 times. A typical discharge curve is shown in Fig. 23-39.

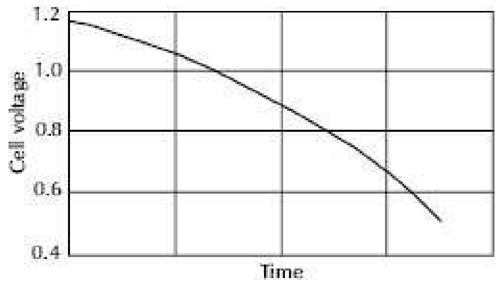


Figure 23-39. Discharge rate of an alkaline-manganese cell, on an arbitrary time scale.

23.10.9 Lithium-Ion Batteries

Lithium-Ion (Li-ion) batteries are popular. The lightest of all metals, lithium offers the great electrochemical potential and provides the large energy density per weight. Rechargeable batteries using lithium metal anodes (negative electrodes) can provide both high voltage and excellent capacity.[8]

Li-ion cells deliver 3.6–4.2V so usually perform better than series-connected battery packs with NiCd batteries. Among the different battery chemistries, Li-ion lends itself best to parallel connection.

All batteries exhibit self-discharge. Nickel-based batteries discharge 10%–15% of their capacity in the first 24 hours after charge, then 10%–15% per month. Self-discharge of Li-ion batteries is about 5% in the first 24 hours and 1%–2% thereafter. The addition of a protection circuit increases the self-discharge to 10% per month. The batteries can be stored for 20 years at room temperature loosing 5% of their charge per year.

Lithium batteries have some safety problems but are safe provided certain precautions are met when charging and discharging. Li-ion batteries require a protection circuit to limit each cell's peak voltage during charge and prevents its voltage from dropping too low on discharge, limit the maximum charge and discharge current, plus monitor the cell temperature. The batteries require more-precise charging because of safety risks associated with lithium batteries as they may leak, catch fire, or explode if improperly charged. Many charging faults can damage either the battery or the devices by allowing voltage or current levels to reach inappropriate levels that can thermally or electrically damage semiconductors, or compromise the battery’s chemical stability. Many of these faults stem from charging circuits that have direct paths from the dc power input to the charging IC, battery, and system.

Resolving input overvoltage involves an external protection circuit with an input over-voltage protection (INOVP) threshold. If the power input voltage exceeds this threshold, the protection circuit blocks the voltage to the charging system for a period called “the immunity time to block voltage surges”. If the power input voltage remains above the threshold beyond the immunity time, the protection circuit blocks the input under the assumption that the circuit is connected to a power input with an incorrect voltage rating.

An overcurrent protection (OCP) circuit must constantly monitor the power input current using a control method similar to INOVP: If the current level rises above the over-current threshold, the current is latched-off. If the current experiences a momentary spike, then keeping the current latched off unnecessarily disables the power input. Protection circuits often allow for a certain number of retries before latching off completely so a minor current issue can resolve itself while limiting system, battery, and charger damage.

Battery overvoltage protection (BOVP) requires monitoring the battery voltage directly, and disconnecting the power source when the battery voltage exceeds a predetermined threshold. Similar to OCP, BOVP allows the voltage to resume after a latch-off period to determine if a voltage spike triggered the BOVP, or if the regulator has malfunctioned, the device latches off after a certain number of retries.

To protect against reverse polarity, a diode or a MOSFET is added between the dc input and the rest of the system (preferably before any protection IC) to prevent a reverse current.

To eliminate reverse leakage, a diode is placed between the charger and the battery to prevent any reverse leakage.

Implementing INOVP, OCP, or BOVP requires adding circuits that actively monitor and rectify a particular fault, Fig. 23-40. The most cost-effective and compact solutions use charger protection ICs and inserting them between the adapter and the rest of the system.

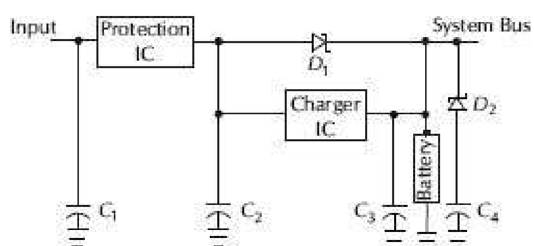


Figure 23-40. Li-ion charging circuit with protection IC. Courtesy Monolithic Power Systems.

The family of Monolithic Power Systems MP267x charger protectors integrate INOVP, OCP, and BOVP into single ICs and provide binary counters for both OCP and BOVI to latch off the current or voltage to the charging system after 16 retries.

23.10.10 Lithium-Ion Polymer Batteries

Lithium-Ion polymer uses a different electrolyte than a conventional battery. This electrolyte resembles a plastic-like film that does not conduct electricity but allows ions to exchange. The polymer electrolyte replaces the traditional porous separator, which is soaked with electrolyte. The commercial lithium-ion polymer cells are very similar in chemistry and materials to their liquid electrolyte counter parts.

Lithium-Ion polymer cells are easier to deform and damage than standard lithium-ion cells so more consideration has to be given in the device design and on how the user will access, use, and replace the battery. Nevertheless, lithium-polymer has found its market niche in thin geometries.

Lithium-polymer, with its high charge density, weighs less than the equivalent energy density of nickel-based and lithium-ion batteries. They are more resistant to overcharge, so there is less chance for electrolyte leakage. They have a low profile and can resemble the profile of a credit card and are not bound by standard cell format.

They do have limitations. They have lower energy density and decreased cycle count compared to lithium-ion. Lithium polymer cells are smaller for an equivalent output and can be charged-discharged more frequently. There are no standard sizes, most cells are produced for high volume consumer markets. Many devices that use lithium-polymer make the battery pack nonreplaceable in order to prevent potential consumer damage to the more fragile battery pack.

23.10.11 Lithium Thionyl Chloride (Li-$SOC1_2$)

Li-$SOCl_2$ cells have an anode of lithium metal and a liquid cathode. The cathode structure is a porous carbon current-collector, filled with thionyl chloride ($SOCl_2$). They are packaged in a cylindrical form factor which approximates a D-size cell and deliver 3.6 V with a capacity of better than 18Ah. They have high energy densities of as much a 1220Wh/L or 760Wh/kg. Self-discharge is less than 1% per year, so they are popular for standby use.[9]

Bobbin-type lithium thionyl chloride (Li-$SOCl_2$) batteries are the preferred choice for many remote wireless applications, delivering up to 40 years of maintenance-free service life along with the high current pulses required for powering advanced two-way communications.

Unfortunately, these batteries are not created equal, one battery brand may have an annual self-discharge rate as low as 0.75% per year while others have a higher annual self-discharge rate of 2.5% to 3%, resulting in shorter service life.

Wireless devices increasingly require high current pulses to power advanced two-way communications. To conserve energy, these devices operate mainly in a “dormant” phase, periodically switching into “active” mode that may require current pulses of several amps for data acquisition and transmission.

If the application involves dormant phases at elevated temperatures, alternating with periodic high current pulses, lower transient voltage readings, called transient minimum voltage (TMV) can occur during the initial phases of battery discharge. This is correlated to the make-up of the electrolyte and/or the cathode.

A hybrid lithium thionyl chloride battery, known as PulsesPlus combines a bobbin-type cell with a patented hybrid layer capacitor (HLC). The HLC stores and generates short-term high current pulses. This type of battery is useful for applications like earthquake sensors that may require a battery that remains dormant for decades then delivers very high current pulses.

23.10.12 Lithium Sulfur Dioxide (Li-SO_2)

Li-SO_2 cells are like Li-$SOC1_2$ cells, except the solution in the porous carbon current collector is sulfur dioxide (SO_2). They come in the same range of standard cylindrical battery sizes but deliver a lower voltage of 2.8 V, rather than 3.6 V, and a lower energy density of 250Wh/kg.

23.10.13 ZPower Batteries

ZPower batteries have an intrinsically safe water-based chemistry that contains no lithium or flammable solvents. Unlike Li-ion and Li-polymer batteries, these batteries are free from the problems of thermal runaway, fire, and danger of explosion. They're also free from the regulations that limit the size of lithium-containing batteries on airplanes.

Besides offering high-performance and safety, ZPower batteries use an environmentally friendly chemistry that allows recycling of battery cells and reuse of contents. Unlike other traditional batteries, ZPower batteries have no heavy metals and no toxic chemicals.

23.10.14 Mercury Dry Cells

The *mercury dry cell* using a zinc-mercury oxide alkaline system was invented by Samuel Ruben during World War II. There are two kinds of mercury cells: one with a voltage of 1.35 V and one with 1.4

V.

The 1.35 V cell has a pure mercuric-oxide cathode. On discharge its voltage drops only slightly until close to the end of the cell life when it then drops rapidly. The 1.4 V cell has a cathode of mercuric oxide and manganese dioxide. On discharge, its voltage is not quite as well regulated as the 1.35 V cell, but it is considerably better than the manganese-alkaline or zinc-carbon cell.

Mercury cells have excellent storage stability. A typical cell will indicate a voltage of 1.3569 V, with a cell-to-cell variation of only 150μV. Variation due to temperature is 42μV/°F, ranging from −70°F to +70°F (−56°C to +21°C), with a slight increase of voltage with temperature. The internal resistance is approximately 0. 75Ω. Voltage loss during storage is about 360μV per month; therefore, a single cell can be used as a reference voltage of 1.3544V ±0.17%. The voltage is defined under a load condition of 5% of the maximum current capacity of the cell. Normal shelf life is on the order of three years.

Recharging of mercury cells is not recommended because of the danger of explosion. A typical stability curve for a single cell over a period of 36 months is shown in Fig. 23-41. The drop in voltage over this period is 13 mV.

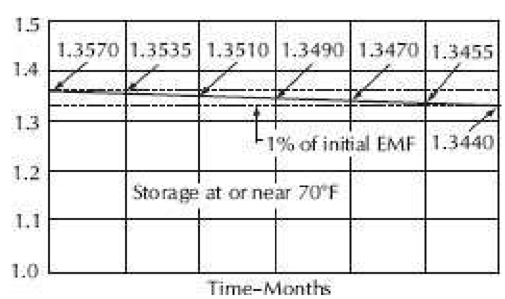


Figure 23-41. Stability curve for a single-cell mercury battery.

23.11 Battery Characteristics

Batteries offer a means of producing a smooth, ripple-free, hum-free, portable power supply. A battery's capacity is rated in ampere-hours (Ah). Three facts about batteries are:

1. An ampere-hour can be a 1 A drain for 1 h, 0.5 A drain for 2h, or 2A drain for 0.5h, etc.
2. A 12 V liquid battery is generally considered completely discharged when its voltage reaches 10.5 V.
3. Batteries for cycling service—i.e. powering amplifiers etc.—are normally rated with:
   - A 20h discharge rate.
   - A reserve capacity stated in minutes for a 25 A discharge rate.

In electronic circuits the cathode is considered the negative terminal, in batteries the positive terminal is called the cathode. This makes sense in batteries because the positively charged ions in the battery are attracted to the cathode and the negatively charged ions are attracted to the anode. Inside and outside the battery electron current flows in a consistent direction.

A cell or battery is an electrochemical system that converts chemical energy into electrical energy. When the chemical action is reversible, the battery is a secondary or rechargeable system.

To be rechargeable, the positive and negative electrodes of a battery must be capable of being converted back to their original state following a discharge. Thus, the battery must be electrically recharged by reversing the process that occurred during its discharge cycle.

23.11.1 Temperature Effects

The standard rating for batteries is at 25°C (77°F). Battery capacity is reduced at lower temperature. At freezing, Ah capacity is reduced to 80%. At −27°C (−22°F), Ah capacity drops to 50%. At 122°F, capacity is increased by 12%.

Battery charging voltage is also affected by temperature. It will vary from about 2.74V/cell (16.4 V for a 12V battery) at −40°C (−40°F) to 2.3V/cell (13.8 V) at 50°C (122°F).

Temperature also affects battery life. While battery capacity is reduced by 50% at −22°F, battery life increases about 60%. For every 8.3°C (15°F) over 25°C (77°F), battery life is cut in half. This holds true for all types of lead-acid batteries, sealed, gelled, and AGM.

23.11.2 Cycles versus Battery Life

A battery cycle is one complete discharge and recharge cycle and is often considered a discharge from 100% to 20%, and then recharged back to 100%. Other ratings for depth of discharge (DOD) cycles are 10%, 20%, and 50%.

Battery life is directly related to how deep the battery is cycled each time. If a battery DOD is 50% every cycle, it will last twice as long as if the DOD is 80%. If the DOD cycle is only 10%, it will last about five times as long as one cycled to 50%. A 50% DOD is usually recommended. A battery that has a DOD cycle of 5% or less usually does not last as long as one cycled down 10% because at very shallow cycles, the lead dioxide tends to build up in clumps on the positive plates rather in an even film.

23.11.3 Battery Voltage

All lead-acid batteries supply about 2.14V/cell, or 12.6 to 12.8V for a 12 volt battery when fully charged.

Batteries that are stored for long periods will eventually self-discharge. This varies with battery type, age, and temperature. Self-discharge can range from 1%–15% per month. Batteries should never be stored in a partly discharged state for a long period of time. A float charge should be maintained if they are not used.

23.11.4 State of Charge

State of charge, or conversely, the depth of discharge (DOD), can be determined by measuring the voltage and/or the specific gravity of the acid with a hydrometer. Voltage on a fully charged battery is 2.12 V to 2.15V/cell, or 12.7 V for a 12 volt battery. At 50% DOD the voltage is 2.03V/cell, and at 0% DOD it is 1.75V/cell or less. Specific gravity is 1.265 for a fully charged cell and 1.13 or less for a totally discharged cell. Many batteries are sealed, therefore, hydrometer reading cannot be taken.

23.11.5 False Capacity

A battery can meet all the tests for being at full charge, yet be lower than its original capacity because the plates are damaged, sulfated, or partially gone from long use. In this case it acts like a battery of much smaller size.

23.11.6 Ampere-Hour Capacity

Deep cycle batteries are rated in ampere hours (Ah). An Ah is a 1A drain for 1h, 10A for 0.1h, etc. It is calculated with the equation A ×

h. Drawing 20A for 20min would be 20A × 0.333h, or 6.67Ah. The accepted Ah rating time period for batteries used in solar electric and backup power systems and for nearly all deep cycle batteries is the 20h rate. This is defined as the battery being discharged to 10.5 V over a 20h period while the total actual Ah it supplies is measured.

23.11.7 Battery Charging

Batteries can be charged by constant current or constant voltage. When charged by the constant-current method, care must be taken to eliminate the possibility of overcharging; therefore, the condition of the battery should be known before charging so that the charger can be removed when the ampere-hour rate of the battery is met.

Charging with the constant voltage method reduces the possibility of overcharging. With the constant voltage method, charge current is high initially and tapers off to a trickle charge when the battery is fully charged. Two requirements must be met when using the constant-voltage method:

- The charging voltage must be stable and set to 2.4 V per cell for a lead-acid battery and 2.30 V per cell for a gel cell battery. Gel cell open-circuit voltage is 2.12 V per cell.
- A current-limiting circuit must be employed to limit charge current when the battery is fully discharged.

A good battery charger charges in three steps. In the first stage, charge current is at the maximum safe rate the batteries will accept until the voltage rises to 80–90% of full charge level. Voltages at this stage typically range from 10.5V to 15V. There is no correct voltage for bulk charging, but there may be limits on the maximum current that the battery and/or wiring can accept.

In the second stage, accept, the voltage remains constant and current gradually tapers off as internal resistance increases during charging. Voltages are typically 14.2 V to 15.5 V.

After batteries reach full charge, the third stage charging voltage is reduced to a lower level, 12.8V to 13.2 V, to reduce gassing and prolong battery life. This is often referred to as a *maintenance, float*, or *trickle charge*, since its main purpose is to keep an already charged battery from discharging, Fig. 23-42. An ideal charging state table is shown in Table 23-14.

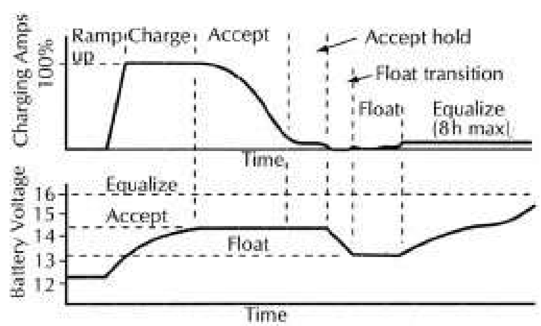


Figure 23-42. Ideal charge curve.

PWM, or *pulse width modulation* is sometimes used as a float or trickle charge. In PWM chargers, the controller circuit senses small voltage drops in the battery and delivers short charging cycles (pulses) to the battery. This may occur several hundred times per minute and is called *pulse width* because the width of the pulses varies from a few microseconds to several seconds.

Most flooded batteries should be charged at no more than the C/8 rate for any sustained period. C/8 is the battery capacity at the 20h rate divided by 8. For a 220Ah battery, this would equal 26A. Gelled cells should be charged at no more than the C/20 rate, or 5% of their amp-hour capacity. AGM batteries can be charged at up the

C × 4 rate, or 400% of the capacity for the bulk charge cycle.

Lead acid batteries require 15.5 V for 100% charge. When the charging voltage reaches 2.583V/cell, charging should be stopped or reduced to a trickle charge. Flooded batteries must bubble (gas) to insure a full charge, and to mix the electrolyte. Float voltage for flooded batteries should be 2.15V to 2.23V/cell, or 12.9V to 13.4V for a 12 volt battery. At higher temperatures, over 85°F, charge voltage should be reduced to 2.10V/cell. Float and charging voltages for gelled batteries are usually about 0.2V less than for flooded batteries.

Table 23-14. Ideal Charging State

Cycle	Voltage	Current
Charge	12.0-14.3 Rising	Maximum
Accept	14.4 Constant	Falling
Float	13.5 Constant	Small (<2% capacity)
Equalize	13.2-16.0 Rising	Constant until 16.0V

23.11.7.1 Equalizing

The equalize cycle puts a controlled overcharge to remove lead sulfate from the plates that is not removed during the normal charging of the battery. Flooded battery life can be extended if an equalizing charge is applied every month or so. This is a charge that is about 10% higher than the normal full charge voltage, and is applied for about 8 hours to be sure that all cells are equally charged and the gas bubbles mix the electrolyte. If the liquid in the cells is not mixed, the electrolyte becomes stratified, creating a strong solution at the top and weak solution at the bottom of the cell. AGM and gelled batteries should be equalized a maximum of two to four times a year.

23.11.7.2 Charging Voltage versus Temperature

Battery charging is sensitive to temperature. As the ambient temperature decreases, the charging voltage must be increased, Table 23-15.

23.11.7.3 State of Charge

Table 23-16 shows no-load typical voltages versus state of charge for a 12 V battery. These voltages are for batteries that have been at rest for 3 hours or more. Note the large voltage drop in the last 10%.

Table 23-15. Temperature Compensation for Various Types of Batteries

Temp		Liquid		Gel—Std.		Gel—Fast		AGM	
°F	°C	Accept	Float	Accept	Float	Accept	Float	Accept	Float
120	49	12.5	12.5	13.0	13.0	13.0	13.0	12.9	12.9
110	43	13.6	12.7	13.5	13.0	14.0	13.4	13.9	12.9
100	38	13.8	12.9	13.7	13.2	14.1	13.5	14.0	13.0
90	32	14.0	13.1	13.8	13.3	14.2	13.6	14.1	13.1
80	27	14.2	13.3	14.0	13.5	14.3	13.7	14.2	13.2
70	21	14.4	13.5	14.1	13.6	14.4	13.8	14.3	13.3
60	16	14.6	13.7	14.3	13.8	14.5	13.9	14.4	13.4
50	10	14.8	13.9	14.4	13.9	14.6	14.0	14.5	13.5
40	5	15.0	14.1	14.6	14.1	14.7	14.1	14.6	13.6
30	−1	15.2	14.3	14.7	14.2	14.8	14.2	14.7	13.7

Table 23-16. No Load Voltage versus State of Charge for a 12V Battery

State of Charge	12 Volt battery	Volts per Cell
100%	12.7	2.1
90%	12.5	2.1
80%	12.4	2.1
70%	12.3	2.1

60%	12.2	2.0
50%	12.1	2.0
40%	11.9	2.0
30%	11.8	2.0
20%	11.6	1.9
10%	11.3	1.9
0%	10.5	1.8

23.11.7.4 Internal Resistance

All batteries have internal resistance that causes the battery voltage to fluctuate with the load. To calculate the internal resistance of a single cell or battery, the open-circuit voltage V_1 is measured using a voltmeter with an internal resistance of at least 1000 Ω/V. The battery or cell is then loaded with resistor R_1, and the voltage V_2 across the resistor is measured. R_1 should be at least 10 times the battery resistance. The current through the resistor R_1 is

$$I = \frac{V_2}{R_1} \qquad (23\text{-}29)$$

The internal resistance, R_i, of the battery may now be calculated using

$$R_i = \frac{V_1}{I} - R_1 \qquad (23\text{-}30)$$

23.11.7.5 Wireless Battery Charging

Batteries provide power to a variety of applications. Often a charging connector is difficult or impossible to use such as products that require sealed enclosures to protect sensitive electronics from harsh environments or to allow for convenient cleaning or sterilization. Other products are too small to include a connector, or the battery-powered application includes movement or rotation.

Three standards organizations are vying for acceptance in wireless charging of batteries in portable devices: the Alliance of Wireless Power (A4WP), the Power Matters Alliance (PMA), and the Wireless Power Consortium (WPC). The IEEE is also involved with its Wireless Power and Charging Systems Working Group (WPCS-WG).[10]

All of these charging methods are based on inductive coupling using a transmitter driving a primary winding in a charging base which creates a magnetic field that induces a voltage into a nearby secondary winding and receiver inside a portable electronic device. Most configurations use a flat charging base with one or more transmitter coils which the device to be charged lies on. The PMA and WPC standards use a 100kHz to 200kHz ac frequency while the A4WP–standard uses a more efficient resonant solution on the 6.78MHz industrial, scientific, and medical (ISM) band frequency.

First-generation Qi (pronounced “Chee”) products required a tight inductive coupling (magnetic induction) between the transmitter/receiver pair and could only charge one device at a time. They required specific positioning on the pad, and had to be in contact with the pad to charge. Loose inductive coupling that seeks to supplant it, uses magnetic resonance, offering inches of charging range, simultaneous charging of multiple devices, even those with different power requirements, and free-form positioning, ideal for charging vehicles and wireless charging in countertops, etc.

Competing wireless charging approaches have until now maintained a unique identity, but differences diminish as we move to next-gen products. It is not clear whether the three main standards will converge. However, magnetic resonance technology will play a starring role in the future.

As shown in Fig. 23-43, a wireless power system is composed of two parts separated by an air gap:

1. The transmit circuitry, including a transmit coil.
2. The receive circuitry, including a receive coil.

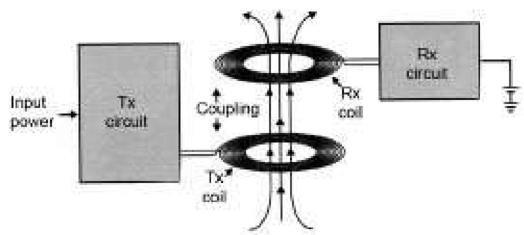


Figure 23-43. Wireless battery charger diagram. Courtesy Linear Technology.

The transmit circuitry generates a high frequency alternating magnetic field around the transmit coil which is coupled to the receive coil and converted to electrical energy to charge a battery or power other circuitry.

A key parameter is the amount of charging power that actually adds energy to the battery and depends on the amount of power being transmitted, the distance and alignment between the transmit coil and the receive coil, also known as the coupling between the coils, and the tolerance of the transmit and receive components.

The primary goal is to delivery of the required power under worst-case power transfer conditions while avoiding thermal and electrical overstress in the receiver which occurs when the output power requirements are low such as when the battery is fully charged or nearly fully charged. This excess power leads to high rectified voltages and a need to dissipate the excess power as heat.

To remedy this problem, the rectified voltage can be clamped with a power Zener diode or transorb which unfortunately is usually large and generates considerable heat. Without feedback from the receiver, the maximum transmitter power can be reduced, but this will limit the available received power or reduce the transmit distance. It is possible to communicate received power back to the transmitter and adjust real-time transmit power accordingly which is used by the Wireless Power Consortium Qi standard.

However, it is also possible to solve this issue in a compact and efficient manner without resorting to complicated digital communication techniques. To efficiently manage the power transfer from transmitter to receiver under all conditions the LTC4120 wireless power receiver by Linear Technology integrates technology including PowerbyProxi's patented dynamic harmonization control, or DHC, a technique that enables high-efficiency contact-less charging without thermal or electrical overstress concerns in the receiver. Using this technology, up to 2W can be transmitted at a distance of up to 1.2cm.

By modulating the resonant frequency of the receiver from a "tuned" condition to a "detuned" condition, DHC guarantees delivery of power under worst-case conditions without worrying about unloaded best-case conditions. This allows the LTC4120-based wireless charging system to operate over a wide transmit distance with significant coil misalignment. Furthermore, by controlling power transfer on the receiver-side only, the LTC4120-based system eliminates all potential communication interference issues, which might disrupt power delivery.

When choosing a transmitter, several factors should be considered. Is transmitter standby power (when a receiver is not present) important? Does the transmitter need to differentiate between a valid receiver and unrelated foreign metal objects? How sensitive is surrounding circuitry to EMI?

The basic transmitter is a very simple solution. Due to passive resonant filtering, the spectrum of EMI is well controlled at the fundamental transmitter frequency (about 130kHz). However, it transmits at full power whether a receiver is present or not, therefore its standby power is relatively high. It also does not differentiate between a receiver and foreign metal objects and can cause unrelated metal objects to warm up through induced eddy currents.

Transmitters manufactured by PowerbyProxi have transmit distance and alignment tolerance performance virtually identical to that of the basic transmitter an can detect whether a valid receiver is present or not. This feature allows them to reduce standby power if no receiver is present and terminate power transmission if unrelated foreign metal objects are nearby.

References

1. Fuse Characteristics, Terms and Consideration Factors, Littelfuse, Inc., 2009.

2. M. Simon, Everything you need to know about picking the right circuit protection, *ECN*, September 2013.

3. Don Tuite, Controller Lets Designers Replace Rectifier Bridge Diodes With MOSFETs, *Electronic Design*, August 8, 2013.

4. David Norton, Power Supply “Remote Sense” Mistakes and Remedies, *EDN*, July 10, 2013.

5. Synchronous Rectification Aids Low-Voltage Power Supplies, *Maxim Application Note 652*, Jan. 31, 2001, Maxim Integrated Products.

6. Bob Zollo, What Is An Autoranging Power Supply?, *Electronic Design*, March 7, 2013.

7. Kerry Lanza, What is the Best Type of Battery, *Electronic Products*, March 2011.

8. Min Xu, Safety and the Lithium Battery, *Electronic Products*, December 2012.

9. Sol Jacobs, Lithium Batteries and energy Harvesting, *ENC*, August, 2013.

10. Trevor Barcelo, Charging Batteries Efficiently Without Wires, *Electronic Products*, March, 2014.

Chapter 24

Amplifier Design

by Bob Cordell

24.1 Introduction

Amplifiers of all sorts are at the heart of audio systems. In this chapter the circuit building blocks of amplifiers will be explained and the applications of different types of amplifiers will be illustrated. In many audio systems, very small signals need to be brought up to a nominal “line” operating level, often on the order of a volt. These signals are often then manipulated by gain or volume controls and other functions like tone controls and filters. Finally, the line-level signals must be boosted by a power amplifier to provide the high voltages and currents required to drive the loudspeakers.

Fig. 24-1 illustrates a typical consumer audio amplifier and processing chain. The input to the chain includes signal sources, like a CD player, an audio DAC or a magnetic phono cartridge. Low-level signal sources require an additional pre-amplification stage like a phono preamp or a microphone preamp to boost the input signal up to the line level. Such sources often produce signals on the order of millivolts. Some must include equalization, like the RIAA equalization for a phono preamp.

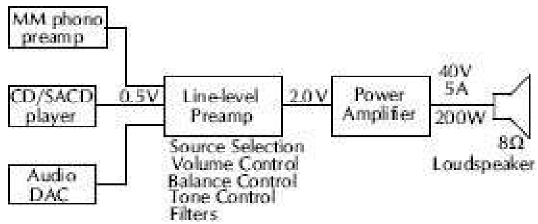


Figure 24-1. Consumer audio processing chain.

The preamplifier includes switching to select the desired input, signal level controls to set the volume, and often a left-right balance control. Some preamplifiers include bass and treble controls to alter the frequency response to the listener's taste. Apart from any included low-level preamplifier sections, the preamplifier provides little or no gain. Line-level signals enter the preamplifier and line-level signals exit the preamplifier.

The power amplifier brings the line-level signal up to the amplitude required to drive the loudspeaker system. It must deliver tens of volts at up to tens of amps to the loudspeaker load.[1] Rated loudspeaker load impedance is typically 8 Ω, but nominal values include 2Ω, 4Ω and occasionally 16 Ω. A 100W amplifier must deliver 28Vrms and 3.5A to an 8 Ω load. An 800W amplifier must deliver 80 V and 10A into an 8 Ω load. 800W into a 2 Ω load corresponds to 40 V and 20A.

The loudspeaker is an integral part of the signal processing chain, and it can influence the performance of, and demands placed on, the power amplifier. While a loudspeaker may have a nominal rating of 8Ω, the actual impedance often varies considerably over frequency; it might fall as low as 3 Ω and rise as high as 50 Ω. These impedance dips and peaks often occur at woofer resonance frequencies and crossover frequencies. The output current demands placed on the amplifier thus may be considerably greater than the 3.5A cited for a 100W amplifier driving an 8 Ω resistive load to 100W.

Sometimes an equalizer is placed in front of the power amplifier to alter the frequency response to compensate for room acoustics, loudspeaker shortcomings or listening taste. Active crossovers may also be placed ahead of a multiplicity of amplifiers where each amplifier may be dedicated to a particular driver section of the loudspeaker, such as the woofer, midrange and tweeter. Professional audio signal chains are not all that different, but they usually include numerous microphone inputs and employ a mixing desk instead of a preamplifier.

24.2 Transistors

The *Bipolar Junction Transistor* (BJT) is the primary building block of most audio amplifiers. If a small current is sourced into the base of an NPN transistor, a much larger current flows in the collector. The ratio of these two currents is the current gain, commonly called beta (β) or h_{fe}. Similarly, if one sinks a small current from the base of a PNP transistor a much larger current flows in its collector. The current gain for a typical small-signal transistor often lies between 50 and 200. For an output transistor, β typically lies between 10 and 100. Beta can vary quite a bit from transistor to transistor and is also a mild function of the transistor current and collector voltage.

24.2.1 Transistor Characteristics

Fig. 24-2 illustrates the output current of a transistor as a function of collector-emitter voltage (V_{ce}) for several different values of base current. The upward slope of each curve with increasing V_{ce} reveals the mild dependence of β on collector-emitter voltage. The spacing of the curves for different values of base current reveals the current gain. Notice that this spacing tends to increase as V_{ce} increases, once again revealing the dependence of current gain on V_{ce}, called *Early effect*. The spacing of the curves may be larger or smaller between different pairs of curves. This illustrates the dependence of current gain on collector current. The transistor shown has β of about 50.

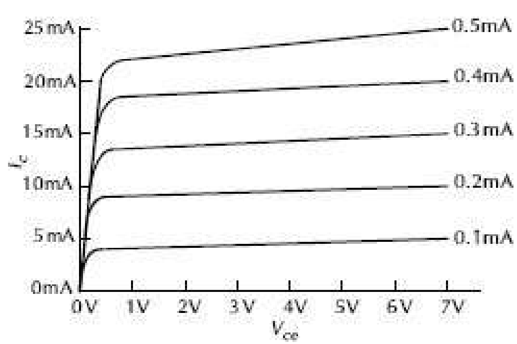


Figure 24-2. Transistor collector current characteristic.

Because transistor β can vary quite a bit, circuits are usually designed so that their operation does not depend heavily on the particular value of β for its transistors. Rather, the circuit is designed so that it operates well for a minimum value of β and for very high β. Because β can sometimes be very high, it is usually bad practice to design a circuit that would misbehave if β became very high.

The bipolar junction transistor requires a certain forward bias voltage at its base-emitter junction to begin to conduct collector current. This turn-on voltage is usually referred to as V_{be}. For silicon transistors, V_{be} is usually about 0.6 V. The actual value of V_{be} depends on the transistor device design and the amount of collector current (I_c).

The base-emitter voltage increases by about 60mV for each decade of increase in collector current. This reflects the logarithmic relationship of Vb_e to collector current. For the popular 2N5551, for example, Vb_e = 600mV at 100μA and rises to 720mV at 10mA. This corresponds to a 120mV increase for a two-decade (100:1) increase in collector current.

24.2.2 The Gummel Plot

If the log of collector current is plotted as a function of Vb_e, a very revealing picture results.[2] It is ideally a straight line. The diagram becomes even more useful and insightful if base current is plotted on the same axes. This is now called a *Gummel* plot, shown in Fig, 24-3.

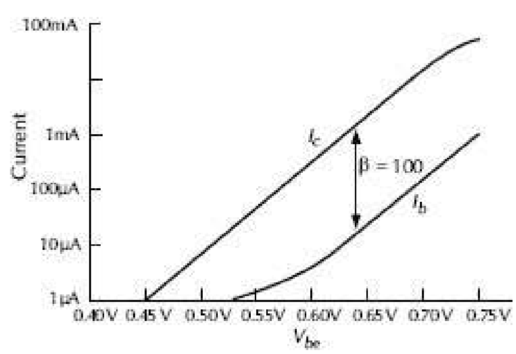


Figure 24-3. Transistor Gummel plot.

In practice, neither the collector current nor the base current plots are straight lines over the full range of V_{be}, and the bending illustrates various non-idealities in the transistor behavior. The vertical distance between the lines corresponds to the β of the transistor, and the change in distance shows how β changes as a function of Vb_e and, by extension, I_c. The curves in Fig. 24-3 illustrate the typical loss in transistor current gain at both low and high current extremes.

24.2.3 Collector Current vs. Base-emitter Voltage

Small amounts of collector current actually begin to flow at quite low values of forward bias (V_{be}). Indeed, the collector current increases exponentially with Vb_e. That is why it looks like there is a fairly well-defined turn-on voltage when collector current is plotted against Vb_e on linear coordinates. It becomes a remarkably straight line over many decades of collector current when the log of collector current is plotted against Vb_e. Some circuits, like multipliers, make great use of this logarithmic dependence of Vb_e on collector current. Collector current increases exponentially with base-emitter voltage, and we have the approximation:

$$I_c = I_S e^{V_{be}/(V_T)} \qquad (24\text{-}1)$$

where the voltage V_T is called the *thermal voltage*.[2] Here V_T is about 26mV at room temperature, and is proportional to absolute temperature. This plays a role in the temperature dependence of Vb_e. However, the major cause of the temperature dependence of Vb_e is the strong increase with temperature of the *saturation current* I_S. This ultimately results in a negative temperature coefficient of Vb_e of about −2.2mV/°C.

24.2.4 Transconductance

The *transconductance* (*gm*) of the transistor describes how much the collector current changes when the base voltage is changed a small amount, when the transistor is operated at a given dc bias current. Thus $gm = \Delta I_c \Delta V_{be}$. Transconductance is actually the more predictable and important design parameter (as long as β is high enough not to matter much). Transconductance has the units of *Siemens*, S (amps per volt). If the base-emitter voltage of a transistor is increased by 1 mV, and the collector current increases by 39 μA, the *gm* of the transistor is 39 milli-Seimens (mS).

While transistor current gain is an important parameter and largely the source of its amplifying ability, the transconductance of the transistor is perhaps the most important characteristic used by engineers when doing actual design. It is sometimes easier to visualize circuit operation and design by taking account of finite *gm* by instead using what may be called the intrinsic emitter resistance $re' = 1/gm$.[1] We can then think of an "internal" emitter that is moving completely with the base voltage, that emitter then being in series with re'. We will see this concept used in the next section.

24.2.5 Hybrid π Transistor Model

Those more familiar with transistors will recognize that much of what has been discussed above contributes to the *hybrid π* small-signal model of the transistor, shown in Fig. 24.4.[2] The fundamental active element of the transistor is a voltage-controlled current source; namely a transconductance (gm). Everything else in the model is essentially a passive *parasitic* component. Small-signal current gain is taken into account by the base-emitter resistance r_π. Early effect is taken into account by *ro*. Collector-base capacitance is shown as C_cb. Current gain roll-off with frequency (f_T) is modeled by C_π. Because this is a small-signal model, element values will change with the operating point of the transistor.

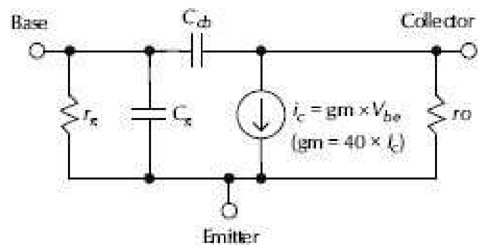


Figure 24-4. Hybrid π transistor model.

24.2.6 JFET Characteristics

JFETs operate on a different principle than BJTs. Picture a bar of N-type doped silicon connected from source to drain. This bar will act like a resistor. Now add a PN junction somewhere along the length of this bar by adding a region with P-type doping. This is the gate. As the P-type gate is reverse biased, a *depletion region* will be formed and this will begin to pinch off the region of conductivity in the N-type bar. This reduces current flow. This is called a *depletion-mode* device. The JFET is nominally on and its degree of conductance will decrease as reverse bias on its gate is increased until the channel is completely pinched off.

The reverse gate voltage where pinch-off occurs is referred to as the *threshold voltage* V_t. The threshold voltage is often on the order of −0.5 V to −4 V for most small-signal N-channel JFETs. Note that control of a JFET is opposite to the way a BJT is controlled. The BJT is normally off and the JFET is normally on. The BJT is turned on by application of a forward bias to the base-emitter junction, while the JFET is turned off by application of a reverse bias to its gate-source junction.

The reverse voltage that exists between the drain and the gate can also act to pinch off the channel. At V_{dg} greater than V_t, the channel will be pinched in such a way that the drain current becomes self-limiting. In this region the JFET no longer acts like a resistor, but rather like a voltage-controlled current source. These two operating regions are referred to as the *linear region* and the *saturation region*, respectively. JFET amplifier stages usually operate in the saturation region.

Fig. 24-5a shows how drain current changes as a function of gate voltage in the saturation region, while Fig. 24-5b illustrates how transconductance changes as a function of drain current in the same region. The device shown here is one-half of a Linear Integrated Systems LS844 dual JFET. Threshold voltage for this device is nominally about −1.8 V.

The JFET I-V characteristic (I_d vs. V_{gs}) obeys a square law, rather than the exponential law applicable to BJTs.[2] The simple relationship below is valid for $Vds > V_T$ and does not take into account the influence of Vd_s that is responsible for output resistance of the device.

$$I_d = \beta(V_{gs} - V_t)^2 \qquad (24\text{-}2)$$

The equation is valid only for positive values of ($V_{gs} - V_t$). The factor β governs the transconductance of the device. When $V_{gs} = V_t$, the $V_{gs} - V_t$ term is zero and no current flows. When $V_{gs} = 0$ V, the term is equal to V_t^2 and maximum current flows.

Notice that *gm* for this JFET is about 2mS at 1 mA. This contrasts with the approximate 39mS of a BJT at 1mA. The *gm* of a JFET is often 1/10 or less than that of a BJT operating under similar conditions.

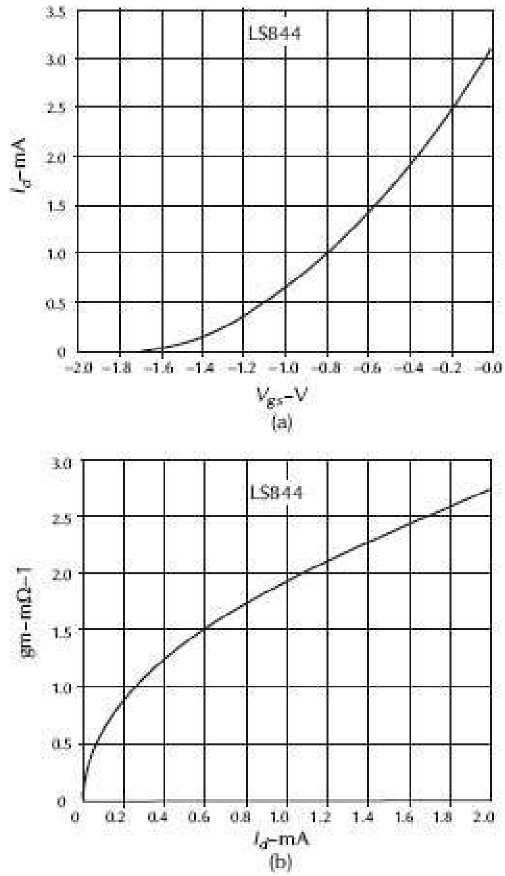


Figure 24-5. JFET drain current and transconductance curves.

24.3 Amplifier Circuits

An audio amplifier is composed of just a few important circuit building blocks put together in many different combinations. Once each of those building blocks can be understood and analyzed, it is not difficult to do an approximate analysis by inspection of a complete amplifier. Knowledge of how these building blocks perform and bring performance value to the table permits the designer to not only analyze circuits, but to also synthesize circuits.

24.3.1 Common Emitter Stage

The *common emitter* (CE) amplifier is possibly the most important circuit building block, as it provides basic voltage gain. Assume that the transistor's emitter is at ground and that a bias current has been established in the transistor. If a small voltage signal is applied to the base of the transistor, the collector current will vary in accordance with the base voltage. If a load resistance is provided in the collector circuit, that resistance will convert the varying collector current to a voltage. A voltage-in, voltage-out amplifier is the result, and it likely has quite a bit of voltage gain. A simple common emitter amplifier is shown in Fig. 24-6.

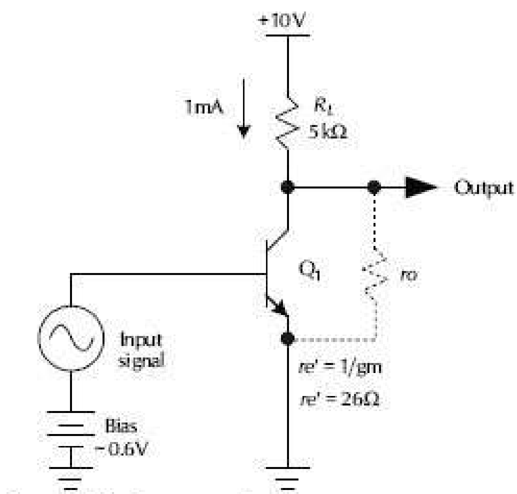


Figure 24-6. Common emitter stage.

The voltage gain will be approximately equal to the collector load resistance times the transconductance, *gm*. Recall that the intrinsic emitter resistance $re' = 1/gm$. Thus, more conveniently, assuming the ideal transistor with intrinsic emitter resistance re', the gain is simply R_I/re'.[1]

Consider a transistor biased at 1 mA with a load resistance of 5000 Ω and a supply voltage of 10 V, as shown in Fig. 24-6. The transconductance will be about 39mS and the intrinsic emitter resistance re' will be about 26 Ω. The gain will be approximately

5000/26 = 192. This is quite a large value. However, any loading by other circuits that are driven by the output has been ignored. Such loading will reduce the gain.

The *Early effect* has also been ignored. It effectively places another resistance *ro* in parallel with the 5000Ω load resistance. This is illustrated by the dashed resistor drawn in the figure. The value of *ro* for a 2N5551 operating at 1mA will be on the order of 100kΩ, so the error introduced by ignoring Early effect here will be about 5%. The value of *ro* is a function of the operating point of the transistor, so it can change with signal swing and thus introduce distortion.

Because *re′* is a function of collector current, the gain will vary with signal swing and the gain stage will suffer from some distortion. The gain will be smaller as the collector current swings low and the output voltage swings high. The gain will be larger as the current swings high and the output voltage swings low. This results in second harmonic distortion.

If the input signal swings positive so that the collector current increases to 1.5 mA and the collector voltage falls to 2.5 V, *re′* will be about 17.3 Ω and the incremental gain will be 5000/17.3 = 289. If the input signal swings negative so that the collector current falls to 0.5 mA and the collector voltage rises to 7.5 V, then *re′* will rise to about 52 Ω and incremental gain will fall to 5000/52 = 96. The incremental gain of this stage has thus changed by over a factor of three when the output signal has swung 5Vp-p. This represents a high level of distortion.

24.3.1.1 Emitter Degeneration

If external emitter resistance is added as shown in Fig. 24-7, then the gain will simply be the ratio of R_L to the total emitter circuit resistance consisting of re' and the external emitter resistance R_e. Since the external emitter resistance does not change with signal, the overall gain is stabilized and is more linear. This is called emitter degeneration. It is a form of local negative feedback.

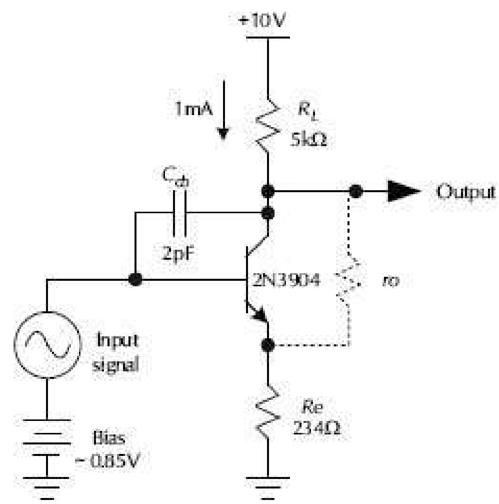


Figure 24-7. CE stage with emitter degeneration.

The CE stage in Fig. 24-7 is essentially the same as that in Fig. 24-6 but with a 234Ω emitter resistor added. This corresponds to 10:1 emitter degeneration because the total effective resistance in the emitter circuit has been increased by a factor of 10 from 26 Ω to 260 Ω. The nominal gain has also been reduced by a factor of 10 to a value of approximately 5000/260 = 19.2.

Consider once again what happens to the gain when the input signal swings positive and negative to cause a 5Vp-p output swing. If the input signal swings positive so that the collector current increases to 1.5 mA and the collector voltage falls to 2.5 V, total emitter circuit resistance R_e will become 17 + 234 = 251Ω and the incremental gain will rise to 5000/251 = 19.9.

If the input signal swings negative so that the collector current falls to 0.5mA and the collector voltage rises to 7.5 V, then R_e will rise to about 234 + 52 = 287Ω and incremental gain will fall to 5000/287 = 17.4. The incremental gain of this stage has now swung over a factor of 1.14:1, or only 14% when the output signal has swung 5 Vp-p. This is indeed a much lower level of distortion than occurred in the un-degenerated circuit of Fig. 24-6. This illustrates the effect of local negative feedback without resort to any negative feedback theory. We thus have, for the CE stage, the approximation:

$$\text{Gain} = \frac{R_L}{(re' + R_e)} \qquad (24\text{-}3)$$

where,
R_L is the net collector load resistance,
R_e is the external emitter resistance.

The emitter degeneration factor is defined as

$$\frac{re' + R_e}{re'}$$

In this case that factor is 10:1.

24.3.1.2 Early Effect

Emitter degeneration also mitigates nonlinearity caused by the Early effect in the CE stage. As shown by the dotted resistance *ro* in Fig. 24-7, most of the signal current flowing in *ro* is returned to the collector by way of being injected into the emitter. If 100% of the signal current in *ro* were returned to the collector, the presence of *ro* would have no effect on the output resistance of the stage. In reality, some of the signal current in *ro* is lost by flowing in the external emitter resistor R_e instead of through emitter resistance *re′* (some is also lost due to the finite current gain of the transistor). The fraction of current “lost” depends on the ratio of *re′* to R_e, which in turn is a reflection of the amount of the emitter degeneration. As a rough approximation, the output resistance due to Early effect for a degenerated CE stage is:

$$R_{out} \sim ro \times degeneration\ factor \qquad (24\text{-}4)$$

If *ro* is 100kΩ and 10:1 emitter degeneration is used as in Fig. 24-7, then the output resistance of the CE stage due to Early effect will be on the order of 1 Bear in mind that this is just a convenient approximation. In practice, the output resistance of the stage cannot exceed approximately *ro* times the current gain of the transistor. It has been assumed that the CE stage here is driven with a voltage source. If it is driven by a source with significant impedance, the output resistance of the degenerated CE stage will decrease somewhat from the values predicted above. That reduction will occur because of the changes in base current that result from the Early effect.

24.3.1.3 Miller Effect

The collector-base capacitance C_{cb} of Q_1 creates a feedback current to the base of Q_1 that reduces input impedance at high frequencies. Because the input and output signals of the stage are out of phase, the signal voltage across C_cb is the sum, which is V_{out} + V_{in}. Since the gain here is 19, the signal across C_cb is 20 times the input signal, meaning that the current in C_{cb} is 20 times what it would be if the right end of the capacitor was connected to ground. The effective value of C_{cb} is thus 20 times as large. This is called Miller effect or Miller multiplication. It can be a significant limiter of high frequency response in a CE stage.

24.3.2 Common Collector Stage

Fig. 24-8 illustrates the common collector stage, often referred to as the emitter follower. Instead of voltage gain, the emitter follower provides current gain. Its voltage gain is nominally just shy of unity. It is most often used as a buffer stage, permitting the high impedance output of a CE stage to drive a heavier load.

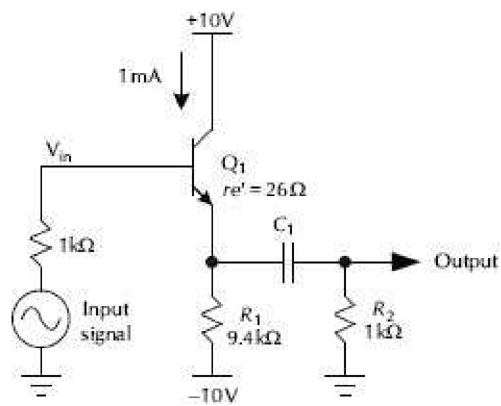


Figure 24-8. Common collector stage.

The signal current in the emitter will be equal to V_{out}/R_L, while the signal current in the base of Q_1 will be this amount divided by β of the transistor. It is immediately apparent that the input impedance seen looking into the base of Q_1 is equal to the impedance of the load multiplied by the current gain of Q_1. This is the most important function of the emitter follower.

The voltage gain of the emitter follower is nearly unity. Suppose R_1 is 9.4kΩ and the transistor bias current is 1mA. The intrinsic emitter resistance re' will then be about 26Ω. Suppose R_2 is 1 kΩ, making net R_L equal to 904Ω. The voltage gain of the emitter follower is then approximately,

$$G = \frac{R_L}{(R_L + re')}$$
$$= 0.97$$

At larger voltage swings the instantaneous collector current of Q_1 will change with signal, causing a change in re'. This will result in a change in incremental gain that corresponds to distortion. Suppose the signal current in the emitter is 0.9 mA peak in each direction. This is a large fraction of the idle current of 1 mA, so we can expect some distortion. The resulting output voltage will be about 814mV peak. At the negative peak swing, emitter current is only 0.1 mA and re' has risen to 260Ω. Incremental gain is down to about 0.78. At the positive peak swing the emitter current is 1.9 mA and re' has fallen to 13.7Ω, resulting in a voltage gain of 0.985.

Voltage gain has thus changed by about 21% over the voltage swing excursion, causing considerable second harmonic distortion. One solution to this is to reduce R_1 so that a greater amount of bias current flows, making re' a smaller part of the gain equation. This of course also reduces net R_L somewhat. A better solution is to replace R_1 with a constant current source of significantly greater than 1 mA.

The transformation of low-value load impedance to much higher input impedance by the emitter follower is a function of the current gain of the transistor. The β is a function of frequency, as dictated by the f_T of the transistor. This means, for example, that a resistive load will be transformed to impedance at the input of the emitter follower that eventually begins to decrease with frequency as β_{ac}

decreases with frequency. A transistor with a nominal β of 100 and f_T of 100 MHz will have an f_β of 1 MHz. The ac β of the transistor will begin to drop at 1 MHz. The decreasing input impedance of the emitter follower thus looks capacitive in nature, and the phase of the input current will lead the phase of the voltage by an amount approaching 90° at high frequencies.

The simplicity of the emitter follower, combined with its great ability to buffer a load, accounts for it being the most common type of circuit used for the output stage of power amplifiers. An emitter follower will often be used to drive a second emitter follower to achieve even larger amounts of current gain and buffering. This arrangement is sometimes called a Darlington connection. Such a pair of transistors, each with a current gain of 50, can increase the impedance seen driving a load by a factor of 2500. Such an output stage driving an 8Ω load would present an input impedance of 20,000Ω.

24.3.3 Cascode Stage

A cascode stage is implemented by Q_2 in Fig. 24-9. The cascode stage is also called a common base stage because the base of its transistor is connected to ac ground. Here the cascode is being driven at its emitter by a CE stage comprising Q_1. The most important function of a cascode stage is to provide isolation. The cas-code keeps the output signal away from the collector of Q_1, mitigating Early effect and Miller effect of Q_1. It provides near-unity current gain, but can provide very substantial voltage gain. In some ways it is like the dual of an emitter follower.

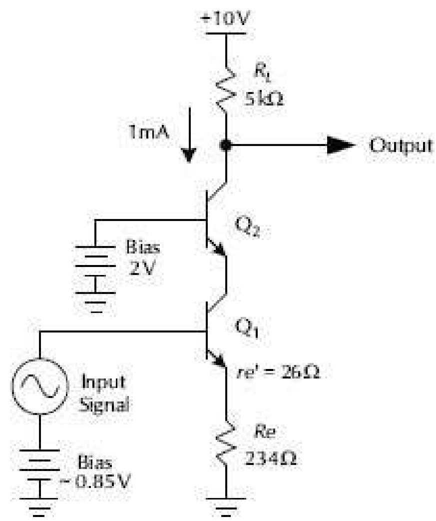


Figure 24-9. Cascode stage.

24.3.4 Three-stage Amplifier

A three-stage amplifier is depicted in Fig. 24-10. It includes 2 common emitter stages in tandem followed by an emitter follower output stage. Feedback through the combination of R_4 and R_3 establishes the closed-loop gain of the amplifier. The higher the attenuation of the output signal by the combination of R_3 and R_4, the higher the gain of the amplifier. The operation of negative feedback will be covered in more detail in the next section. The amplifier is biased by the voltage divider formed by R_1 and R_2. The bias point is stabilized by the negative feedback. Gain of the amplifier is about 10. Bandwidth is about 13 MHz.

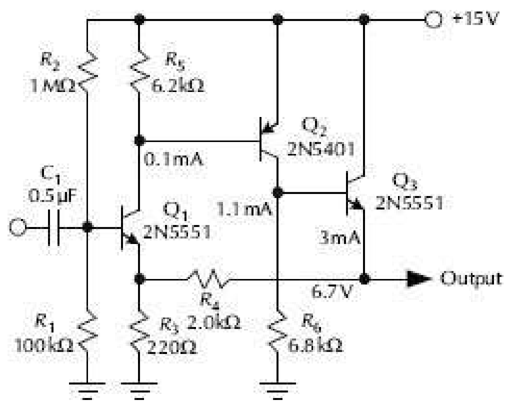


Figure 24-10. Three-stage amplifier.

24.3.5 Differential Pair

The differential amplifier is illustrated in Fig. 24-11. It is much like a pair of common emitter amplifiers tied together at the emitters and biased with a common current. This current is called the *tail current*. The arrangement is often referred to as a *long-tailed pair*, or LTP.

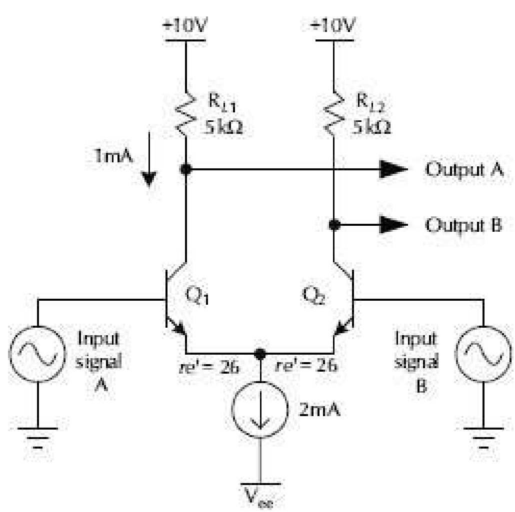


Figure 24-11. Differential pair.

The differential amplifier routes its tail current to the collectors of Q_1 and Q_2 in accordance with the voltage differential across the bases of Q_1 and Q_2. If the base voltages are equal, then equal currents will flow in the collectors of Q_1 and Q_2. If the base of Q_1 is more positive than that of Q_2, more of the tail current will flow in the collector of Q_1 and less will flow in the collector of Q_2. This will result in a larger voltage drop across the collector load resistor R_{L1} and a smaller voltage drop across load resistor R_{L2}. Output A is thus inverted with respect to Input A, while Output B is non-inverted with respect to Input A.

Visualize the intrinsic emitter resistance re' present in each emitter leg of Q_1 and Q_2. Recall that the value of re ‘ is approximately 26Ω divided by the transistor operating current in mA. With 1mA flowing nominally through each of Q_1 and Q_2, each can be seen as having an emitter resistance re' of 26Ω. Note that since $gm = 1/re'$ is dependent on the instantaneous transistor current, the values of gm and re' are somewhat signal dependent, and indeed this represents a nonlinearity that gives rise to distortion.

Having visualized the ideal transistor with emitter resistance re', one can now assume that the idealized internal emitter of each device moves exactly with the base of the transistor, but with a fixed dc voltage offset equal to V_{be}. Now look what happens if the base of Q_1 is 5.2mV more positive than the base of Q_2. The total emitter resistance separating these two voltage points is 52Ω, so a current of 5.2mV/52Ω = 0.1mA will flow from the emitter of Q_1 to the emitter of Q_2. This means that the collector current of Q_1 will be 100 µA more than nominal, and the collector current of Q_2 will be 100 µA less than nominal. The collector currents of Q_1 and Q_2 are thus 1.1mA and 0.9mA, respectively, since they must sum to the tail current of 2.0mA (assuming very high β for the transistors).

If input signals move together in phase, this is referred to as a common-mode input signal. It is easy to see that ideally this produces no change in the collector currents of Q_1 and Q_2, and thus no output. The differential stage thus has *common mode rejection*.

24.3.6 Current Sources

Current sources are used in many different ways in amplifiers, and there are many different ways to make a current source. The distinguishing feature of a current source is that it is an element through which a current flows wherein that current is independent of the voltage across that element. The current source in the tail of the differential pair is a good example of its use. Most current sources are based on the observation that if a known voltage is impressed across a resistor a known current will flow.

In Fig. 24-12 a Zener diode provides a reference voltage for the base of Q_1. The Zener diode is biased with 0.5mA supplied by R_2. About 5.5 V is impressed across the 1.1kΩ resistor R_1, causing 5 mA to flow. The output impedance of this current source is approximately 2MΩ.

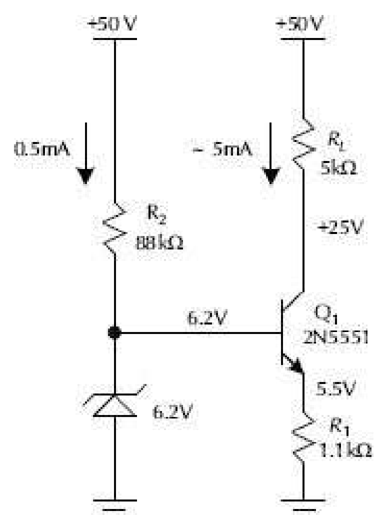


Figure 24-12. Zener reference current source.

Fig. 24-13 illustrates a clever two-transistor feedback circuit that is used to force one V_{be} of voltage drop across R_1, causing a current flow of 5 mA. It does so by using transistor Q_2 to effectively regulate the current of Q_1. If the current of Q_1 is too large, Q_2 will be turned on harder and pull down on the base of Q_1, adjusting its current downward appropriately. As in Fig. 24-12, 0.5mA is supplied to bias the current source. This current flows through Q_2. The output impedance of this current source is an impressive 3MΩ.

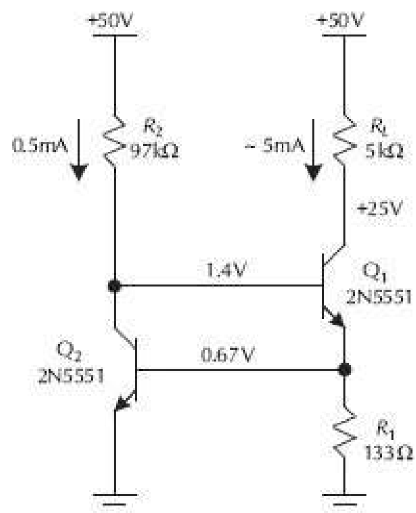


Figure 24-13. Feedback current source.

This circuit will work satisfactorily even if less than 0.5 mA (one tenth the output current) is supplied as bias for Q_2, but then the output impedance will fall to a lower value and the “quality” of the current source will suffer somewhat. This happens because at lower collector current Q_2 has less transconductance and its feedback control of the current variations in Q_1 as a result of Early effect is less strong. If the bias current is reduced to 0.1mA, for example, the output impedance falls to about 1MΩ.

24.3.7 Current Mirrors

Fig. 24-14 depicts a very useful circuit called a current mirror. This particular one is called a Widlar current mirror. If a given amount of current is sourced into Q_1, that same amount of current will be sunk by Q_2, assuming that the emitter degeneration resistors R_1 and R_2 are equal, that the transistor V_{be} drops are the same, and that losses through base currents can be ignored. The values of R_1 and R_2 will often be selected to drop about 100mV to ensure decent matching in the face of unmatched transistor V_{be} drops, but this is not critical.

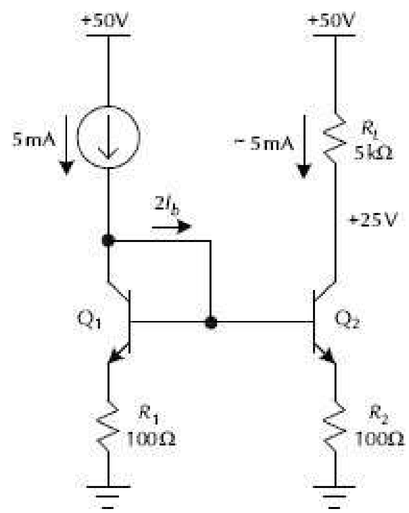


Figure 24-14. Widlar current mirror.

If R_1 and R_2 are made different, a larger or smaller multiple if the input current can be made to flow in the collector of Q_2. In practice, the base currents of Q_1 and Q_2 cause a small error in the output current with respect to the input current. In the example above, if transistor β is 100, the base current I_b of each transistor will be 50 μA, causing a total error of 100 μA, or 2% in the output current. This effect can be reduced by adding an emitter follower "*helper*" transistor between the collector of Q_1 and the base node of Q_1 and Q_2.

24.3.8 Discrete Operational Amplifier

If the output of a differential input stage is fed to a common emitter stage whose output is then buffered by an emitter follower, a differential amplifier with high voltage gain and fairly low output impedance is formed. This is the beginning of a simple operational amplifier. In Fig. 24-15 the differential pair is loaded with a current mirror. The high impedance of the load on the collector of Q_1 results in large voltage gain. Current from both collectors is used in a push-pull fashion and this further increases gain. Notice that multiple current sources are created by slaving Q_8 and Q_{10} off of the feedback current source formed by Q_3 and Q_4.

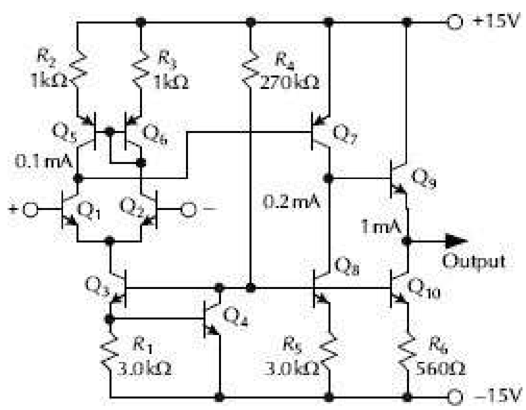


Figure 24-15. Discrete operational amplifier.

Open loop gain is about 113dB, while open-loop bandwidth is about 400Hz. The unity gain frequency is 20MHz with a phase angle of 230°.

24.4 Negative feedback

Negative feedback is employed in virtually every kind of amplifier. It acts to reduce distortion and make gain more predictable and stable. Invented by Harold Black in 1927, it operates by comparing the output of an amplifier stage to the input, creating an error signal that drives the heart of the amplifier.[3]

24.4.1 *Negative Feedback—How It Works*

Fig. 24-16 shows a simplified block diagram of a negative feedback amplifier. The basic amplifier has a forward gain A_{ol}. This is called the *open-loop gain* (OLG) because it is the gain that the overall amplifier would have from input to output if there were no negative feedback.

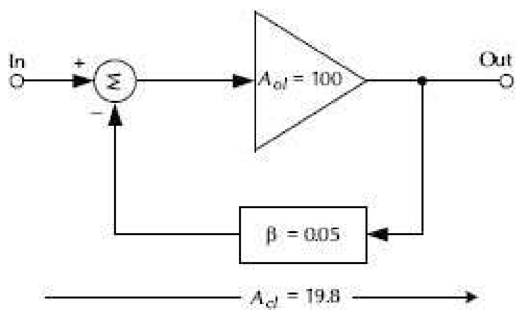


Figure 24-16. Negative feedback.

A portion of the output is fed back to the input with a negative polarity. The fraction governing how much of the output is fed back is referred to as *Beta* (β). The negative feedback *loop gain* is the product of A_{ol} and β. The overall gain of the closed-loop amplifier is called the *closed-loop gain* (CLG) and is designated as A_{cl}. The action of the negative feedback opposes the input signal and makes the closed-loop gain smaller than the open-loop gain, often by a large factor.

The output of the subtractor at the input of the amplifier is called the error signal. It is the difference between the input signal and the divided-down replica of the output signal. The error signal, when multiplied by the open-loop gain of the amplifier becomes the output signal. As the gain A_{ol} becomes large, it can be seen that the error signal will necessarily become small, meaning that the output signal will become close to the value v_{in}/β. If A_{ol} is very large and β is 0.05, it is easy to see that the closed-loop gain A_{cl} will become 20. The important thing to notice here is that the closed-loop gain in this case has been determined by β and not by the open-loop gain A_{ol}. Since β is usually set by passive components like resistors, the closed-loop gain has been stabilized by the use of negative feedback. Because distortion can often be viewed as a signal-dependent variation in amplifier gain, it can be seen that the application of negative feedback also reduces distortion.

24.4.2 Closed-loop Gain

The closed-loop gain for finite values of open-loop gain is shown in the equation below. As an example, if A_{ol} is 100 and β is 0.05, then the product $A_{ol}\ \beta = 5$, which is the loop gain. The closed-loop gain (A_{cl}) will be equal to 100/(1+5) = 100/6 = 16.7. This is somewhat less than the closed loop gain of 20 that would result for an infinite value of A_{ol}. If A_{ol} were 2000, A_{cl} would be 2000/(1+100) = 2000/101 = 19.8. Here we see that if $A_{ol}\ \beta = 100$, the error in gain from the ideal value is about 1%. The product $A_{ol}\beta$ is called the loop gain, since it is the gain around the feedback loop. The net gain around the feedback loop is a negative number when the negative sign at the input subtractor is taken into account.

$$A_{cl} = \frac{A_{o1}}{1 + A_{o1}\beta} \qquad (24\text{-}5)$$

Notice that if the sign of $A_{ol}\beta$ is negative for some reason, the closed-loop gain will become greater than the open-loop gain. For example, if $A_{ol}\beta = -0.6$, the closed-loop gain will become $A_{cl}\beta = A_{ol}/(1 - 0.6) = 2.5\ A_{ol}$. The closed-loop gain will ultimately go to infinity and oscillation will result if the product $A_{ol}\ \beta$ reaches −1. The potential for positive feedback effects is of central importance to feedback compensation and stability. Negative feedback is said to have a phase of 180°, corresponding to a net inversion as the signal circles the loop. Positive feedback has a phase of 0°, which is the same as 360°.

24.4.3 Input-referred Feedback Analysis

A useful way to look at the action of negative feedback is to view the circuit from the input, essentially answering the question, "*what input is required to produce a given output?*" Viewing negative feedback this way essentially breaks the loop[1]. We will also find later that viewing distortion in an *input-referred* way can also be helpful and lend insight.

In the example above with $A_{ol} = 100$ and $\beta = 0.05$, assume that there is 1.0 V at the output of the amplifier. The amount of signal fed back will be 0.05 V. The error signal driving the forward gain path will need to be 0.01V to drive the forward amplifier. An additional 0.05 V must be supplied by the input signal to overcome the voltage being fed back. The input must therefore be 0.06 V. This corresponds to a gain of 16.7 as calculated in Eq. 24-5.

Suppose that instead $\beta = -0.009$. This corresponds to positive feedback, which reinforces the input signal in driving the forward amplifier. With an output of 1.0 V, the feedback would be −0.009 V. The drive required for the forward amplifier to produce 1.0 V is only 0.01V, of which 0.009 V will be supplied by the positive feedback. The required input is then only 0.001V. The closed loop gain has been enhanced to 1000 by the presence of the positive feedback. If β was −0.01, the product $A_{ol}\beta$ would be −1 and the denominator in Eq. 24-5 would go to zero, implying infinite gain and oscillation.

24.4.4 Non-inverting Amplifier

A non-inverting feedback amplifier is illustrated in Fig. 24-17. The open-loop gain (OLG) is 50, while the feedback factor β is 0.1. This means that the closed loop gain (CLG) will be just shy of 10. If open loop gain is infinity, the closed-loop gain will be 10. It is easy to see that the 20 mV required at the input of the amplifier to produce the assumed output of 1.0 V is what causes the gain to be less than the ideal 10. This observation is a form of input-referred feedback analysis.

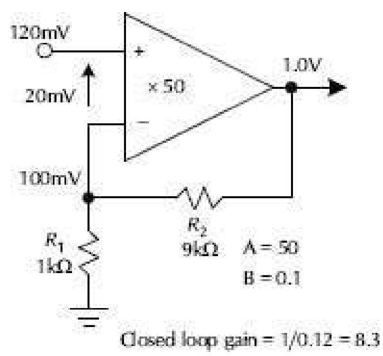


Figure 24-17. Non-inverting amplifier.

24.4.5 Inverting Amplifier

The inverting feedback amplifier shown in Fig. 24-18 is also set for an ideal gain of 10×, but it also has an amplifier with a forward gain of only 50, so once again its actual CLG will be smaller than 10×. In analyzing this circuit, it is useful to recognize that all of the current in R_1 also flows in R_2 if the input impedance of the op-amp is high.

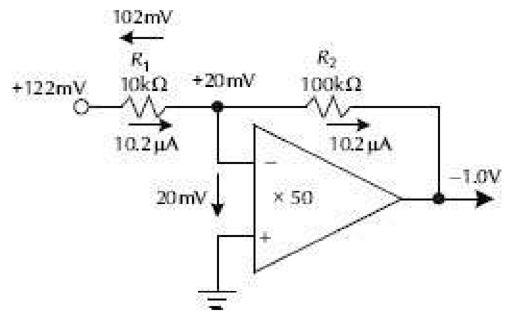


Figure 24-18. Inverting amplifier.

Notice that the feedback factor β is 0.091 as a result of the voltage divider formed by R_2 and R_1. Loop gain is thus 4.55, a fairly small value. It is easy to see by the input-referred analysis shown in the figure that the actual gain is 8.197.

24.4.6 Feedback Compensation and Stability

Negative feedback must remain negative in order to do its job. Indeed, if for some reason the feedback produced by the loop becomes positive, instability or oscillation may result. As we have seen above, simple negative feedback relies on a 180° phase shift (an inversion) located somewhere in the loop. High-frequency *roll-off* within the loop can add additional lagging phase shift that will cause the loop phase to be larger than 180°. These excess phase shifts are usually frequency-dependent and increasing with frequency.

As shown by the plot in Fig. 24-19, the loop gain is often rolled off at 6dB/octave as frequency increases so that the loop gain goes below unity before enough phase shift accumulates around the loop to reach 360°. Note that this deliberate compensation roll-off introduces 90° of phase shift at frequencies well above the pole, so any additional phase shift in the circuit must be well less than 90° before the unity loop gain frequency is reached. The feedback compensation depicted in Fig. 24-19 is called dominant pole compensation because a single pole located at a relatively low frequency (here about 600 Hz) is the major contributor to gain roll-off with frequency.

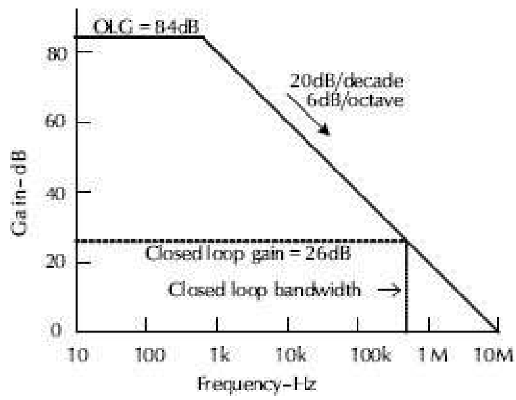


Figure 24-19. Feedback compensation

If the total phase lag is 135° at the frequency where the loop gain falls to unity (here 500 kHz), then the *phase margin* (PM) is 45°. This is the margin against 180° phase shift where oscillation will occur. Similarly, if the loop gain at the frequency where loop phase lag is 180° has fallen to –6dB, the *gain margin* (GM) is 6dB. These are the minimum margins that are generally acceptable in a design.

At the 3 dB frequency f_p of a pole, the additional lagging phase shift is 45°. At very high frequencies a pole will contribute 90° of phase shift. If a system has multiple poles this added phase shift may reach 180° at some high frequency. The total phase shift around the loop will then equal 360°, and there will be positive feedback. If the gain around the loop at this frequency is unity, oscillation will occur.

24.5 Operational Amplifiers

Operational amplifiers are high-gain circuits whose larger function is defined by negative feedback. In essence, they are a universal gain block that can be used in a great many different ways.[4,5] They are virtually all implemented conveniently in integrated circuits. They come in many different varieties to suit different applications, but the operating and application principles are generally the same for all of them.

With a few exceptions, operational amplifiers operate in the small-signal domain, where signal voltages are often less than about 10Vrms. Indeed, many operational amplifiers are operated from ±15V power supplies.

24.5.1 The Ideal Op-amp

The ideal op-amp has differential inputs and infinite gain, with no input dc offset voltage and no input bias current. It may be characterized by finite bandwidth, however. This is the parameter referred to as gain-band-width product. It is the frequency where the voltage gain of the op-amp falls to unity, usually at a rate of 6dB per octave. Most real op-amps have very high gain at dc and low frequencies, often on the order of 120dB or more. A typical modern op-amp designed for audio may have a gain-bandwidth product of about 10MHz. This means that the op-amp open-loop gain at 1kHz will be about 10,000, or 80dB (10MHz divided by 1 kHz).

24.5.2 Inverting Amplifiers and the Virtual Ground

Consider the simple op-amp inverter in Fig. 24-20. Assume its open-loop gain is 80dB at dc (it is actually much more). When the op-amp described above is delivering +1V at its output, the signal voltage at the inverting input must be only −0.1 mV, since the positive input is connected to ground. This number is so small and close to ground that we refer to the inverting input node of the op-amp in such a feedback arrangement as a "*virtual ground*". The currents in R_1 and R_2 must be virtually identical. It is then easy to see that a −1 V input is required to produce the +1 V output.

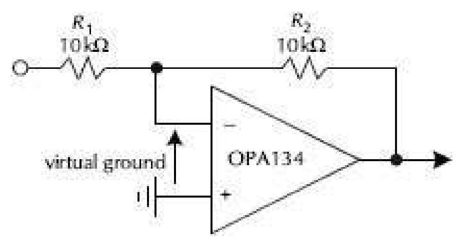


Figure 24-20. A unity-gain inverter

In essence, the negative feedback forces these relationships to hold. In practice, the voltage across R_1 will be only 999.9mV as a result of the 0.1 mV that actually exists at the inverting input of the op-amp. This means that the gain is ever so slightly less than unity, as expected in a negative feedback system with less than infinite loop gain. Notice that, for the feedback loop gain, the combination of R_1 and R_2 forms a 6dB attenuator. The loop gain in this arrangement is thus 74dB, corresponding to a ratio of 5000. The closed-loop bandwidth will be on the order of 5 MHz, since the loop gain has been divided by 2 by the combination of R_1 and R_2.

Fig. 24-21 shows an amplifier with an inverting gain of 10, with R_1 = 10k and R_2 = 100k. Since the currents flowing in R_1 and R_2 are virtually identical, and since the voltage at the inverting input of the op-amp is virtually zero, it can be seen by inspection that the gain is −10. When the output voltage is +1V, the input voltage will be almost exactly −100mV. Note here that the combination of R_1 and R_2 introduces 21 dB of attenuation in the feedback loop: $R_1/(R_1 + R_2) = 0.09$. This means that the loop gain at 1kHz is now only 59dB.

The voltage error is still about 0.1 mV at the inverting input of the op-amp, but it is now larger in comparison to the smaller input signal. The combination of R_1 and R_2 attenuates the feedback signal by a factor of 11, so the closed loop bandwidth will be on the order of 10MHz/11 = 910kHz.

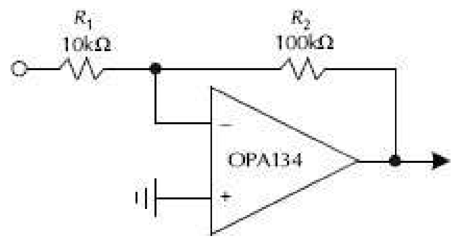


Figure 24-21. A × 10 inverting amplifier.

Fig. 24-22 illustrates a simple mixer circuit where three different signal sources can be mixed together in equal amplitudes. In this circuit, each input sees a gain of −1 to the output. If different resistor values are used, different gain can be applied to each signal source. The virtual ground suppresses crosstalk among the sources.

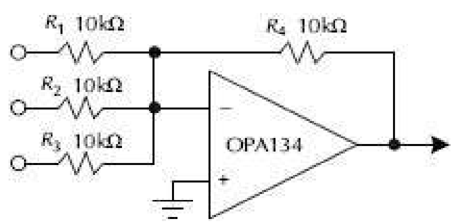


Figure 24-22. A three-input mixer.

Here the loop gain has been divided by 4 by the action of R_4 against the parallel combination of R_1, R_2 and R_3. The closed-loop bandwidth will thus be only about 2.5MHz. If you visualize all three inputs driven by a single source, you can see that the amplifier looks like an ×3 inverter.

Circuits like this are important in pro sound mixers where often many inputs are summed together. The reduced loop gain in such situations must be kept in mind, since bandwidth and distortion will be adversely affected. Amplifier input noise will also be increased. To amplifier noise, the above arrangement looks like a non-inverting amplifier with a gain of 4, and the *noise gain* is said to be 4. This makes it all the more important to use low-noise amplifiers in mixing stages with a large number of inputs.

24.5.3 Non-inverting Amplifiers and the Virtual Short

If both the inverting and non-inverting inputs of the op-amp are driven in a negative feedback arrangement, the very small voltage that will exist across those two inputs allows us to think of there being a "virtual short" between those two inputs, meaning that in a feedback arrangement those two voltages will always be virtually the same. This approximation makes the design of such circuits very easy.

Consider the unity-gain non-inverting buffer in Fig. 24-23. If there is 1V at its output at 1kHz, then there will only be 0.1 mV across the inputs, meaning that the input signal and output signal will differ by only 0.1 mV. In fact, the input will be 1.0 V plus 0.1 mV, or 1.1 V, reflecting a very small departure in gain from +1. Note that in many of these analyses, we are working backwards from the assumed output to the input. This often makes the analysis of feedback circuits easy. This circuit is often called a voltage follower.

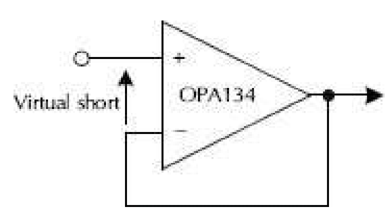


Figure 24-23. Non-inverting unity gain buffer.

Now consider the +10× buffer in Fig. 24-24. R_1 and R_2 form a 10× voltage divider in the feedback path. When the output is 1 V, the voltage at the inverting input will be 100mV and the voltage at the main input will be 100.1 mV. By inspection, the gain is thus ever so slightly less than 10×. The 10× attenuation in the feedback path means that the loop gain will fall to unity at 10MHz/10 = 1MHz, and this will be the approximate closed loop bandwidth.

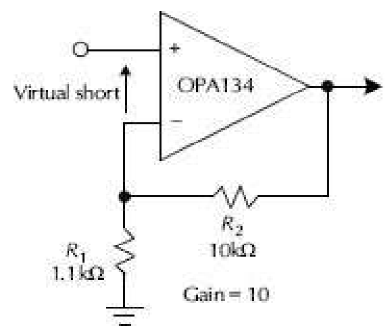


Figure 24-24. Non-inverting ×10 amplifier

24.5.4 Differential Amplifier

The circuit in Fig. 24-25 is a popular op-amp-based differential amplifier arrangement. It can be analyzed in a number of different ways, but one way is to see it as a combination of an inverting amplifier and a non-inverting amplifier. The concept of the virtual short once again makes analysis easy. With 10kΩ input resistors (R_1 and R_3) and a 10kΩ feedback resistor (R_2) and a 10kΩ shunt resistor (R_4) to ground, the amplifier is configured for a differential gain of 1×.

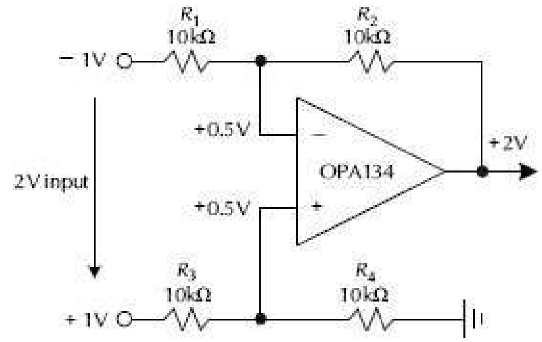


Figure 24-25. Single op-amp differential amplifier.

Once again, assume a 1 V output and work your way backwards to the input. To make things easy, now ignore the 0.1mV that must appear across the op-amp input at 1kHz. If the positive input to the circuit is grounded, it is easy to see that the circuit defaults to an inverter with a gain of −1×. If the negative input to the circuit is grounded, it is easy to see that the circuit defaults to a non-inverting amplifier with a gain of +1× (a 2× non-inverting amplifier preceded by a 0.5× attenuator consisting of R_3 and R_4). It is easy to see that if the same signal is applied to both the positive and negative inputs of the circuit, the net gain will be zero. This means that the circuit provides common mode rejection. Looking at it another way, if the output is zero, both input signals create the same voltage at the inverting and non-inverting inputs of the op-amp, so there is no net op-amp input signal to create an output signal.

Notice that the plus and minus input currents to the differential amplifier circuit are not the same if only one or the other input is driven. This means that the input impedances to ground of the plus and minus inputs are not the same. This can be a source of imperfection in the arrangement when it is driven by a source with finite source impedances to ground. This may compromise the common mode rejection of the circuit.

24.5.5 Instrumentation Amplifier

The circuit of Fig. 24-26 is a version of a differential amplifier called an *instrumentation amplifier*. The first two op-amps are configured as a non-inverting differential buffer with a gain of 3×. Each one can be thought of as a simple non-inverting amplifier with the virtual center-tap of R_3 grounded. With a pure differential input this point will be at ground potential anyway. Letting this point float decreases the common-mode gain to unity, thus giving a 3× improvement of differential gain as compared to common mode gain (e.g., CMRR). The buffering effect of U1 and U2 also means that the input impedance of the circuit on the plus and minus inputs is the same. This preserves the common mode rejection in the face of finite source impedances. The second half of the instrumentation amplifier is merely a conventional differential amplifier arrangement like that in Fig. 24-25.

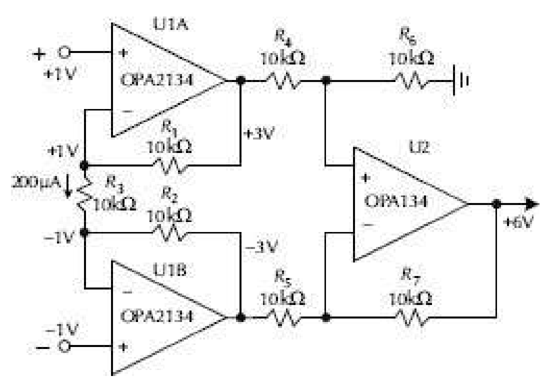


Figure 24-26. Instrumentation amplifier.

24.5.6 Power Amplifiers

As we will see later, most power amplifiers are like discrete op-amps arranged as non-inverting amplifiers with a gain of about 20–30. A typical power amplifier will be set to have a closed-loop bandwidth of about 1 MHz and a closed loop gain of 20.[1] This means that its op-amp-like core will have a gain-bandwidth of about 20MHz. Such an amplifier might have R_1 = ΛΩ and R_2= 19kΩ

24.5.7 Compensation

As with any arrangement using negative feedback, the loop gain must fall to unity at some high frequency for reasons of stability. Most op-amps are internally compensated so that their gain falls to unity at their rated gain-bandwidth frequency. Their forward gain will ideally begin to fall off at their open-loop bandwidth frequency. It will usually fall at 6 dB per octave until the gain-bandwidth frequency is reached. For example, a modern op-amp with gain-bandwidth of 10 MHz and dc gain of 120dB (1 million) will have its gain begin to fall off at 10Hz.

Most modern op-amps are designed to be unity-gain stable. This means that if the output is connected directly to the inverting input, as in a non-inverting unity-gain buffer, the arrangement will be stable, with adequate gain and phase margin. This usually means that an op-amp with a 10 MHz gain-bandwidth will have unity gain at 10MHz with less than perhaps 120° of phase shift, leaving a 60° phase margin.

24.6 Noise

All passive and active circuits create noise, and noise is undesirable. Noise limits the dynamic range of a system. In order to design circuits for the lowest noise, the noise generating phenomena must be understood. There are numerous tradeoffs, and the best low-noise design is usually very dependent on the particular application.[6] More complete discussions than those here on noise as it pertains to audio amplifiers can be found in references 1,6,7,8.

24.6.1 Thermal Noise

The most common type of noise in electronic circuits is *thermal noise*, often called *Johnson noise*. Every resistor, regardless of its value, creates the same amount of noise power. It is the most basic source of noise in electronic circuits. It is most often modeled as a noise voltage source in series with the resistor. The noise power in a resistor is dependent on temperature.[6]

$$\begin{aligned} P_n &= 4kTB \text{ watts} \\ &= 1.66^{-20} \text{ watts/Hz} \end{aligned} \tag{24-6}$$

where,
k is Boltzman's constant = 1.38^{-23} J/K,
T is the temperature in K = 300K @ 27°C),
B is the bandwidth in Hz.

The open-circuit rms noise voltage across a resistor of value R is simply

$$\begin{aligned} e_n &= \sqrt{4KTRB} \\ &= 0.129 \text{ nV}/\sqrt{\text{Hz}} \text{ per } \sqrt{\Omega} \end{aligned} \tag{24-7}$$

Noise voltage for a resistor thus increases as the square root of both bandwidth and resistance.

The noise power is distributed evenly over the full frequency spectrum, and is referred to as white noise. Each hertz of frequency spectrum contains the same amount of noise power. Since we most often refer to noise voltage or noise current, the noise spectral density goes as the square root of the bandwidth in which the noise is being measured. Noise voltage is usually expressed in $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and noise current is usually expressed in $\mathrm{pA}/\sqrt{\mathrm{Hz}}$. A convenient rule of thumb is that a 1kΩ resistor produces about $4\mathrm{nV}/\sqrt{\mathrm{Hz}}$ into an open circuit.[1] A 10Ω resistor will produce $0.4\mathrm{nV}/\sqrt{\mathrm{Hz}}$, while a 100kΩ resistor will produce $40\mathrm{nV}/\sqrt{\mathrm{Hz}}$.

The rms voltage noise in a given noise bandwidth increases as the square root of the bandwidth. For example, in a 20 kHz bandwidth there are $141\sqrt{\mathrm{Hz}}$, so the 1kΩ resistor will produce a noise voltage of 4 × 141 = 564nVrms.

24.6.2 *Shot Noise*

The second most common noise component is shot noise.[6] It results from the quantization of electric charge in the flow of current. Shot noise is also evenly distributed across the frequency spectrum, so it is also white. The shot noise current is usually stated in $\mathrm{pA}/\sqrt{\mathrm{Hz}}$ and has the rms value of

$$\begin{aligned} I_{shot} &= \sqrt{2qI_{dc}B} \\ &= 0.57\,\mathrm{pA}/\sqrt{\mathrm{Hz}}/\sqrt{\mu\mathrm{A}} \end{aligned} \qquad (24\text{-}8)$$

where,
q is 1.6^{-19} Coulombs per electron,
I_{dc} is in A,
B is bandwidth in Hz.

It is easily seen that shot noise current increases as the square root of bandwidth and as the square root of current. A good example of shot noise is the current flow in the collector of a transistor. A collector current of 1mA will have a shot noise component of$\mathrm{pA}/\sqrt{\mathrm{Hz}}$.

If the transistor has a beta of 100, base current will be 10 μA. This corresponds to $3.16\sqrt{\mu\mathrm{A}}$. Input noise current will thus be $1.8\mathrm{pA}/\sqrt{\mathrm{Hz}}$. Put another way, base shot noise is collector shot noise divided by $\sqrt{\beta}$. If the collector shot noise is divided by the transconductance of the transistor, we will have an effective input noise voltage. The transistor biased at 1 mA will have *gm* = 39mS, so the input-referred noise voltage due to collector shot noise will be $0.46\mathrm{nV}/\sqrt{\mathrm{Hz}}$.

Notice that *re′* for this transistor is 26Ω. The noise voltage for a 26Ω resistor is $0.66\mathrm{nV}/\sqrt{\mathrm{Hz}}$. We see that the input-referred voltage noise of a transistor is equal to the Johnson noise of a resistor of half the value of *re′*.[1] This is a very handy relationship. Of course, transistors have base resistance, which adds to the effective input noise. If the transistor operated at 1 mA has base resistance R_B of 13Ω, then R_B will contribute $0.46\mathrm{nV}/\sqrt{\mathrm{Hz}}$, increasing the transistor's effective input noise voltage by 3 dB to $0.66\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Low-noise transistors for use in low-impedance circuits feature small base resistance. Good examples are the 2N4401 and 2N4403. Low-noise transistors for high-impedance circuits, however, have high β in order to keep input noise current small. The 2N5089 is an example of such a transistor.

Noise in JFETs

The voltage noise in JFETs is usually greater than that in a BJT, especially when they are operated at the same current. The primary noise source in JFETs is *thermal channel noise*.[7,8] That noise is modeled as the Johnson noise of an equivalent input resistor r_n

whose resistance is equal to approximately 0.67/*gm*.[7,8] Voltage noise in JFETs increases at very low frequencies. This effect is referred to as *1/f* noise (or *flicker* noise) because its power spectral density increases as the inverse of frequency below the so-called 1/*f* *corner frequency*. The 1/*f* noise voltage thus increases by 3dB per octave as frequency decreases below the 1/*f* noise frequency. The 1/*f* region has the same spectral shape as pink noise. Many JFETs have 1/*f* corner frequencies between 30Hz and 300Hz. The 1/*f* noise phenomenon results from small defects in the semiconductor crystal structure. Input current noise in JFETs is very low, since it is just the shot noise associated with the very small gate leakage current.

24.6.3 Signal-to-Noise Ratio

Consider a power amplifier with a voltage gain of 30 and input-referred noise of $10\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Assume that the signal-to-noise ratio is being measured in a 20kHz bandwidth, which contains $141\sqrt{\mathrm{Hz}}$. Output noise will be 4.23 µV Amplifier *SNR* is usually specified with respect to 1W into 8Ω, or 2.83 V. The *SNR* is then $2.83/4.23\times10^{-6}$, or about 116dB—a very good number.

24.6.4 A-weighted Noise Specifications

The frequency response of the A-weighting curve is shown in Fig. 24-27. It weights the noise in accordance with the human ear's perception of noise loudness.

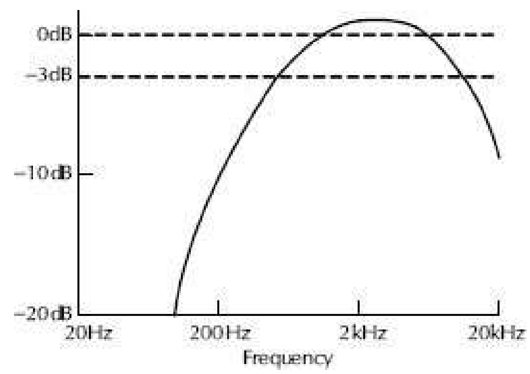


Figure 24-27. A-weighting frequency response.

The A-weighted noise specification for an amplifier will usually be quite a bit better than the un-weighted noise because the A-weighted measurement tends to attenuate noise contributions at higher frequencies and hum contributions at lower frequencies. A good power amplifier might have an un-weighted signal to noise ratio (*SNR*) of 90dB with respect to 1W into 8Ω, while that same amplifier might have an A-weighted *SNR* of 105dB with respect to 1W.[1] A fair amplifier might sport 65dB and 80dB *SNR* figures, respectively. The A-weighted number can often be 10–20dB better than the un-weighted number.

24.7 Preamplifiers and Mixers

Various kinds of small-signal amplifier circuits play a critical role in consumer preamplifiers and professional sound mixers. In this section we'll take a look at several of those amplifier functions and their applications.

Preamplifiers boost small signals to line-level signals and/or manipulate line level signals. They are also used to perform equalization, such as in phono or tape preamplifiers where RIAA or NAB equalization must be implemented. There are thus two broad classes of preamplifiers, often implemented together in a single piece of equipment: low-level and line-level. A mixing desk can be thought of as an elaborate preamplifier that is arranged with a plurality of preamplifier-like channel modules whose outputs can be mixed together in different proportions into the left and right output channels. The level control for each channel is called a *fader* and the control that distributes the signal between the left and right channels to position the source in the sound field is called a *pan pot*.

Line-level preamps typically provide only a modest line-level gain, perhaps 12dB to boost a 0.5 V line level signal to a 2 V signal to drive a power amplifier. The main function of a line-level preamp is the control of volume and the selection of input sources, which may include CD/SACD players, audio DACs, phono stages and microphone preamplifiers.

The volume control is most often implemented with a dual audio-taper potentiometer. Other approaches to volume controls include ICs and switched attenuators. The location of the volume control in the signal chain can affect noise and overload performance. If the volume control is early in the chain, overload performance will be improved by the nominal attenuation of the volume control, but the gain after the volume control will contribute noise that is not attenuated when the volume is set low. If the volume control is toward the output of the preamp, the converse will be the case.

24.7.1 Typical Preamplifier Signal Path

Fig. 24-28 shows a block diagram of a preamplifier signal chain, including a low-level phono preamp, input selection, volume control, balance control and tone control.

24.7.2 Simple Line Level Preamplifier

A simple line-level preamp is shown in Fig. 24-29.[9] It includes only a balance control and a volume control. Not shown are functions like input selection, tone controls, equalizers, etc.

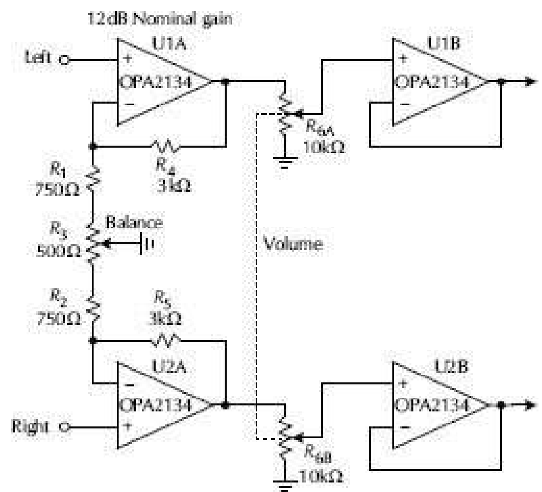


Figure 24-29. Simple line level preamp.

The preamp includes 2 stages in each channel, both implemented with the well-regarded OPA2134 dual JFET op amp. The first stage provides a nominal gain of 12 dB, which is varied about 3.5dB peak to peak by the position of the wiper on the balance control. The simple balance control above is susceptible to some small level of crosstalk due to finite wiper resistance, typically about 5 Ω for a quality pot of this value. Crosstalk will be about −46dB. Other preamp designs use separate balance pots for each channel, some specifically designed for the balance function and including center-position detents. These do not suffer the wiper crosstalk problem.[9,10]

The volume control is shown implemented by the usual audio-taper pot, many of which suffer from channel-to-channel mismatch that can be as high as 1–2dB. Many alternatives exist, including stepped attenuators, integrated circuit stepped attenuators and voltage-controlled amplifiers (VCAs). Circuits that employ linear pots with inherently less tracking error in circuits that approximate a log taper have also been described.[10,11]

24.7.3 Phono Preamplifiers

Fig. 24-30 illustrates a classic phono preamp design.[12] The circuit is a non-inverting feedback circuit whose gain is shaped by the feedback network to conform to the RIAA equalization characteristic. A typical phono preamp will provide about 35dB of gain at 1kHz. Relative to 1 kHz, the RIAA curve is up about 20 dB at 20 Hz and down about 20dB at 20kHz. Thus, an equalization gain range of 40dB is required. The RIAA curve falls with frequency, with low-frequency time constants at 3180 µs and 318 µs. The high-frequency part of the equalization is characterized by a pole with a 75 µs time constant. The low-distortion, low-noise LM4562 is used as the operational amplifier. C_4 establishes the 3180 µs time constant while C_3 establishes the 75 µs time constant. Further discussion and a more sophisticated phono preamp that handles both moving coil and moving magnet cartridges is described in reference 13.

24.7.4 Dynamic Microphone Preamplifiers

The requirements for a dynamic microphone preamp are not unlike those of a phono preamp, since the voltage level and source impedance are similar.[8,9,10] A typical dynamic microphone may produce 2mV at 94 dB SPL (1 Pascal), and may have a source impedance of about 300Ω. Low input voltage and current noise is important. The microphone preamp does not require equalization, but it must accept a balanced input and have a very large controllable gain range.

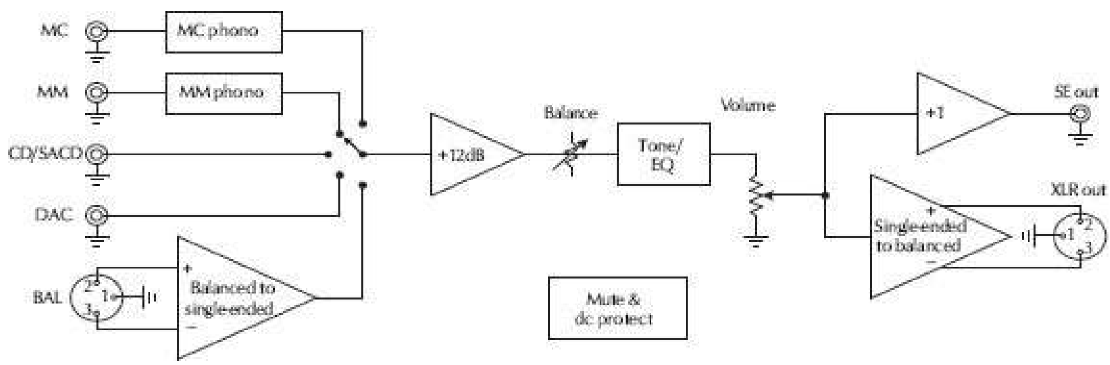


Figure 24-28. Typical preamplifier signal path.

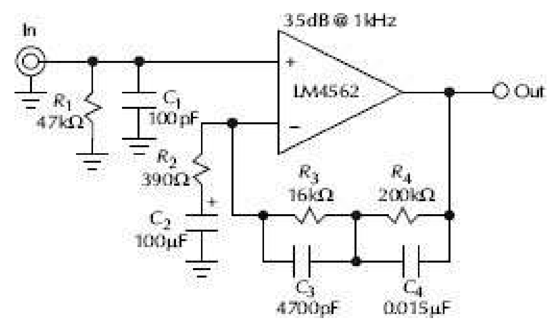


Figure 24-30. Phono preamp.

The front-end can be implemented as a differential amplifier without negative feedback. This provides a true high-impedance balanced input with soft overload characteristics. The input stage consists of a degenerated differential a pair of JFETs, each with a current source connected to the source. The gain is set by a variable resistance connected from source to source. Drain load resistors create a differential output that is fed to an op-amp configured as a differential amplifier. In such a design, the relatively low transconductance of the JFETs limits the amount of gain that can be obtained.

An improved design is shown in Fig. 24-31.[8,9] Each JFET is configured as a complementary feedback pair (CFP) by adding a PNP transistor in the drain circuit. This arrangement increases the effective transconductance of the JFET by a factor of about 50 and provides local distortion-reducing feedback. The JFETs are biased at 1mA and the BJTs are biased at 2mA. Care should be taken to minimize the amount of noise from current sources I_1 and I_2 that is not in the common mode. Dc offset is controlled by a servo that trims I_1 and I_2 (not shown).

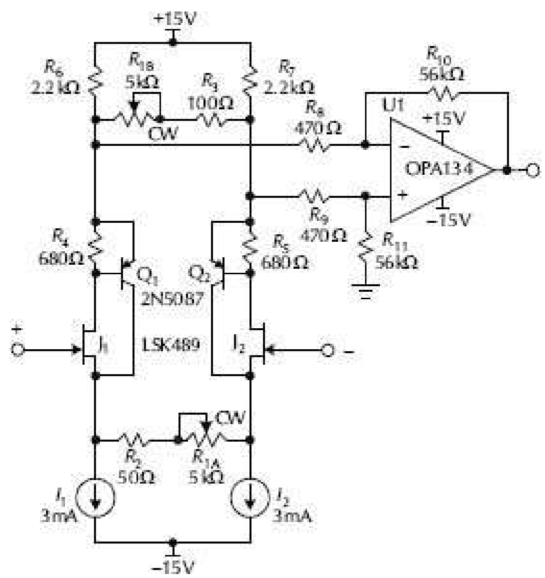


Figure 24-31. Dynamic microphone preamp.

Gain control over a wide range is difficult to achieve with a single variable resistance in the JFET source circuit, so a second pot ganged with the first is added in a differential shunt arrangement in the drain circuit. A gain adjustment range of 6–60dB is achieved with a single knob, and is reasonably distributed with respect to pot rotation even when using linear pots.[9] The use of log-taper pots can provide an even more uniform distribution of attenuation vs. rotation.

Input noise is only $5\text{nV}/\sqrt{\text{Hz}}$ at the highest gain setting. Harmonic distortion is no more than 0.012% at an output level of 5 V peak. At a nominal line level of 1 Vrms, THD is 0.003% at a gain setting of 60dB. Even though the circuit appears differential and symmetrical, distortion is dominated by the much more benign second harmonic. This is due to the asymmetry created in the drain circuit by the differential op-amp configuration. It results in a difference in signal amplitude at the drains of J_1 and J_2. This effect can be eliminated by use of an instrumentation-type differential amplifier. Distortion above the 3rd harmonic is virtually absent.

24.7.5 Condenser Microphone Preamplifiers

A condenser microphone consists of a microphone capsule and a built-in preamplifier. The output of the condenser microphone is fed by a microphone cable to the input of a conventional dynamic microphone preamp located in the mixing console. The condenser microphone capsule comprises a diaphragm that is the plate of a capacitor that is charged to 40–60V. It produces a voltage when the acoustic vibrations of the diaphragm change the capacitance while charge is conserved. The capacitance may be as small as 5pF but is often in the neighborhood of 50pF. The output impedance of the capsule is thus extremely high at audio frequencies (160MΩ to 1.6GΩ at 20 Hz). The microphone preamp functions mainly as a buffer, since the output voltage of the capsule is fairly high in comparison to the output voltage of a dynamic microphone.[8,9,10] A typical number is 20mV at 94dB SPL (1 Pascal).

The extremely high capacitive source impedance of the capsule means that the amplifier must present a very high load resistance in order to preserve low frequency response. This loading may need to be on the order of 1-100Ω. A typical input circuit comprises a 10Ω polarizing resistor at the capsule ac-coupled to a 10Ω gate return resistor, resulting in a net 5Ω input resistance. A JFET buffers the signal. The input resistance of the preamp must also be high in order to minimize the noise contribution of that resistance. The capsule capacitance low-pass filters the Johnson noise of the 10GΩ polarizing and gate-return resistors. A capsule with higher capacitance will thus provide a better *SNR* for a given preamp input resistance. The Johnson noise of the preamp input resistance (passive self-noise) often dominates the noise of the preamp. For a 20pF capsule loaded with 5GΩ, this noise will be $17\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 1 kHz and $700\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 20Hz.

A simple condenser microphone preamp is shown in Fig. 24-32. The preamplifier input capacitance must be very low, especially if it is nonlinear like that of a semiconductor junction. The drain of the JFET is often bootstrapped with signal to further reduce the effect of C_{rss}. While most condenser microphone preamps use a single-ended JFET input stage, differential arrangements using dual monolithic JFETs can provide lower distortion in the presence of the fairly high input voltages that can be present with a condenser microphone capsule under high *SPL* conditions.[8]

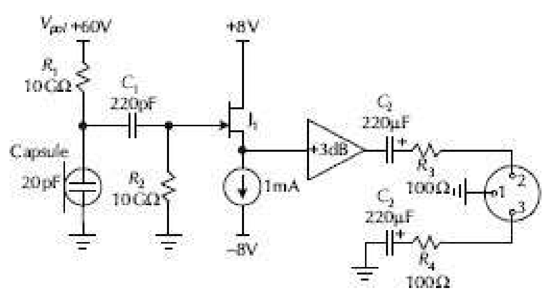


Figure 24-32. Simple condenser microphone preamplifier.

The gate leakage current of a JFET can come into play in a condenser microphone preamplifier. The input is ac-coupled and the gate is biased with a resistor as large as 10GΩ. The JFET's gate input current flows into the resistor, creating a positive voltage offset. A maximum gate current might be 25pA at 25°C. This gate current can create an offset of +250mV through 10GΩ. The gate leakage current doubles every 10°C. Consider a condenser microphone in the hot sun reaching a temperature of 45°C. Maximum gate current could reach 100pA, with a resulting input offset voltage of +1V. Such an offset can be disruptive to some circuits. Offset voltage can be controlled by a dc servo connected to the return end of the 10GΩ gate return resistor.[8,9]

The voltage gain of the preamp may often be on the order of 10dB or less. In fact, for more sensitive capsules and high sound levels, attenuation of the signal may be necessary. A useful approach to attenuation is to switch in a shunt capacitor across the capsule, creating an attenuation of perhaps 20dB. A 20pF capsule with a 180pF shunt capacitor will have 20dB of attenuation. An alternative means to attenuation is to reduce the polarizing voltage; this approach does not preserve the *SNR*, but that may not be an issue where high *SPL* conditions require attenuation.

24.7.5.1 A Practical Preamp Front-end

The preamplifier input capacitance must be kept very low. Fig. 24-33 illustrates a design where the drain of the JFET is bootstrapped with the output signal to further reduce the effect of C_{rss}. J_1 still acts essentially as a source follower, but it is cascoded by J_2, whose gate is driven by the output signal raised by $2V_{be}$ with diodes D_1 and D_2. The low capacitances of the LSK489 dual monolithic JFET further helps keep effective input capacitance very low. Finally, PNP transistor Q_1 makes the circuit into a complementary feedback pair

(CFP), raising the effective transconductance of J_1 by a factor of about 10 and providing the forward bias for D_1 and D_2 in the process. The local negative feedback provided by the CFP greatly reduces distortion. It also keeps the gain of the stage extremely close to unity, maximizing the effectiveness of the bootstrapping, even when the circuit is driving a load. Input voltage noise of the circuit is about $3.7\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$.

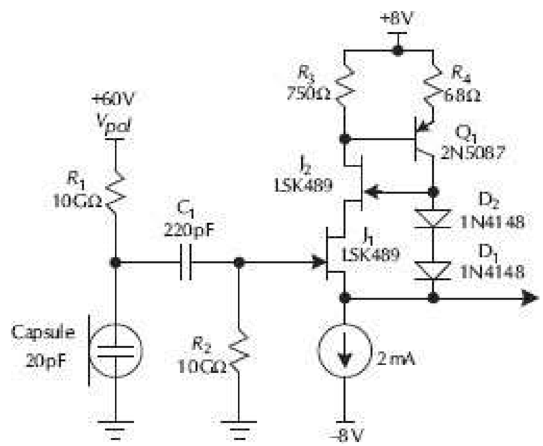


Figure 24-33. Improved condenser microphone preamplifier.

24.7.5.2 Phantom Powering

The condenser microphone electronics must be powered by so-called "*phantom*" powering. Phantom power is usually supplied to the balanced microphone lines through a pair of 6.8PΩ resistors from 48 V. A preamp requiring 7 mA, for example, will drop 24 V on the phantom source, leaving only 24 V for the preamp electronics. For this reason, low-power electronics are a must. A dc-dc converter comprising an R-C oscillator or L-C oscillator and a capacitance voltage multiplier can be used to supply the 40–60V polarizing voltage for the capsule.[9] Capsule sensitivity is proportional to the polarizing voltage.

24.7.5.3 Driving the Balanced Line

Many condenser microphone preamps drive the balanced microphone cable with a single-ended signal, ac-coupling the *cold* wire to shield ground, sometimes through a terminating resistor, as shown in Fig. 24-32. This is adequate but not optimal. A better approach is to drive the *hot* and *cold* (pins 2 and 3) lines with the true balanced output from a differential pair of BJT or JFET transistors, essentially using the pair of 6.8kΩ phantom drive resistors at the far end as the collector or drain load. A local differential shunt can be used to establish the microphone preamp's differential output impedance.[9] Some designs drive the balanced line with emitter followers.

24.7.5.4 The Printed Wiring Board

The extremely high input impedance at the input dictates that the input nodes be carefully guarded and/or shielded in the layout and that high quality printed wiring board material be used to minimize dielectric absorption losses, *hook* effects and parasitic resistance.[9] I recommend using Teflon or Panasonic Megtron 6 PWB material. Meticulous cleaning of the PWB after assembly is also a must to avoid contamination that can introduce leakage resistance.

24.7.6 Balanced Input Circuits

Many consumer preamps and most pro-audio preamps and mixing boards include balanced inputs with XLR connectors. Because most of the internal functions are single-ended, these inputs must pass through a differential-to-single-ended conversion. The single op-amp differential amplifier of Fig. 24-25 or the instrumentation amplifier of Fig. 24-26 can be used for this purpose. However, the

InGenious™ 1200 chip from THAT Corporation implements this function without external resistors and provides higher performance.[14]

A balanced input circuit using the THAT 1200 IC is shown in Fig. 24-34. The circuit inside the 1200 is akin to an instrumentation amplifier where the input amplifiers are operating at unity gain. However, a common-mode signal is developed from the outputs of U1A and U1B of Fig. 24-26 that is fed back through a fourth op-amp and C_1 to bootstrap the input bias resistors with the common mode signal. This largely takes them out of the picture in respect to their creating a path to ground that can reduce common mode rejection due to source impedance imbalances. This circuit was developed and patented by Bill Whitlock of Jensen Transformers.[14]

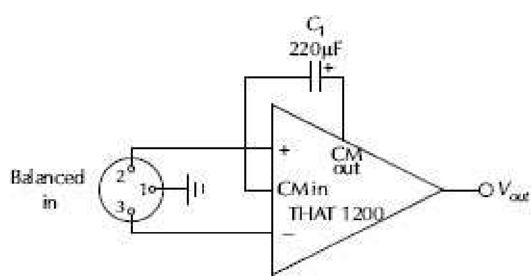


Figure 24-34. Balanced input circuit.

24.7.7 Balanced Output Circuits

Balanced outputs can be implemented with a transformer or by merely adding an inverter to a single-ended output. The transformer provides a floating balanced signal with excellent common-mode immunity, but is expensive and can introduce distortion and bandwidth limitations. The approach using an inverter is inexpensive, but merely provides two opposite-polarity outputs that are actually just single-ended signals referenced to ground. The 1646 *OutSmarts*™ chip from THAT Corporation provides the best of both worlds in a single chip that requires no external components.[15] A balanced output circuit using this chip is shown in Fig. 24-35.

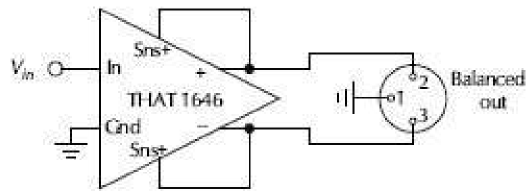


Figure 24-35. Balanced output circuit.

24.7.8 Voltage-controlled Amplifiers

Voltage-controlled amplifiers (VCAs) are widely used in the analog portions of mixing consoles and other pro-audio processing functions. They are especially useful in supporting automation. The *Blackmer*® VCA produced by dbx Inc. has long been the most widely used VCA. It has been made available in IC form by THAT Corporation since the company's inception.[16]

A simple VCA using the THAT 2180 chip is shown in Fig. 24-36. The 2180 is a current-in, current-out device. $R1$ converts the input voltage to a current at the low-impedance input of the 2180 chip. The output signal current is fed to the virtual ground of U2 and the signal is converted to a voltage by means of feedback resistor R_5. The 2180 provides decibel gain/attenuation that is linear with control voltage. In the circuit shown, the control coefficient is 100mV/dB. A fader can be implemented with a pot that merely produces a low-impedance dc voltage to control the gain of the VCA. The VCA is also available in a dual version, the 2162. The circuit shown is simplified, and the control port circuit shown can require significant current. See reference 16 and related application notes for better approaches to control port buffering. The 2180 and 2162 are pre-trimmed for low distortion, but they allow the use of a trimming circuit that injects a small current into the *SYM* pin that can be adjusted for lowest distortion. The trim circuit is not shown here, but can be found in reference 16.

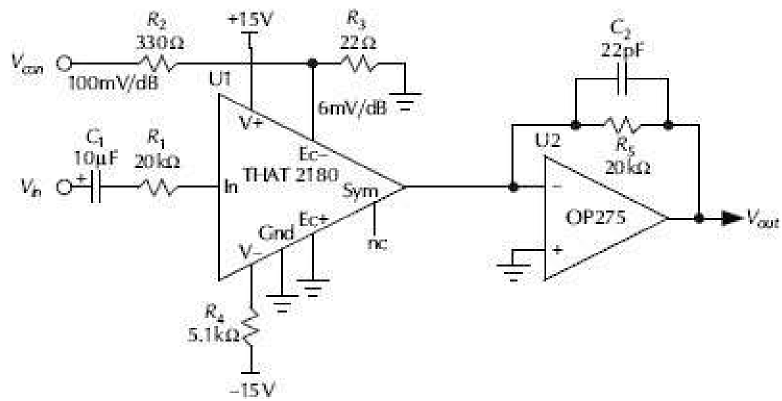


Figure 24-36. Voltage-controlled amplifier.

24.7.9 VCA-based Mixers

The current output nature of the 2180 makes it handy in mixers, since multiple 2180 VCAs can be arranged to drive the virtual ground of U2 in Fig. 24-36. An additional advantage is that the noise gain of the U2 mixing circuit is not increased as channels are added to the mixer, as was the case described in the simple mixer of Fig. 24-22. This is because the input mixing resistors are effectively replaced by signal current sources. Each output of the 2180 has about 15pF of capacitance to ground, so C_2 may have to be increased a bit for stability as additional VCAs are added to the bus.

24.7.10 Panning Circuits

The pan pot in a mixer distributes a channel signal to the left and right summing busses in a proportion that determines the location of the source on the sound stage. This is usually done in such a way that the power sum for that signal in the left and right channels is constant with changes in position. A source panned to the center would then be down 3dB in each channel. In reality, the loudness in the left and right channels does not always sum on a power basis, and can under some conditions be perceived to sum on more like a sound pressure level (*SPL*) basis, which would suggest constant-voltage panning where a centered source would be down 6dB in each channel. A frequent compromise is to have the centered source be down 4.5dB.[10]

Designing pan pot circuits with a pair of potentiometers that implement the desired signal distribution law can be a challenge involving numerous compromises. Using a pair of VCAs, with their convenient linear decibel control law, can make the design easier and better-performing.[17] The THAT 2162 dual VCA is a good choice for this application. The two VCAs are driven with opposing nonlinear control voltages arranged so that the dc voltage provided by the pan pot when set for center causes the output to each channel to be down by 3.0dB for a constant-power law. The dc control voltage to each VCA can be created with nonlinear circuits to achieve the desired pan pot control law. Once again, the current output of the VCA makes it convenient to drive the virtual ground right and left summing busses directly without summing resistors that increase mixer noise gain.

The VCA has two control inputs, one of each control polarity. They can be driven differentially or independently. The negative control port can be driven with the pan pot control signal while the positive control port can be driven by the fader signal, implementing the fader and pan pot functions together with a single dual VCA. The ability to do this while keeping the pan position the same with different fader levels is made possible by the linear decibel control law of the VCA.

A simplified diagram of such a circuit is shown in Fig. 24-37.[9] The +1 unity gain control buffers must be able to deliver up to 27 mA at −10 V to drive the low-impedance VCA control port resistive divider. Once again, better approaches to buffering the control port signals can be found in reference 16. The resistive dividers of the panning potentiometer and its associated resistors create nonlinear control voltages as a function of rotation that are remarkably close to what is needed to implement the constant-power panning law. Power is constant within a ±0.2dB window over the full rotation of the pan pot, with maximum attenuation of the low side approaching 100dB.[9] The panning law can be changed to a −4.5 dB compromise law merely by changing R_3 and R_4 to 5200Ω.

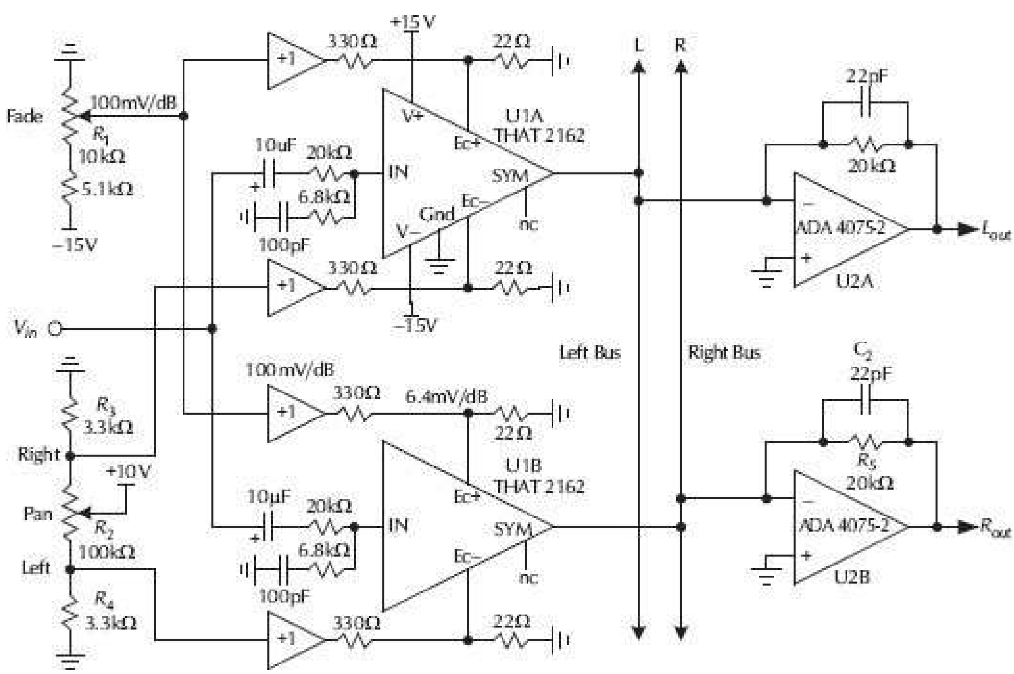


Figure 24-37. Panning and fader circuit.

24.8 Linear Power Amplifiers

Power amplifiers are used to drive the loudspeakers in sound systems. As the name implies, they create audio power. They generally must amplify the line-level signal to many tens of volts to drive the loudspeaker. For example, a 100W amplifier must generate up to 40 V peak to drive an 8Ω loudspeaker. A 1000W professional power amplifier must deliver 126 V peak into an 8Ω loudspeaker.

At the same time, large amounts of current must be supplied. The 100W amplifier must supply peak current of 5A into an 8Ω load. Loads of lower impedance, like 4Ω or 2Ω, require higher current for the same operating power level. A 1000W amplifier must deliver 16A into an 8Ω load, but it must deliver 32A peak into a 2Ω load at an operating level of 1000W.

24.8.1 Driving the Loudspeaker Load

A flat frequency response is desirable to avoid tonal coloration, but a flat response may not always be obtained when the amplifier is driving a real-world loudspeaker load. The input impedance of real loudspeakers can vary dramatically as a function of frequency, while the output impedance of the power amplifier is non-zero. The amplifier output impedance and the loudspeaker input impedance comprise a voltage divider, as illustrated in Fig. 24-38.[1] Here the amplifier is modeled as an ideal amplifier with zero output impedance in series with impedance Z_{out} that describes its actual output impedance. This is referred to as a Thevenin equivalent circuit.

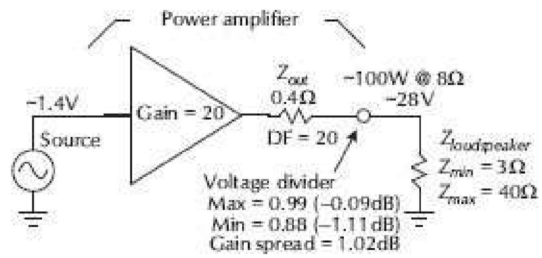


Figure 24-38. Driving loudspeaker load with varying impedance.

This is where the *Damping Factor* (DF) comes into play. In spite of its important-sounding name, this is just a different way of expressing the output impedance of the amplifier. While amplifiers ideally act like voltage sources with zero output impedance, they all have finite output impedance. The term *Damping Factor* came from the fact that a loudspeaker is a mechanically resonant system, and that the low output impedance of an amplifier damps that resonance via the resistance of the loudspeaker's voice coil and *Electromotive Force* (EMF). An amplifier with higher output impedance will provide less damping of the loudspeaker cone motion because it adds to the total amount of resistance in the circuit.

Damping factor is defined as the ratio of 8Ω to the actual output impedance of the amplifier. Thus, an amplifier with an output impedance of 0.2Ω will have a damping factor of 40. Most vacuum tube amplifiers have a damping factor of less than 20, while many solid-state amplifiers have a DF in excess of 100. It is important to bear in mind that DF is usually a function of frequency, often being larger at low frequencies. This is consistent with the need to damp the cone motion of woofers, but ignores the influence of DF on frequency response at higher frequencies. Many loudspeakers have a substantial peak or dip in their impedance at or near their crossover frequencies. This could result in coloration if amplifier DF is low.[1]

The effect of damping factor and output impedance on frequency response must not be under-estimated in light of the large impedance variations seen in many contemporary loudspeakers. It is not unusual for a loudspeaker's impedance to dip as low as 3Ω and rise as high as 40Ω across the audio band. Consider this wildly varying load against the 0.4Ω output impedance of a vacuum tube amplifier with a DF of 20. This will cause an audible peak-to-peak frequency response variation of ± 0.5dB across the audio band.

24.8.2 Amplifier Block Diagram with Feedback

A simple block diagram of a power amplifier is illustrated in Fig. 24-39. The core of the amplifier that provides all of the *open-loop gain* (OLG or A_{ol}) is shown as a gain block just like an operational amplifier. For purposes of illustration it is shown with a gain of 1000 (60dB). The feedback network is depicted as a block that attenuates the signal being fed back by a factor of 20. Suppose the output of the amplifier is 20 V. The amount fed back will then be 1 V. The input across the differential inputs of the gain block will be 20 mV if the forward gain is 1000. The required input from the input terminal will then be 1.02 V.

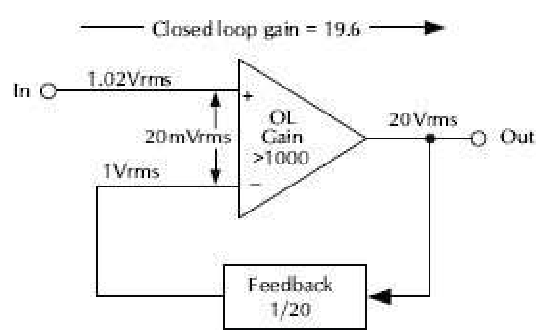


Figure 24-39. Amplifier block diagram with feedback.

This simplified approach to looking at a feedback circuit is sometimes referred to as *input-referred* feedback analysis because we start at the output and work our way back to the input to see what input would have been required to produce the assumed output.[1] The closed loop gain is thus 20/1.02 = 19.6. This is just 2% shy of what we would get if we assumed that the closed loop gain were just the inverse of the attenuation in the feedback path. Notice that the loop gain of the amplifier (the product of open-loop gain and the feedback factor) is 50, corresponding to a 2% error. The open-loop gain of most power amplifiers is greater at low frequencies and less at high frequencies, often being more than 10,000 (80dB) at 1kHz.

24.8.3 The Class AB Output Stage

A simplified class AB output stage is shown in Fig. 24-40. Q_1 delivers current to the load for positive signals and Q_2 sinks current from the load for negative signals. At no signal and for very small signals both Q_1 and Q_2 conduct current and both contribute to the output signal current in a class A region. The box labeled *bias* provides a voltage of about 2 Vb_e that spans the bases of Q_1 and Q_2 to keep them on at idle and establish the idle current, which here is 100 mA. This circuit is called a *bias spreader*.

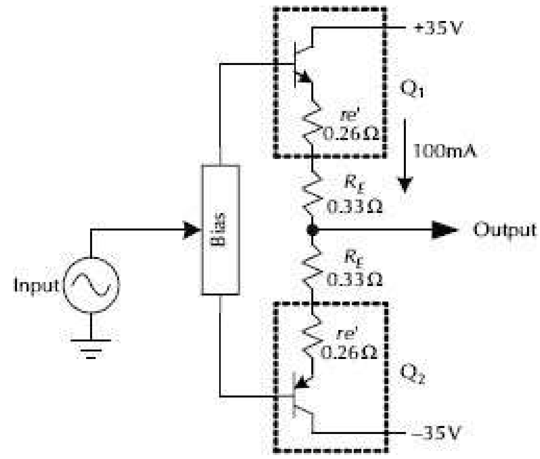


Figure 24-40. Class AB output stage.

A key observation is that the signal takes a different path through the output stage on positive and negative half cycles. If the voltage or current gains of the top and bottom parts of the output stage are different, distortion will result. Moreover, the "splice point" where the signal current passes through zero and crosses from one path to the other can be tricky, and this can lead to so-called *crossover distortion*.

As an aside, if the bias spreader is set to provide a very large output stage idle bias current, both the top and bottom output transistors will conduct on both half cycles of the signal. One will be increasing its current as the other decreases its current, with the difference flowing into the load. In this case we have a so-called *class A* output stage. The fact that the signal is then always taking the same path to the output (consisting of two parallel paths) tends to result in less distortion because there is no crossover from one half of the output stage to the other as the signal swings from positive to negative. The price paid is very high power dissipation as a result of the high output stage bias current.

The voltage gain of the output stage is determined by the voltage divider formed by the output stage emitter follower output impedance and the loudspeaker load impedance. The output impedance of each half of the output stage is approximately equal to *re′* plus R_E. This is illustrated in Fig. 24-40.

Since the two halves of the output stage act in parallel when they are both active at idle and under small-signal conditions, the net output impedance will be about half that of each side.

$$Z_{out(small\,signal)} \approx \frac{re'_{idle} + R_E}{2} \tag{24-9}$$

If the output stage is biased at 100mA, then *re′* of each output transistor will be about 0.26Ω. The summed resistance for each side will then be 0.26 + 0.33 = 0.59Ω. Both output halves being in parallel will then result in an output impedance of about 0.3Ω. Because voltage gain is being calculated, these figures assume that the output stage is being driven by a voltage source. The voltage divider action governing the output stage gain is illustrated in Fig. 24-41. If the load impedance is 8Ω, then the voltage gain of the output stage will be 8/(8 + 0.3) = 0.96. If instead the load impedance is 4Ω, the gain of the output stage will fall to 0.93.

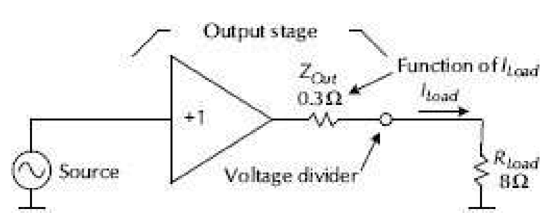


Figure 24-41. Output stage gain.

Bear in mind that the small-signal gain of the output stage has been calculated at its quiescent bias current. The value of *re′* for each of the output transistors will change as transistor currents increase or decrease, giving rise to complex changes in the output stage gain. Moreover, at larger signal swings only one half of the output stage is active. This means that the output impedance under those conditions will be approximately $re' + R_E$ rather than half that amount. These changes in incremental output stage gain as a function of output signal current cause what is called *static crossover distortion*.

$$\begin{aligned} Z_{out\,(large\,signal)} &\approx re'_{high\,current} + R_E \\ &\approx R_E \end{aligned} \qquad (24\text{-}10)$$

At high current, re' becomes very small. At 1A, re' is just 0.026Ω, much smaller than a typical value of R_E. At 10A, re' is theoretically just 0.0026Ω. That is why $\approx R_E$. If R_E is chosen so that

$$R_E = re'_{idle}$$

then

$$\begin{aligned} Z_{out\,(large\,signal)} &\approx Z_{out\,(small\,signal)} \\ &\approx R_E \end{aligned}$$

and crossover distortion is minimized by making the large-signal and small-signal output stage gains approximately equal. This is only a compromise solution and does not eliminate static crossover distortion because the equality does not hold at intermediate values of output current as the signal passes through the crossover region. This variation in output stage gain as a function of output current is illustrated in Fig. 24-42[1,18]

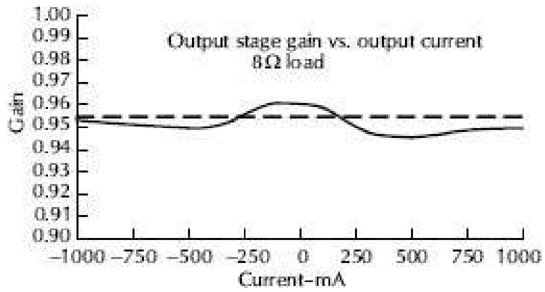


Figure 24-42. Wingspread plot.

24.8.3.1 Output Stage Bias Current

The idle bias current of the output stage plays a critical role in controlling crossover distortion. It is important that the right amount of bias current flows through the output stage, from top to bottom, when the output is at zero volts and not delivering any current to the load. Notice that together the two driver and two output transistors require at least two V_{be} voltage drops from the base of Q_1 to the base of Q_2 to begin to turn on. Any additional drop across the output emitter resistors will increase the needed bias spreading voltage.

The optimum class AB idle bias for a conventional output stage like this is that amount of current that produces a voltage drop of approximately 26mV across each of the output emitter resistors.[1,18] Recall that

$$\begin{aligned} re' &= \frac{V_T}{I_c} \\ &= 26\,\mathrm{mV}/I_c \end{aligned}$$

then,

$$\begin{aligned} I_c &= 26\,\mathrm{mV}/re' \\ &= 26\,\mathrm{mV}/R_E \end{aligned}$$

and

$$\begin{aligned} V_{RE} &= I_c \times R_E \\ &= 26\,\text{mV} \end{aligned}$$

This amount of bias current makes *re′* of the output transistor equal to the resistance of its associated emitter resistor. With 0.33Ω emitter resistors, this corresponds to about 79 mA. In the example here the output stage has been over-biased slightly to a current of 100 mA. This means that small-signal gain will be slightly larger than large-signal gain in this case. This is evident in Fig. 24-42.

There is a caveat to the assertion above that the optimum bias point occurs when 26mV is dropped across each output emitter resistor. The actual output impedance of an emitter follower is slightly greater than *re′*. The additional resistance results from physical base and emitter resistances inside the transistor. This additional resistive component acts as an extension of the external emitter resistor R_E. This means that the optimum voltage drop across the external emitter resistor may be somewhat less than 26mV. The bias spreader that usually comprises a Vb_e multiplier sets the voltage that appears across R_E. A trimmer in the bias spreader is adjusted to set the output stage bias current to the desired value.

The objective of the bias spreader design is temperature stability of the bias point of the output stage. The temperature coefficient of the voltage produced by the Vb_e multiplier should match that of the base-emitter junction voltages of the driver and output transistors. Since the Vb_e of a transistor decreases by about 2.2mV/°C, it is important for thermal bias stability that these junction drops track one another reasonably. The output transistors will usually heat up the most. Because they are mounted on a heat sink, the Vb_e multiplier transistor should also be mounted on the heat sink so that it is exposed to the same temperature. This approach is only an approximation because the temperature of the heat sink changes more slowly than that of the power transistor junctions. See chapter 14 in reference 1 for more details.

24.8.3.2 Crossover Distortion

It is generally understood that crossover distortion becomes worse at high frequencies and when lower-impedance loads are being driven. Total harmonic distortion at 20kHz (THD-20) is a better indicator of crossover distortion because there is less negative feedback available at 20kHz and its harmonic frequencies to reduce distortion. THD-20 also captures both static and dynamic (switching) crossover distortion.

Fig. 24-43 shows THD-20 as a function of power when a high-power amplifier is driving a 4Ω load. The graph illustrates how crossover distortion peaks at a relatively low power. Here the crossover distortion manifests itself as a peak in THD-20 at a power level of 11W for an amplifier that is rated at 350W into the 4Ω load. Full power distortion thus does not tell the whole story.

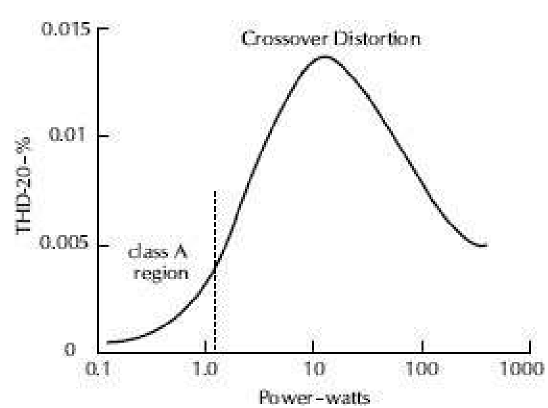


Figure 24-43. THD-20 vs. power.

Fig. 24-43 also illustrates where the transition occurs from class A operation to class B operation for this amplifier when driving the 4Ω load. With an idle bias current of 400 mA in the output stage, this amplifier can drive about 800mA into the load before either the top or bottom power transistors go into cutoff. This 800mA of peak current into a 4Ω load corresponds to 1.28W.

24.8.4 Simple Power Amplifier

Fig. 24-44 illustrates a very simple power amplifier that serves to explain the operation of the most common topology of audio power amplifier.[1]

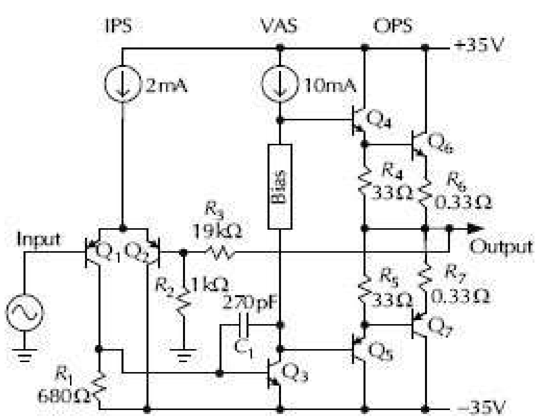


Figure 24-44. Simple power amplifier.

24.8.4.1 Input Stage

Transistors Q_1 and Q_2 form the input stage differential pair. This arrangement is often called a *Long Tailed Pair* (LTP) because it is supplied with a so-called *tail current* from a very high-impedance circuit like the current source shown. The amplifier's input stage is referred to as the *IPS*. The input stage usually has a fairly low voltage gain, typically ranging from 2 to 15.

The IPS compares the applied input signal to a fraction of the output of the amplifier and provides the amount of signal necessary for the remainder of the amplifier to create the required output. This operation forms the essence of the negative feedback loop. The fraction of the output to which the input is compared is determined by the voltage divider consisting of R_3 and R_2. If the fraction is 1/20, and the forward gain of the amplifier is large, then very little difference need exist between the input and the feedback signal applied to the IPS in order to produce the required output voltage. The gain of the amplifier will then be very nearly 20. This is referred to as the *closed loop gain* of the amplifier (CLG or A_{cl}).

In practice, most well-designed amplifiers incorporate emitter degeneration in the input stage, often reducing transconductance by a factor of 5–10. This increases linearity and reduces the value of C_1 required for stability. This in turn increases amplifier slew rate.[1]

24.8.4.2 The VAS

Transistor Q_3 in Fig. 24-44 forms what is called the *Voltage Amplifier Stage*, or VAS. It is a high-gain *common-emitter* (CE) stage that provides most of the voltage gain of the amplifier. It is loaded with a current source rather than a resistor so as to provide the highest possible gain. It is not unusual for the VAS to provide a voltage gain of 100 to 10,000. This means that the difference signal needed to drive the input stage does not need to be very large to drive the output to its required level.

24.8.4.3 Output Stage

The output stage (OPS) is composed of transistors Q_4 through Q_7.

Its main job is to provide buffering in the form of current gain between the output of the VAS and the loudspeaker load. Most output stages have a voltage gain of slightly less than unity. The output stage here consists essentially of two pairs of *emitter followers* (EF), one for each polarity of the output swing. This is called a complementary push-pull output stage. Transistors Q_4 and Q_5 are referred to as the *drivers*, while Q_6 and Q_7 are the output devices.

The two-stage OPS will typically provide current gain between 2500 and 10,000 if the transistors have β between 50 and 100. This means that an 8Ω load resistance will look like a load resistance between 2kΩ and 80kΩ to the output of the VAS. Other output stages, like so-called *Triples*, can provide current gain of 100,000 to 1 million, greatly reducing the load on the VAS.[1]

24.8.5 A More Complete Amplifier

The amplifier described has been simplified in some ways. There are some necessary additions to make it complete. This added circuitry is shown in Fig. 24-45, where the core amplifier is merely shown as a block of open-loop gain. Feedback network resistors R_2 and R_3 that were included in the core amplifier designs above are shown here for clarity.

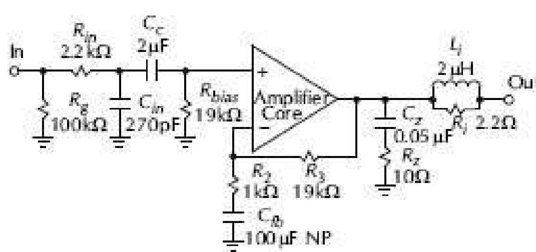


Figure 24-45. A more complete amplifier.

24.8.5.1 Input Network

Resistor R_{in} and capacitor C_{in} form a first-order input low-pass filter that keeps out unwanted radio frequencies. This filter is often designed to limit the bandwidth of the complete amplifier to a frequency that is somewhat less than the actual closed-loop bandwidth of the amplifier-proper. A typical 3-dB frequency for the input filter might be 300kHz, as shown in Fig. 24-45 with R_{in} = 2.2kΩ and C_{in} = 270pF. Resistor R_g keeps the input terminal from floating at dc.

Coupling capacitor C_c blocks any dc that might be present from the source, while R_{bias} provides a return path for the input bias current of the amplifier's input stage. This keeps the non-inverting input node of the amplifier near 0 V. Notice that if Q_1 in the amplifier is biased at 0.5 mA and it has β = 100, its base current will be approximately 5 μA. This will cause a voltage drop of 95mV across R_{bias}. The value of R_{bias} will often be set to equal that of R_3 so that voltage drops across these two resistors balance out any dc offset. In this case, R_{bias} would be set to 19kΩ. The capacitance of C_c against R_{bias} forms a high-pass filter whose 3-dB frequency we wish to keep below 5Hz. Here a 2pF capacitor provides a lower 3-dB frequency of 4Hz. This capacitor should be of high quality.

24.8.5.2 Feedback ac Decoupling Network

The core amplifier illustrated in Fig. 24-44 is dc coupled. It has the same gain of 20 at dc as for ac signals. This means that any offset at the input of the amplifier will be amplified by 20 times. In the case where there is an input R_{bias} whose voltage drop is not fully compensated in some way, that voltage will be amplified by a factor of 20. A 10mV offset will become a 200mV offset at the output. This is excessive. For this reason capacitor C_{fb} is placed in series with R_2. This causes the gain of the amplifier to fall to unity at dc, greatly reducing output offset voltage.

C_{fb} in combination with R_2 creates a high-pass filter effect, causing amplifier gain to fall at low frequencies. Once again, we desire that the 3 dB frequency of this filter be at 5 Hz or less. Because $C_{\beta b}$ is working against a smaller resistor, it will have to be a significantly larger value than was used for C_c. Because R_2 at 1kΩ is 19 times smaller than R_{bias} at 19kΩ, the value of C_{fb} will have to be 19 times as large to achieve a comparable low-frequency cut-off. This comes out to 38 μF. In practice, a 100 μF non-polarized electrolytic capacitor would typically be used for C_{fb}. Once again, capacitor quality comes into play with few easy solutions here. The quality of this capacitor is every bit as important as that of coupling capacitor C_c. A non-polarized electrolytic with a voltage rating of at least 50 V will work well. As an alternative, dc servo circuits that can eliminate the need for C_{fb} are described in reference 1.

24.8.5.3 Output Network

Most solid-state power amplifiers include an output network to keep them from becoming unstable under unusual load conditions. The emitter follower output stage itself can sometimes become unstable at high frequencies under a no-load condition. This problem is avoided by inclusion of the shunt *Zobel* network consisting of R_z and C_z. This network assures that at very high frequencies the output stage is never loaded by less than the load provided by R_z. Typical values for R_z and C_z are 10Ω and 0.05 μF, respectively.

Capacitive loads can de-stabilize an emitter follower stage or a feedback loop. For this reason, most amplifiers incorporate a parallel R-L isolating network in series with the output of the amplifier. At very high frequencies the impedance of the inductor becomes large, leaving the series resistance of the resistor to isolate any load capacitance from the emitter follower output stage. Typical values of L_i and R_i might be 2 μH and 2.2Ω.

24.8.6 Power Dissipation

The power dissipation of a class AB output stage is greatest at an intermediate value of power, often at about 1/3 power. At this power level, the output transistors spend more time with both significant current through them and voltage across them. Fig. 24-46 shows the ideal and actual power dissipation of a 100W amplifier as a function of output power. The actual dissipation is higher because the output stage cannot swing all the way to the rail and it has other losses, such as those in the emitter resistors and drivers. The IPS and VAS also consume a fixed amount of power.

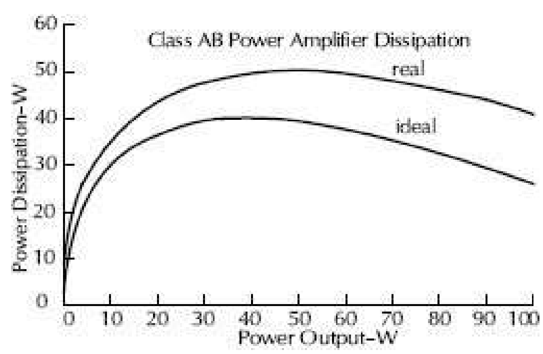


Figure 24-46. Power dissipation vs. output power.

24.8.7 A 225W Amplifier

Fig. 24-47 shows a practical 225W amplifier. The design employs 64 V power rails and 4 pair of output transistors. The overall design is largely a refined and completed version of the simple topology illustrated in Fig. 24-44. The IPS, VAS and OPS are easily identified, as are the input and output networks. In most amplifiers with paralleled output transistors, so-called *base stopper* resistors are placed in series with each base. These resistors, often on the order of 2–5Ω are not shown for clarity. Protection circuits are also not shown.

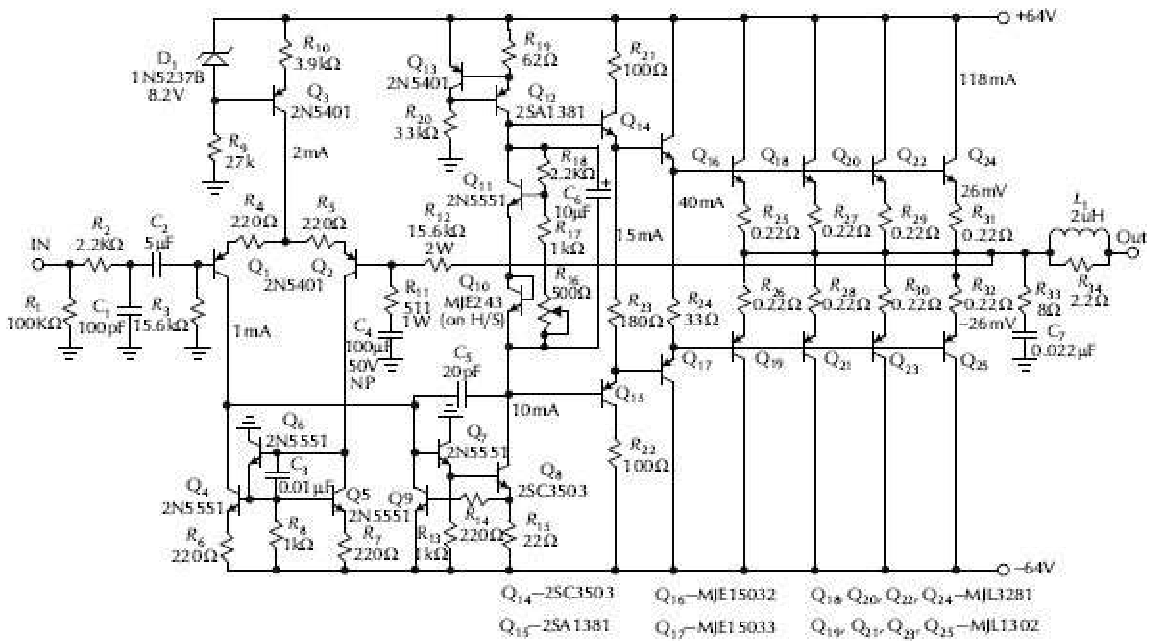


Figure 24-47. 225W amplifier.

24.8.8 Bridged Amplifiers

Bridged amplifiers are used widely where high power is required. This is especially so in the pro-sound arena. A simple bridged amplifier arrangement is illustrated in Fig. 24-48. Two channels of a stereo amplifier are driven out-of-phase and the loudspeaker is connected across the hot outputs of the two amplifiers. A bridged power amplifier can theoretically produce four times the power into a given load compared to its non-bridged counterpart (one channel of the stereo amplifier) using the same rail voltages. This is because the voltage across the loudspeaker is doubled and power goes as the square of voltage. Under these conditions, each of the two amplifiers "sees" an effective load resistance of half that of the loudspeaker impedance. For this reason, the bridged amplifier may produce somewhat less than four times the power.

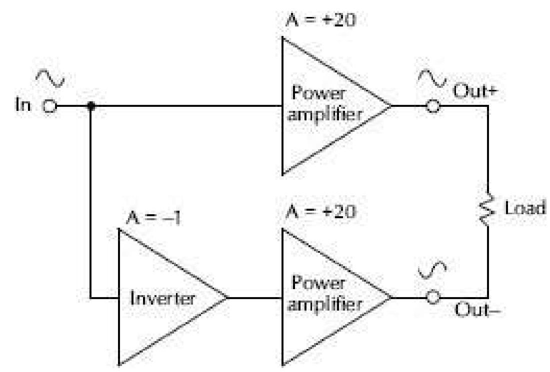


Figure 24-48. Bridged power amplifier.

Power Supply Advantages

Amplifiers operating in bridged mode often enjoy some advantage in regard to the power supply. Lower rail voltages are required for a given power output. This also relaxes output transistor voltage rating requirements. Single-ended amplifiers draw power from either the positive rail or the negative rail on each half cycle of the signal. This current flows through ground. The duty cycle of current flow from each rail is only 50%. In bridged mode, current is being drawn from both rails simultaneously and little or none of the output current flows through ground. Each rail is being used 100% of the time. The waveform of the current being sourced by each rail is full-wave rather than half-wave.

24.8.9 Class G Amplifiers

Heat is always a concern, especially in high-power professional power amplifier applications. The conventional class AB power amplifier is not very efficient. Considerable power is dissipated when the power transistors are conducting high current with high voltage across them. The class G amplifier was developed to address this issue by reducing the voltage across the main output transistors when it is not needed. The waveforms depicted in Fig. 24-49 show the output transistors operating with a 40V rail under low power conditions.[1,19] If the signal voltage becomes too high to be supported by this rail voltage, an upper (outside) set of power transistors lifts the rails in proportion to the larger output voltage demanded. When this transition occurs, the main output transistors operate with a constant collector-emitter voltage.

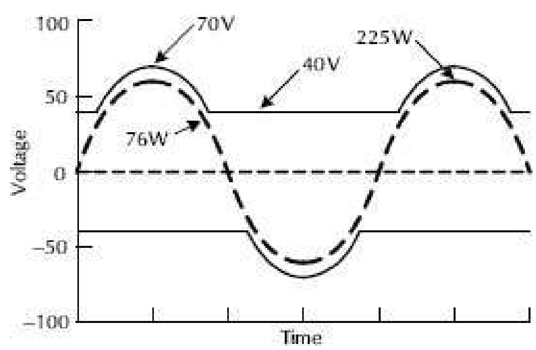


Figure 24-49. Class G waveforms.

Fig. 24-50 shows a simplified class G output stage that illustrates the concept. At low power levels, Q_1 is supplied from the +40V rail via commutating diode D_1. If the drive signal becomes large enough at higher power, the base of Q_2 will be driven above the 40 V rail voltage and lift the collector voltage supplied to Q_1. The floating voltage source determines the point at which the collector voltage of Q_1 is raised and determines how much voltage headroom will be made available to Q_1.

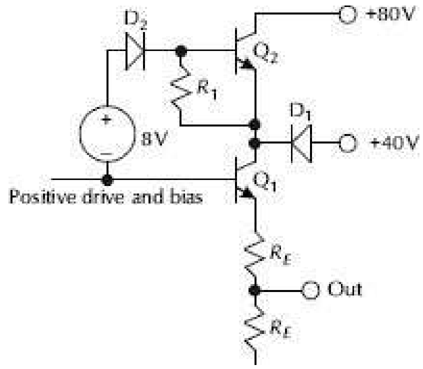


Figure 24-50. Class G concept-positive half.

Fig. 24-51 shows an actual class G output stage.[1] In this design the drive to the outer transistors Q_{11} and Q_{12} is derived from the output signal in a bootstrap fashion. Zener diodes D_5 and D_6 establish the drive offset to the bases of Q_{11} and Q_{12} and hence the headroom for output transistors Q_5 and Q_6 when large signal swings are present. The output stage is a Locanthi T circuit triple emitter follower. The collectors of pre-drivers Q_1 and Q_2 are bootstrapped by Q_7 and Q_8, while the collector voltages to the drivers Q_3 and Q_4 are boosted by Q_9 and Q_{10} when large signal swings are present. D_3 and D_4 prevent the base-emitter junctions of Q_9 and Q_{10} from being excessively reverse-biased.

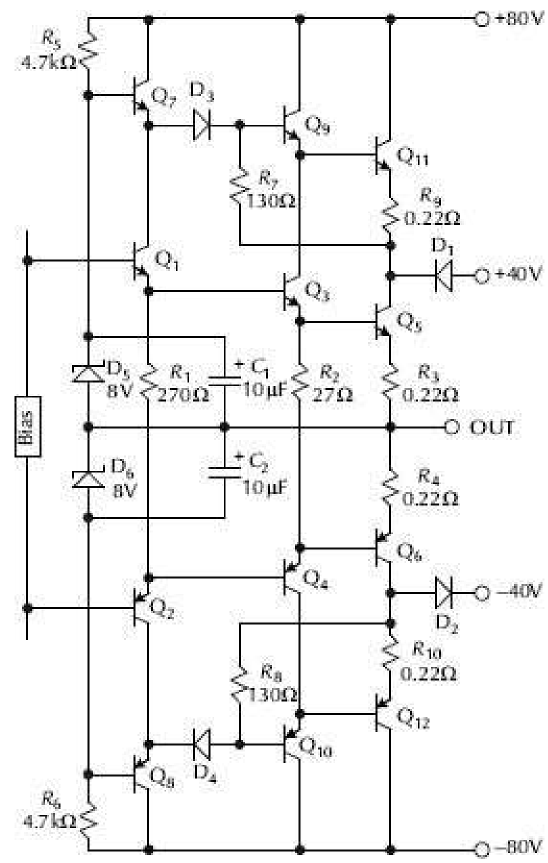


Figure 24-51. Class G output stage.

Fig. 24-52 shows power dissipation as a function of power output for comparable 300W class AB and class G amplifiers. The class G amplifier maintains its greatest dissipation advantage in the power region where program material spends most of its time.

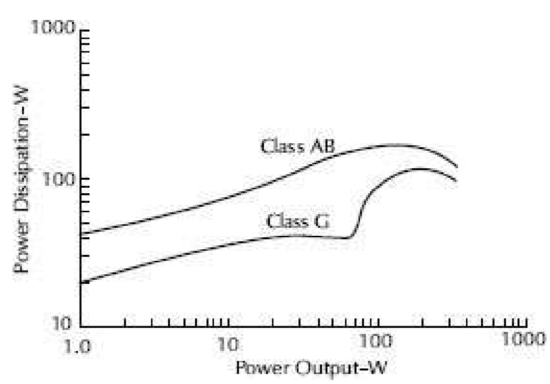


Figure 24-52. Class AB and class G power dissipation.

24.8.10 Class H Amplifiers

Class H amplifiers operate with a similar concept to class G, but the power supply rails to an otherwise-conventional class AB output stage are abruptly switched to a higher voltage when the signal amplitude demands it.[19,20] This sudden switching can lead to distortion due to shoot-though of the switching edge as a result of the limited PSRR of the output stage.

24.8.11 Protection Circuits

Protection circuits serve to protect the loudspeaker and the power amplifier, with the former being the first priority. In some amplifiers, protection is little more than a fuse in series with the output. However, this is usually not sufficient to protect the output devices from failure due to exceeding their safe area of operation (SOA). In such cases, failure of the output devices can occur before the fuse can blow. For this reason, active protection of the output stage is employed to guard against short circuits and SOA failures. Sometimes a simple current limiter will suffice, but more often a V-I limiter is used whose current-limiting set point is a function of the voltage across the output transistor. Such a V-I limiter is shown in Fig. 24-53.

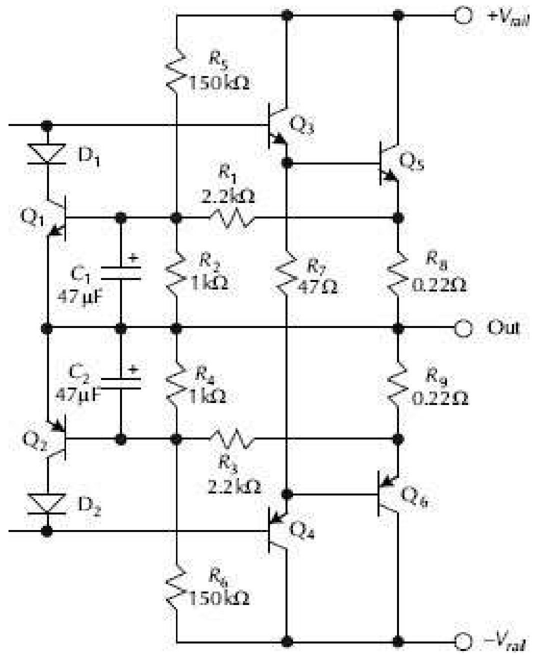


Figure 24-53. Single-slope V-I limiter.

Looking at the top half, the output current is sensed by R_8 and a high current will turn on Q_1 and rob the output stage of base drive. The VAS drive circuit is presumed to be capable of only limited current. D_1 prevents the collector-base junction of Q_1 from being forward-biased on negative signal excursions. C_1 provides some delay in the current-limiting action to accommodate short current transients that would not endanger the output transistors. The current-limiting threshold is primarily set by the ratio of R_1 and R_2. When the attenuated voltage drop across R_8 exceeds one V_{be}, current limiting begins. R_5, however, injects a current into the circuit that increases with the V_{ce} of Q_5, reducing the current limiting threshold when Q_5 is subjected to high voltage.

Returning to the top-priority—protecting the loudspeaker—a simple fuse is often not sufficient, and its tendency to blow sometimes needlessly is problematic. Instead, a relay may be driven from a circuit that senses dangerous conditions at the amplifier output, such as excessive dc, and opens the relay. The relay is also often used to mute the output to prevent turn-on and turn-off transients and thumps. In other designs a crowbar circuit is used to short the output of the amplifier to protect the loudspeaker from serious danger. While it seems destructive to use a crowbar, bear in mind that the amplifier should be designed to tolerate an output short circuit anyway, at least for the time it takes for a rail fuse to blow. A common amplifier failure is an output transistor that has shorted, placing the full rail voltage on the output. In this case it can be argued that a crowbar will do no significant further damage anyway.

24.9 Class D Power Amplifiers

Class D amplifiers operate on an entirely different principle from linear amplifiers. The output stage in a class D amplifier comprises switches that are either *on* or *off*. The switches apply the positive supply to the output for one brief period and then connect the negative supply rail to the output for the next brief period. The process is then repeated indefinitely. This results in a square wave at the output. If the positive and negative intervals are the same, the net output is zero. If the first is longer than the second, the output has a net positive value. A low pass filter extracts the average value to drive the loudspeaker. The cutoff of the LPF often lies in the 30–60kHz range.

This process is referred to as pulse width modulation (PWM). These switching intervals alternate at a high frequency, often in the range of 500kHz. Thus, the average value of the square wave drives the load. Because the switches are either *on* or *off*, they dissipate no power. Virtually all of the input power from the power rails is transferred to the load, so efficiency is very high and power dissipation is very low. Efficiency of 85% to 95% is not uncommon. Small amplifiers that run cool are the result. A big challenge in class D amplifiers is the proper driving of the output stage switches so that the *on* and *off* timing intervals accurately reflect the input signal.

Class D amplifiers have long suffered from poor distortion performance. Matters have improved dramatically since the late 1990s. The need to squeeze more power capability in to a smaller space while generating less heat has driven their development. The implementation of class D amplifiers is far more involved than the simple description above. See reference 1 for more coverage than is presented here. Although the *D* in class D does not stand for digital, it is true that the implementations of class D amplifiers are moving more toward digital. In fact there are approaches that involve direct digital conversion from PCM input streams to class D audio outputs.

There are numerous ways to build a class D power amplifier, but those based on Pulse Width Modulation (PWM) are the oldest and still very popular. Other approaches will be touched on here as well. The emphasis here will be on class D amplifiers designed for high sound quality, as opposed to smaller or more efficient amplifiers designed for more pedestrian applications like hand-held consumer electronics.

24.9.1 The PWM Modulator

Fig. 24-54 shows a simple arrangement that converts an analog input to a digital PWM signal.[21,22,23] This circuit is referred to as a PWM modulator. A triangle waveform at several hundred kHz is applied to one side of a comparator while the input signal is applied to the other side. Whenever the input signal is more positive than the reference triangle wave, a positive pulse is produced that lasts as long as the input signal is above the threshold set by the triangle wave. With a perfect triangle wave it is easy to see that the pulse width is linearly proportional to the input amplitude. Conversely, when the input signal is below the time-varying threshold set by the triangle wave, the output of the comparator is negative.

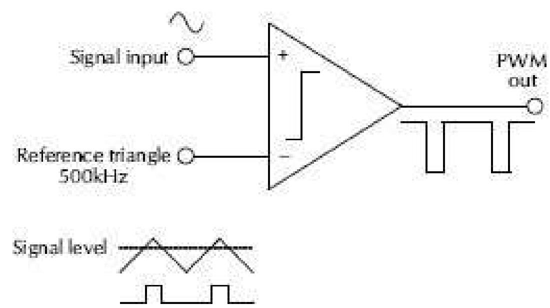


Figure 24-54. Simple PWM modulator.

The output of the comparator is thus a square wave whose duty cycle corresponds to the amplitude of the input signal voltage. The frequency of the square wave is referred to as the carrier frequency. It is easy to see that the average value of the square wave is an accurate representation of the input signal. If the comparator output is used to drive power MOSFETs on and off, the average output will reflect the input signal value multiplied by the power supply rail voltage. The average output is extracted from the high-power pulse stream by a low-pass output filter, as shown in Fig. 24-55. The technical term for this process of retrieving the analog signal from a pulse train is called reconstruction; a discrete-time switched signal is converted to a continuous-time signal. The output filter also suppresses high-frequency EMI that is a part of the square wave.

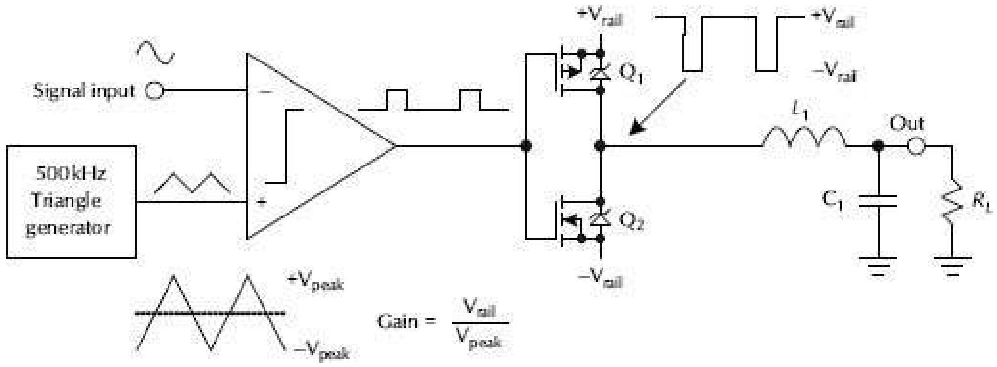


Figure 24-55. PWM class D amplifier.

The output stage shown here employs complementary MOSFETs. An N-channel MOSFET is used for the low side switch and a P-channel MOSFET is used for the high side switch. Both of these devices are in a common-source configuration. This is a simplified conceptual arrangement. The gates of the devices connected this way would not withstand the voltages typically used in a power amplifier.

The gain of this amplifier is equal to the ratio of the power supply rail voltage to the peak voltage of the triangle wave. This means that power supply rejection (PSRR) is poor. If the peak value of the triangle wave is 2 V and the peak value of the input signal is just equal to 2 V, the output of the comparator will be high all of the time and the positive rail will be connected to the load for 100% of the time. If the rail voltage is 40 V, the peak output will be 40 V and the gain of the amplifier will be seen to be 20. The positive duty cycle of the square wave is reflective of modulation depth, which is the relative degree of departure from a 50% duty cycle.

24.9.2 Class D Output Stages

The output stage is where many of the challenges lie in class D amplifiers.[22,24,25] This is where the power supply rail voltages are alternately switched onto the output bus at very high frequencies with very fast rise and fall times. A highly simplified class D output stage is shown in Fig. 24-55. The single-ended output arrangement is commonly referred to as a *half bridge*. Typical PWM carrier frequencies are in the 500kHz range while typical rise and fall times are in the 20ns range. In an amplifier with ± 50 V rails where the output transitions through 100 V in 20ns, the voltage rate-of-change at the output switching node is about 5000V/μs. To put this in perspective, 2.5A is required to drive a 500pF capacitor at this voltage rate-of-change (slew rate).

24.9.2.1 H-bridge Output Stages

The arrangement shown in Fig. 24-56 is what is called a *full bridge* or an *H-bridge*. This circuit is driven so that when one side is high the other side is low, doubling the output voltage available to the load. Interestingly, the H-bridge conveniently has an “off” position where no current flows in the load if both sides of the bridge are high or low. Some more sophisticated modulators can take advantage of this third output state.

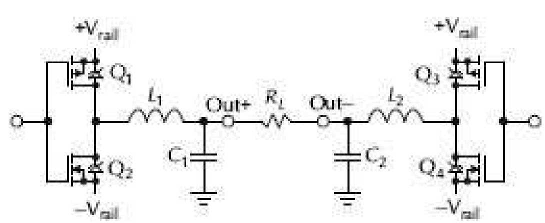


Figure 24-56. Full bridge class D output stage.

The full bridge requires twice as many components, but only half the power supply voltage to realize a given output power capability. This is analogous to linear amplifiers that are bridged. The full bridge also has some technical advantages in regard to the class D amplification process that will be discussed later. MOSFETs with desirable switching characteristics are more readily available with lower voltage ratings (below about 150 V), so high-power class D amplifiers will often employ a full bridge.

24.9.2.2 N-channel Output Stages

N-channel MOSFETs inevitably have better switching characteristics than P-channel devices. For this reason, output stages often employ N-channel devices for both the low side and high side switches, as shown in Fig. 24-57. These designs require more complex drivers that include level shifters and boost supplies.[21,24] The gate drive for the high side N-channel MOSFET must float on top of the output signal, since it is referenced to the source of the device. The boost supply is required because the high side N-channel switch requires gate drive voltages above the positive rail in order to turn on. Fortunately, integrated circuit drivers are available that take care of most of the complexity. An example is the International Rectifier IR2011.[24]

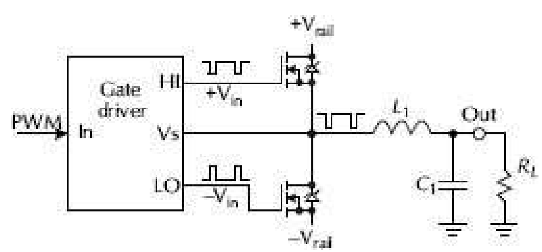


Figure 24-57. Half bridge N-channel output stage.

24.9.2.3 Gate Drive Control

The input impedance of a MOSFET gate is very high at low frequencies, but becomes much lower at high frequencies as a result of gate-source (C_{gs}) and gate-drain (C_{gd}) capacitances, which can be quite high. These gate capacitances must be charged and discharged at a very high speed in class D output stages. Further adding to the gate drive burden is the Miller effect involving the gate-drain capacitance. Peak gate drive currents can lie in the ampere range even in a 100W class D amplifier. Further complicating this is the need to provide highly precise gate timing to minimize distortion.

24.9.2.4 Dead Time Control

When switches are connected from both the positive and negative rails to a single output node there is always the possibility that both switches will briefly be *on* at the same time. This will cause shoot-through current to flow directly from the positive supply to the negative supply. At minimum, this will result in wasted power. In some cases it will result in the destruction of the output stage. For this reason there is a very small dead zone incorporated into the MOSFET drive circuit. This ensures that there is always a very small time when both devices are turned off. With slight variations in timing there should never be a situation where both devices are *on* simultaneously. Unfortunately, the presence of this dead time is a cause of distortion similar in nature to crossover distortion.[24]

Assume that the low side has been on and sinking current from the load through the output inductor. The magnetic field in the inductor represents stored energy. When the low side turns off, the output voltage will quickly transition from the negative rail to the positive rail as the magnetic field collapses. This will forward bias the substrate diode in the top MOSFET and cause *commutation* current to flow into the positive supply, even while the top transistor itself is still off due to any dead time. Commutation current will continue to flow in the same direction through the inductor for some time because inductors seek to keep current flowing in the same direction. That current will continue to flow after the top transistor has turned on.

24.9.3 Negative Feedback

Closing a feedback loop around a class D power amplifier can be challenging.[23,26] The feedback can be taken from before or after the output filter. The use of negative feedback is especially important for improving the very poor PSRR of many class D amplifier architectures. Class D amplifiers that run open loop suffer high distortion.

24.9.3.1 Pre-filter Feedback

If the feedback is taken from before the filter, the signal has not yet been reconstructed. Even at this point there is some effective phase delay due to the sampled data nature of the forward path. Some low-pass filtering or integration must be placed somewhere in the loop for purposes of reconstructing the feedback signal. The signal must be converted from a discrete-time representation to a continuous-time representation (analog). Pre-filter feedback will not reduce output filter distortion or improve high-frequency damping factor.

In Fig. 24-58 an inverting Miller integrator is placed in front of the class D amplifier. The analog input is applied through R_1 while the switched PWM feedback is applied through R_2. The forward path integrator serves the need for reconstruction of the raw switched output that is fed back. The closed loop gain is simply the ratio of R_2 to R_1. The feedback gain crossover frequency is at $\omega_0 = k/R_1 \times C_M$ where k is the forward gain of the class D amplifier. This is a fairly conventional dominant pole approach to compensation of the feedback.

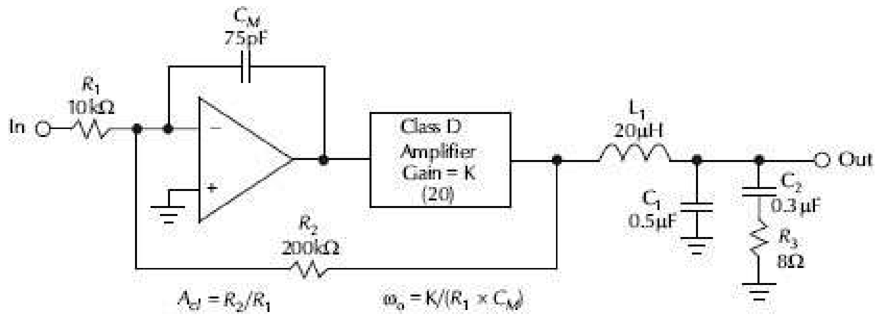


Figure 24-58. Class D amplifier with pre-filter feedback.

There is inevitable delay between the input signal and the output signal in a PWM amplifier, even without considering the effects of a reconstruction filter. The delay is a result of the sampling process that is fundamental to analog-to-PWM conversion. The analog input signal is sampled twice per period of the PWM carrier, once when it slices the positive slope of the triangle reference and once when it slices the negative slope of the triangle. A change in the analog signal will not result in a change in the PWM pulse ratio until an average of one-fourth sample period has elapsed. This added delay introduces instability if the feedback gain crossover frequency is too high. That in turn limits the effectiveness of the feedback in reducing high-frequency distortion. A big disadvantage of pre-filter feedback pick-off is that it does nothing about filter distortion or filter degradation of damping factor.

24.9.3.2 Post-filter Feedback

Post-filter feedback will reduce filter distortion and preserve damping factor, but the loop will experience quite a bit of additional phase lag from the output filter and it will be difficult to achieve loop stability. If the closed loop bandwidth of the feedback loop is restricted to achieve adequate stability, then there may not be enough loop gain at the higher audio frequencies to make much of an improvement.

To make the best use of post-filter feedback the PWM carrier frequency and the filter cutoff frequency should be made as high as possible. The fact that the output filter has already reconstructed the feedback signal is helpful. This can reduce or eliminate the role of the forward path integrator. Such an approach is illustrated in Fig. 24-59. Note that a second, smaller, LPF has been added outside the loop for reduced EMI.

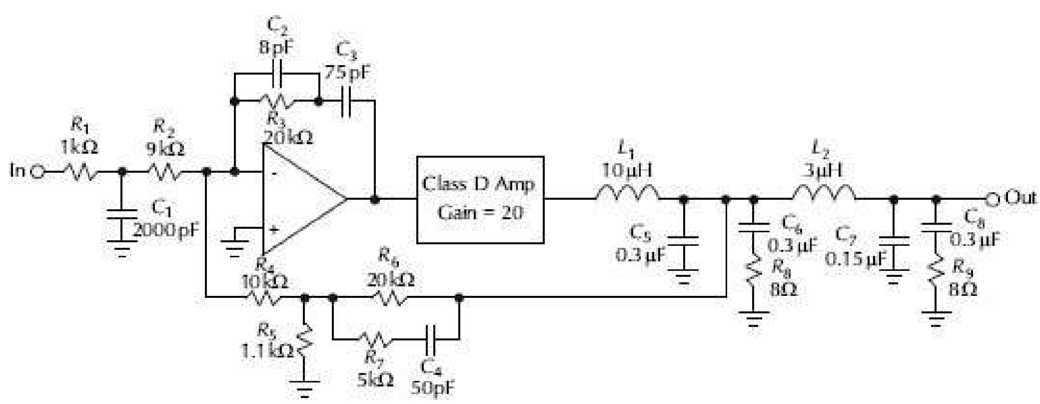


Figure 24-59. Post-filter feedback.

24.9.3.3 Self-oscillation with the Output Filter

It can be a struggle to close the feedback loop around the output filter and maintain adequate feedback stability in light of the phase shift introduced by the output filter. The PWM approach in reference 27 instead takes advantage of this “problem” by allowing the loop to oscillate in a controlled fashion. It is a self-oscillating design wherein the instability resulting from post-filter feedback is put to good use.[27,28] The simplified amplifier illustrated in Fig. 24-

60 oscillates at about 400kHz, over ten times the corner frequency of the second-order output filter. At this frequency the phase shift of the filter is nearly 180 degrees, so it does not take much additional phase shift to push the loop into oscillation. This additional phase shift is provided by the switching delay in the forward path. A phase lead network is placed in the feedback path to control the frequency of oscillation.

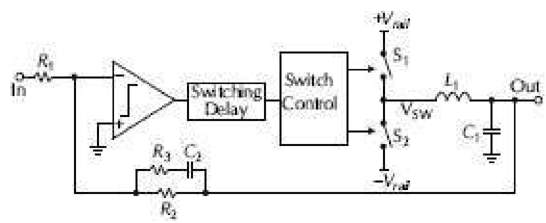


Figure 24-60. Self-oscillating PWM amplifier.

The stroke of genius here is that applying an input signal to an otherwise-oscillating class D amplifier will result in the appropriate PWM duty cycle signal corresponding to the audio waveform. The filter frequency is about 35 kHz and the oscillation frequency is about 400kHz. The loop gain is almost independent of power supply voltage and depends almost solely on the frequency response of the output filter. Loop gain is about 30dB, flat across the audio band out to about 35kHz. The gain crossover frequency is at about 200kHz, while the closed loop frequency response is down 3 dB at about 40kHz. This design profits from post-filter feedback, reducing filter distortion and providing good load invariance.

24.9.4 Bus Pumping

Bus pumping represents the transfer of energy from one power supply rail to the opposite power supply rail when a half bridge output stage is employed.[21] Assume that the low side has been on and sinking current from the load through the output inductor. The magnetic field in the inductor represents stored energy. When the low side turns off, the output voltage will quickly transition from the negative rail to the positive rail as the magnetic field collapses. Commutation current will continue to flow in the same direction through the inductor for some time because inductors seek to keep current flowing in the same direction.

The commutation current flowing in the inductor now must flow through the upper commutation diode into the positive supply, in a direction that actually seeks to make the positive supply more positive. It tends to pump up the positive rail. The process represents a lossless transfer of energy from the negative supply to the positive supply. If the duty cycle of the square wave is negative over a long period of time (as with a low frequency output), more energy will be transferred from the negative supply to the positive supply during this period, representing an average transfer of energy from the negative supply to the positive supply. If the positive supply cannot absorb this energy, its voltage will rise as a result of the pumping.

Bus pumping depends on the reactive behavior of the output inductor, but the reactive nature of the loudspeaker can also contribute to bus pumping. The effects of bus pumping are worse at low frequencies because the reservoir capacitors can change their voltage over the signal cycle if the power supply cannot absorb the pump current. If there is another source of heavy current draw from the supply in excess of the commutation current that is pumping the supply, then there will be no problem.

Fig. 24-61 shows the amount of current flowing back into the positive supply as a function of PWM duty cycle. The pump current is greatest at a duty cycle of about 25%, representing a fairly strong negative output current. The positive supply must be able to sink this current or its voltage will rise. Negative values on the Y-axis indicate current that the positive supply must source when the load is being driven positive. These values are shown for context. The values shown are for a half bridge output stage with ±50V rails driving a 4Ω load.

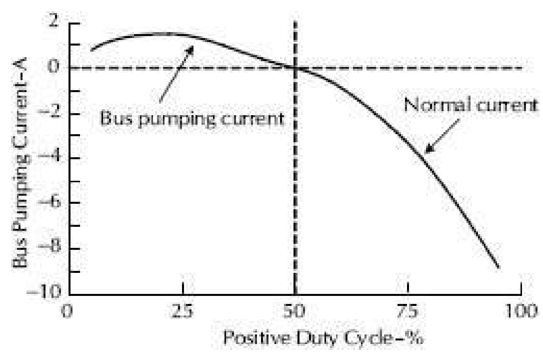


Figure 24-61. Bus pumping current vs. modulation depth.

The full bridge is largely immune to the bus pumping effect because the long-term current of the low side from the negative rail is matched by a similar long-term current of the high side from the positive rail that flows on the other side of the bridge. The commutation current is thus returned to the other rail. In effect, the commutation current flows through a closed loop.[21]

Power supply variation due to bus pumping must be avoided because any variation will influence the audio signal as a result of the 0dB PSRR of the PWM class D output stage. Bus pumping is mainly a problem at low frequencies where the rail capacitance may not be large enough to suppress rail voltage increases. Large reservoir capacitors help to mitigate these effects. Some switching power supply architectures are able to absorb the pump current with very little loss and return it to the opposite supply.

24.9.5 *Power Supply Rejection*

Poor PSRR is the Achilles heel of most class D amplifier designs. In some cases PSRR is literally 0dB.[21,24] When either output switch is on, one of the power supply rails is connected directly to the output of the amplifier. This means that there is virtually no power supply rejection inherent to the class D output stage. Indeed, power supply ripple and noise is sampled by the class D process and applied to the output. Recall that the gain of the simple class D amplifier is also defined in terms of the ratio of the power supply rail voltage to the peak voltage of the triangle reference wave. Changes in power supply voltage thus modulate the gain of the output stage.

All of this means that the power supply for the class D output stage must be very quiet and well regulated. Any variation in the power supply voltage caused by the signal will result in signal-dependent gain variations and thus intermodulation distortion. This effect is not mitigated by the use of a full bridge output stage. Several techniques are available to improve power supply rejection. Most of these involve negative feedback in one form or another.[1]

24.9.6 *The Output Filter*

The raw output of the class D output stage is a sharp-edged rectangular waveform whose pulse width varies. Rise times of the switched output pulses are often between 5ns and 40ns. The output filter serves two very important purposes. First, the filter extracts the low-frequency average from the high-frequency rectangular waveform to provide the audio output signal. In this regard, it is a PWM-to-analog converter. Secondly, it must filter out the very high frequency carrier and its harmonics to prevent EMI radiation from the amplifier. The filter must be designed to operate at high current with low distortion. The radiation of the filter itself within the amplifier must also be considered carefully as this could cause distortion in the analog circuits.

The simplest output filter is the second-order filter with a single inductor and capacitor. Its response will fall off at 12dB per octave (40dB/decade) above its cutoff frequency. It is not unusual to employ an output filter with a cutoff frequency that is about a decade below the PWM carrier frequency. This will provide about 40dB of attenuation at the carrier frequency and about 60dB of attenuation at the third harmonic of the carrier frequency.

Proper operation of the output filter depends on it being loaded with the correct impedance. Unfortunately, the impedance of loudspeakers is all over the map. This is of particular concern at high frequencies where the filter action is taking place. Loudspeakers can become inductive at high frequencies, as can the speaker cable driving them. For this reason a Zobel network is often placed at the output of the filter so that at high frequencies there appears a resistive load.[26] This network will usually comprise an 8Ω resistor in series with a capacitor, much like the Zobel network used in linear amplifiers.

24.9.7 Input Filter and Aliasing

All class D amplifiers involve a sampling process of one kind or another applied to the analog input signal. This means that aliasing is always possible. For this reason the input filter that is commonly found on linear power amplifiers takes on even greater importance. The input filter need not necessarily have a low cutoff frequency, but it must have good attenuation at frequencies approaching half the PWM carrier frequency and above.

24.9.8 Sigma-Delta Modulators

The PWM class D amplifier is essentially an open-loop device around which negative feedback may be applied. The sigma-delta modulator is an alternative means of generating a bit stream whose average value corresponds to the signal amplitude.[29] Negative feedback is intrinsic to its operation. The basis of a commercial implementation of a sigma-delta class D amplifier is described in reference 30.

Fig. 24-62 illustrates a simple first-order sigma-delta modulator (SDM or ΣΔ). It is operated at a fixed frequency. However, instead of producing pulses at a carrier frequency whose width is continuously variable, it produces pulses whose width is in discrete increments of the clock period. This is a form of pulse density modulation (PDM). Because the pulses are in increments of the clock period, this signal is quantized in time. As a result, quantization noise is introduced. This is a problem that must be addressed. The simple answer is that the clock for the sigma-delta modulator is at a much higher frequency than the typical PWM carrier frequency. The ΣΔ modulator fundamentally depends on oversampling, wherein the sampling frequency is larger than twice the maximum frequency of the signal being sampled.

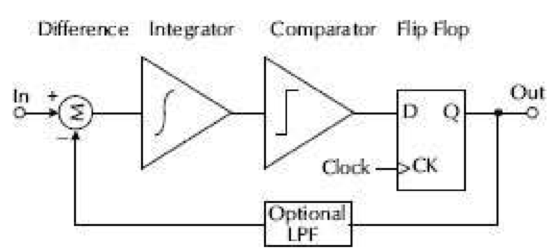


Figure 24-62. Simple analog sigma-delta modulator.

The ΣΔ modulator comprises a summer, an integrator, a comparator and a D-type flip flop. The output of the flip-flop is the bit stream. The average value of the switched output signal represents the analog output. The output signal is fed back to the summer, where the input signal is compared to the average value of the feedback signal. If the input signal is larger than the average value of the feedback signal, the integrator output will move in a positive direction, eventually crossing the threshold of the comparator and causing its output to go high.

On the next positive clock edge the D-type flip flop will be clocked high and the output will go high for this bit period. On subsequent positive clock edges, the output will continue to be clocked high until the fed back output catches up with the more positive input and actually passes it. At that point, the integrator output will fall a sufficient amount to go below the comparator threshold and cause the flip flop to be clocked low on the next positive clock edge. This whole process repeats indefinitely, with the ΣΔ loop always seeking to keep the difference between the input and the reconstructed output small. If the input signal is at zero, the output of the modulator will toggle between high and low levels at the clock rate, producing a square wave at half the clock frequency.

Because the average value of the output bit stream is driven to equal the input signal by feedback, the pulse density in the bit stream will be a faithful representation of the input signal amplitude. In a conventional ΣΔ class D amplifier, the bit stream that is fed back is the bit stream that exists at the output of the amplifier prior to the filter.

It is important to recognize that the feedback signal to the analog summer in Fig. 24-62 is not reconstructed—it is either a positive or negative reference voltage in discrete time. Reconstruction of the feedback signal into a continuous time representation takes place in the integrator. In most class D amplifiers the switched signal fed back is not a fixed reference signal (one-bit A/D), but rather an attenuated version of the actual switched output signal. In this way the pulses fed back are a more faithful representation of the areas of the actual output pulses of the amplifier. If the power supply rail increases, for example, that will be appropriately reflected in the area of the pulse that is fed back. This improves PSRR. In some designs the feedback signal is low-pass filtered before application to the summer.

Notice that when a pulse occurs, it will never have a width less than one clock period. This is in contrast to the output of a PWM modulator, where very small pulse slivers can be created when the signal is near its maximum magnitude. Conversely, there is no limit to the width of a pulse in a sigma-delta bit stream. Moreover, there need be no periodic regularity to the time when pulses go from positive to negative or vice versa. There is no carrier frequency in the output signal.

Oversampling and Noise Shaping

Because sigma-delta modulators create quantization noise in the time domain, it is important that they run at significantly higher clock frequencies than PWM modulators. This allows for finer granularity of the pulse time intervals. The high rate at which the ΣΔ modulator is clocked is referred to as oversampling. The ratio of the ΣΔ clock rate to the required Nyquist sampling rate is referred to as the oversampling ratio, or OSR. The Nyquist sampling rate for a 20kHz analog signal is 40kHz. A ΣΔ modulator operating at 4MHz will have an OSR of 100. Oversampling reduces in-band noise by spreading the total sampling noise power over a much larger range of frequencies.[31] An OSR of two will spread the noise over twice the frequency spectrum, cutting the in-band noise power in half and improving *SNR* by 3dB.

Increasing the clock rate of a first-order ΣΔ modulator reduces noise by 9dB per octave of OSR. This larger decrease than 3 dB per octave results from a process called noise shaping.[31] Increasing the clock rate from 5MHz to 20MHz thus reduces the quantization noise by 18dB. Employing a higher-frequency clock also allows the use of a physically smaller output filter with a higher 3-dB frequency.

There is a limit to the smallest pulse that can be handled by a class D amplifier output stage, especially when necessary dead time margins are considered. This was discussed earlier in connection with sliver pulses produced by PWM modulators at very high or very low duty cycles. The switching rate of the class D output stage also has an upper limit (apart from minimum pulse width) because its power loss increases with the average switching rate. Each time the output state is switched, power is dissipated. The product of the clock frequency and the transition density of the bit stream is what influences heat generation from switching losses.

24.9.9 Digital Modulators

In some ways the power amplifier is one of the last things to go digital in many areas of consumer electronics. For this reason, and given the tremendous processing capability made available at low cost by VLSI chips, it is only natural for class D modulators to be implemented in all-digital form.

Class D amplifiers using digital modulators are a natural fit to digital audio sources where the signal often originates in PCM format from an I^{2}S bus. The I^{2}S format (Inter-IC Sound) is a popular standard developed by Philips Semiconductor for transport of digital audio signals. Use of a digital modulator eliminates entirely the analog interface and potentially some other mixed-signal functions. This can be especially attractive in Home Theater receiver applications where most of the audio signals are handled in digital fashion.

24.9.10 Hybrid Class D Amplifiers

In some cases higher sonic performance can be achieved by combining class D amplifiers with analog power amplification. A simple example of this is to amplify the signal with both class D and class AB amplifiers to the same level. The class AB amplifier is a low-power, high-current amplifier that actually drives the load. Its output stage power supply is driven by the output signal of the class D amplifier.[1] This is somewhat analogous to a linear power amplifier wherein a floating class A amplifier is driven by a class AB amplifier.

Fig. 24-63 is a conceptual illustration of a hybrid class D amplifier. The most straightforward approach is to have the class D amplifier drive the entire power supply of the class AB amplifier. That power supply can be either a linear supply or a switcher.

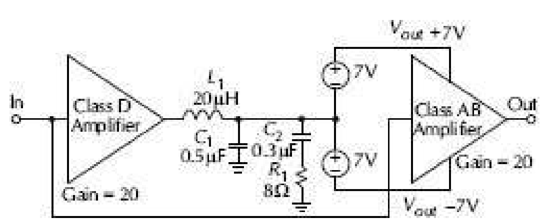


Figure 24-63. Hybrid class D amplifier.

The hybrid class D amplifier has several advantages. It isolates the class D output from the load, greatly reducing EMI and taking the output filter out of the signal path. It also preserves the damping factor that would be attained by a linear amplifier. Finally, it allows the negative feedback to be closed from the output terminals of the amplifier without suffering the consequences of phase shift introduced by the output filter.

In practice, only the output stage of the class AB amplifier needs to be run from the flying rails provided by the class D amplifier. All of the earlier stages can be run from a very clean linear supply because they require very little power. Then, decent PSRR of the class AB output stage is all that is needed.

The hybrid class D amplifier is an intelligent tradeoff, providing improved sound quality in exchange for a reduction in efficiency. The class AB amplifier portion can be run at low voltage, but it must still be designed to be able to deliver the full current produced by the amplifier. The use of small local rail voltages in the class AB amplifier section greatly eases safe area requirements for the output transistors.

Fig. 24-64 shows estimated power dissipation as a function of power for a conventional class AB amplifier, a hybrid class D amplifier and a class D amplifier, all rated at 200W/8Ω. The class AB amplifier within the hybrid class D amplifier is assumed to have ±15V floating rails.

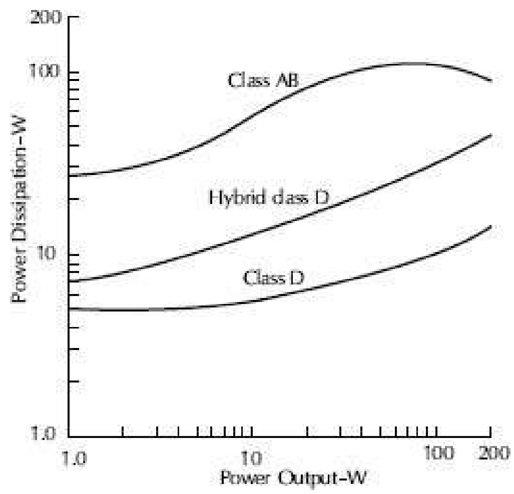


Figure 24-64. Power dissipation of Class AB, Hybrid class D and class D amplifiers.

24.9.11 Measurement of Class D Amplifiers

The proper measurement of class D amplifier performance can present some challenges in two areas in particular. The output filter limits the bandwidth of many class D amplifiers to the point where harmonics of a 20 kHz THD fundamental will be significantly attenuated, leading to erroneously optimistic THD-20 numbers. For this reason, high-frequency intermodulation distortion tests, like the 19+20kHz CCIF IM test should be used, since the distortion products in this test are all in-band.[1]

The potentially large out-of-band EMI voltages at the output of class D amplifiers can upset the input circuitry of sensitive measuring equipment, possibly causing high-frequency overload in the front-ends. For this reason some passive low-pass filtering is required ahead of the test equipment. It is easy to see that such filters will further invalidate high-frequency THD measurements.

The AES17 Filter

To deal with the spurious EMI that may be present at the output of class D amplifiers, the Audio Engineering Society published a filtering recommendation called AES17.[32] The AES17 low-pass filter is placed between the amplifier output and measurement instruments like distortion analyzers. The filter is very sharp, flat to 20 kHz and then down by 60dB at 24 kHz. This usually requires a 7th order elliptic filter, most or all of which should be implemented with passive components. Without the filter, measurements at low signal levels may be especially affected in an adverse way.

The AES17 filter is not unlike an anti-alias filter. In many cases the *brick-wall* nature of the AES-17 filter is overkill. Where test equipment with robust front-ends is used, a much gentler filter function may be adequate. This will be especially the case for class D amplifiers whose EMI spectra lie higher above the audio band. Even a 3rd order (or higher) Bessel filter may be sufficient in some cases, where the real pole is implemented passively. A side benefit of this is better preservation of waveform fidelity for time domain measurements like square wave response.

24.10 Switching Power Supplies

Conventional power supplies are heavy due to the low line frequency of 50 or 60Hz. It makes the power transformer heavy and requires large reservoir capacitors because of the low ripple frequency. Conventional supplies have poor power factor due to their highly non-linear impulsive input current. These rectifier pulses of many tens of Amperes cause voltage drops in the resistive mains supply that reduce the amount of power that can effectively be extracted from the mains.

Switch Mode Power Supplies (SMPS) are lighter and more efficient than conventional power supplies. They can also reduce the presence of 60 Hz or 120Hz (and related harmonics) electrical and magnetic fields within the amplifier. Because the operating frequency is much higher than that of the mains, power supply ripple on the amplifier power rails is typically much smaller for a given amount of reservoir capacitance.

Switching power supplies are beginning to make their way into more power amplifiers. They offer high efficiency while being compact and lightweight. Their use is becoming widespread in pro audio amplifiers and in Home Theater receivers where these qualities are especially important. They are also moving into some audiophile power amplifiers. Their advantages include regulated outputs at no penalty in power dissipation.

Switching supplies operate by rectifying the mains voltage on the line side and storing the dc on a reservoir capacitor.[33,34] Fig. 24-65 shows a simplified arrangement. The intermediate dc bus voltage is then switched at a high frequency (several hundred kHz) to become ac to drive an isolating transformer. The secondary voltage is then rectified and stored on secondary reservoir capacitors. High-frequency transformers are smaller, lighter and less expensive than transformers that operate at the mains frequency.

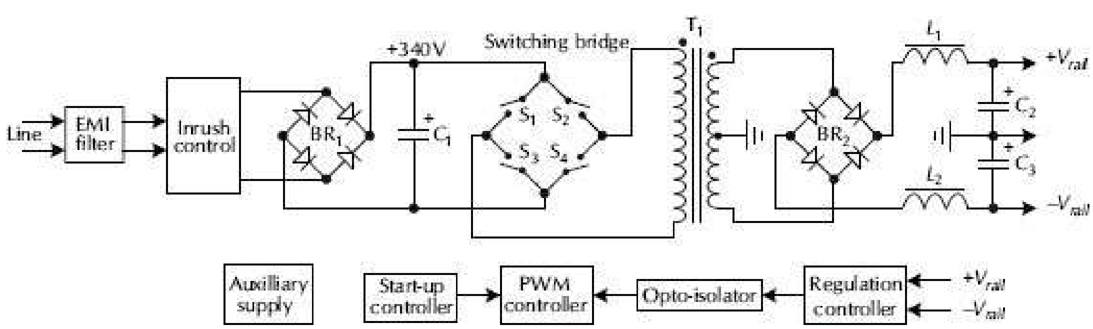


Figure 24-65. Switch mode power supply.

As with a conventional linear power supply, multiple secondary windings can be included on the transformer to create different secondary voltages, each with its own rectifiers and filter capacitors. Alternatively, a single rectified voltage can be created at the secondary side and if lower voltages are required they can be created by non-isolated SMPS converters.

24.10.1 EMI Filter

Switching power supplies create a significant amount of high frequency electrical noise that can result in electromagnetic interference (EMI). This must be prevented from entering the line where it can be radiated and interfere with other electronic equipment.

The SMPS switching frequencies fall in the range of 50kHz to 500kHz with the switching pulses being pulse width modulated (PWM). As a result, harmonics are generated in the radio-frequency range. This property necessitates the incorporation of a sophisticated electromagnetic interference (EMI) filter to prevent the appearance of the switching frequency and its harmonics as common-mode signals on the ac supply lines. A typical EMI filter circuit is shown in Fig. 24-66.

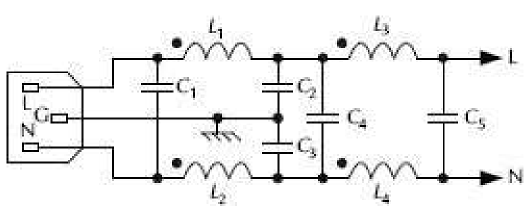


Figure 24-66. EMI filter.

The EMI filter must be designed to prevent common-mode signals that are generated in the power supply from being conducted to the external supply mains while at the same time offering minimum series impedance to the differential 60 Hz ac voltage of the supply mains. This is accomplished by having balanced inductors in each of the supply lines with positive mutual inductance between the inductors in upper and lower lines. This arrangement maximizes the inductance for common-mode currents that flow in the same direction in both conductors. For the oppositely directed differential currents of the 60Hz main supply, the mutual inductance is negative, thus forcing the overall series inductance to a small value.

24.10.2 Line dc Supply

The line supply part of an SMPS converts the mains voltage into a dc voltage of about 340 V that can be used by the subsequent isolated DC-DC converter. It is easy to see that for a 240 V input, the line supply shown in Fig. 24-67 is merely a full wave bridge rectifier feeding a filter capacitor consisting of C_1 and C_2 in series. These each may be on the order of hundreds of µK Notice that SW1 is open for 240 V input operation. For 120 V operation, SW1 is closed, converting the arrangement into a voltage doubler.[33] Under these conditions, D_3 and D_4 are reverse-biased and not used. Inrush control is provided by an NTC resistor that is shunted by relay contacts that are closed shortly after power is applied.

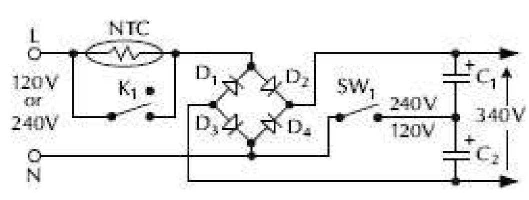


Figure 24-67. Line supply.

24.10.3 Isolated DC-DC Converter

So-called *offline* switching power supplies operate off the mains voltage and almost always provide isolation. Isolated switching converters are quite simple in concept. A high-frequency transformer is used instead of the line-frequency transformer in a conventional supply. The mains voltage is first rectified and filtered to produce a dc voltage. That voltage is then chopped into a square wave to provide the ac signal necessary for feeding the high frequency transformer. The conversion of the dc voltage to the ac square wave voltage is usually accomplished by MOSFET switches. The secondary of the transformer then feeds a rectifier and reservoir capacitor.

Fig. 24-68 shows a more complete off-line DC-DC converter that includes voltage regulation that is implemented via pulse width modulation (PWM). As with class D power amplifiers, smaller *on* times of voltage pulses applied to the transformer supply less energy and thus result in a lower output voltage. The rectified output voltage at the secondary is usually compared to a reference voltage to create an error signal that controls the PWM circuit. In many cases the error signal is conveyed to the PWM controller on the line side by an opto-coupler to provide the necessary isolation. Because the opto-coupler is only conveying error information in a closed-loop feedback arrangement, variation in its coupling efficiency matters little. If multiple secondary windings are used to implement different voltage supplies, all but the one connected in the error loop will have reduced load regulation unless they include their own regulators.

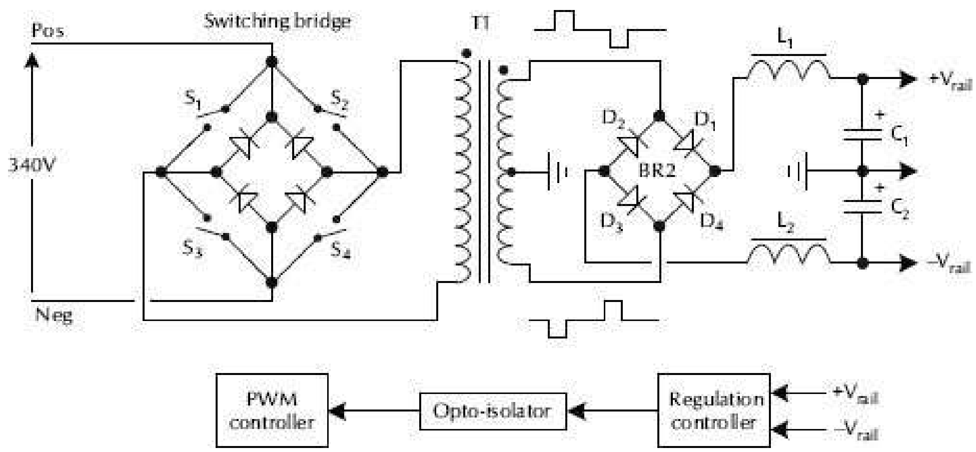


Figure 24-68. Isolated DC-DC converter.

Since the ripple frequency is much higher than in a conventional supply, the storage capacitors can be much smaller for a given amount of ripple voltage. However, fairly large reservoir capacitors are still necessary if the very large brief peak currents demanded by signal transients in the power amplifier are to be supplied. The SMPS would not normally be able to supply such large brief load currents by itself unless it was grossly over-designed.

24.10.4 Buck Converters

The buck converter takes a dc input voltage and steps it down to a lower dc voltage. Understanding it is central to understanding most of the concepts underlying SMPS design. The buck converter uses a transistor switch, a diode, an inductor and a filter capacitor, as shown in Fig. 24-69. It is basically a PWM arrangement that relies on the fact that an inductor tries to keep current flowing as its magnetic field collapses. It operates with high efficiency because its switching device is only on or off. The relative *on* time of the switch controls the amount of energy that is delivered to the load and thus the output voltage.

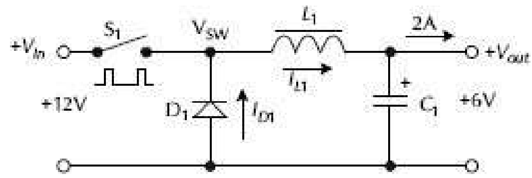


Figure 24-69. Buck converter.

The inductor is placed between the source and the load. If V_{in} – V_{out} is Δ V then the current flowing in L_1 when the switch is closed will rise with a slope of $\Delta V/L_1$, reaching a maximum at the end of the time that the switch is closed. When the switch opens, the voltage at the input side of the inductor flies negative and flows through the catch diode, sometimes called the *freewheeling diode*. Current continues to flow into the load as the magnetic field in the inductor collapses, decreasing with time until the switch turns back on. In the steady state, current is usually flowing continuously in the inductor during both the closed and open states of the switch. In other words, S_1 turns back on before the magnetic field fully collapses. A triangular ripple current thus flows in the inductor. When the current in the inductor does not go to zero at any point in the cycle, the converter is said to be operating in the *continuous* mode of operation and the output voltage is a linear function of the duty cycle.[33,34]

However, if the load current is smaller than half the peak-to-peak value of the ripple current, the converter will enter the so-called *discontinuous* region, where inductor current stops before the end of the *on* time of the switch. This is a less desirable mode of operation. Buck converters operate best in the continuous mode, and this is why buck-type converters are best operated with a minimum load current. If the inductance is too small, or if the frequency is too low, the current in the inductor will fall to zero before the end of the second half cycle where energy is being transferred to the load. In this case, it can be shown that the relationship of output voltage to duty cycle is no longer linear.[34] The time slope of the current in the inductor is determined by the difference in the input and output voltage and the value of the inductance. A larger inductance will result in a shallower slope and less likelihood that current in the inductor will fall to zero.

Operation of the buck converter is further explained by the waveforms in Fig. 24-70 where the converter is delivering 2A to the load. During the first half cycle when S_1 is closed, the input voltage is applied to the inductor and the current will rise linearly with time in accordance with the voltage across the inductor and the inductance. It will reach a value I_{max}.

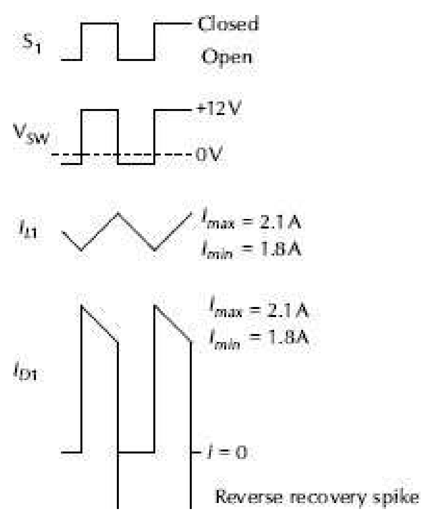


Figure 24-70. Buck converter waveforms.

When S_1 opens, the current in L_1 will continue to flow due to its collapsing magnetic field. This current is usually referred to as the *flyback current* or the *commutation current*. During the second half cycle the only place the current can flow from is through D_1 from ground. The switched output voltage V_{sw} thus snaps from +12V down to about −0.7 V, the forward drop of D_1. It is important to recognize that the current continues to flow through the inductor during this second half cycle. During this time the current in the inductor will fall linearly to I_{min}. Switching frequencies and inductor values are usually chosen so that I_{min} is greater than zero at the end of the second half cycle. The average of I_{max} and I_{min} is the output current into the load. The difference between I_{max} and I_{min} is called the ripple current.

Diode D_1 is often called a *freewheeling, flyback* or *commutating* diode. As shown in Fig. 24-70, D_1 is in a conducting state at the end of the second half cycle when S_1 again closes and raises the switched output from −0.7 V to +12V. Diodes in a conducting state cannot instantly stop conducting when the current is reversed. Silicon diodes that have been conducting contain stored charge in the form of *minority carriers* that must be swept out before the diode can allow reverse voltage across its terminals. This process is called *reverse recovery*. As a result, a brief large current spike will occur when S_1 closes. This is undesirable and represents lost energy and a source of EMI. Fast diodes with small reverse recovery times are thus desirable. Shottky diodes do not involve minority carriers and are largely free from the reverse recovery effect. They are also preferred because of their smaller forward voltage drop.

24.10.5 Synchronous Buck Converter

Fig. 24-71 illustrates a synchronous buck converter. Here D_1 is replaced with a switch S_2 implemented with a MOSFET. This arrangement is more efficient because there is no junction drop when the inductor current is flowing from ground on the second half cycle. That voltage drop represents lost power. S_2 is sometimes referred to as a *synchronous rectifier*. It is a switch that is turned on when it is supposed to be conducting like a rectifier. In this arrangement, S_2 is off when S_1 is on and vice versa. Operation is largely identical to that of Fig. 24-70 with the exception that the switched output voltage falls almost to zero during the second half cycle instead of to −0.7 V. Notice that the gates are driven by pulse transformers. This makes it easy to drive floating switches like S_1.

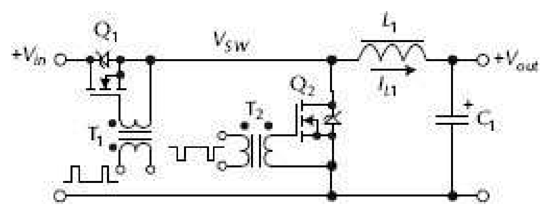


Figure 24-71. Synchronous buck converter.

There is one concern with the synchronous buck converter. If both switches are on even briefly the input power supply will be shorted to ground and a very large *shoot-through* current will flow. This must be avoided by adding *dead time* to the operation of the switches where both switches are off for a brief time. If S_1 opens and S_2 has not yet closed, a negative flyback voltage will be created by the collapsing magnetic field of L_1. The flyback voltage is clamped for this brief interval by adding a freewheeling diode across S_2. The switched output node will then be prevented from going more negative than −0.7 V. This is why the negative portion of the output voltage exhibits negative "ears" at the beginning and end of the second half cycle when both switches are open.

Because the dead time is usually kept very small compared to the period, most of the improved efficiency contributed by S_2 is preserved. Importantly, the R_{DSON} of the MOSFET used for S_2 must be small enough so that the voltage drop across it when it is *on* is considerably smaller than the *on* voltage of the diode. All MOSFETs designed for switching applications have the diode built into them in the form of the source-drain silicon body diode.

Unfortunately, this arrangement does not prevent the reverse recovery current spike when S_1 closes. This is because the freewheeling diode becomes conducting during the dead time just before the end of the second half cycle. The reverse recovery shoot-through current can be reduced if a Shottky diode is connected in parallel with the MOSFET. It will prevent the slower silicon diode from ever turning on.

24.10.6 Forward Converters

The isolated converter of Fig. 24-68 is called a *forward* converter. It is very much like a buck converter where a transformer has been interposed in the forward path. When the dotted ends of the windings are made positive by the PWM pulse, D_1 sources current into L_1, just as did S_1 in the buck converter of Fig. 24-69. At the end of the *on* time, when the voltage across the windings is zero, flyback current from L_1 flows through D_4 and maintains current flow into the load, just as in a buck converter. During the same PWM interval, D_3 and D_2 are servicing the negative rail load via L_2.

There is a second *on* interval during the second half cycle of the ac waveform when the dotted ends of the windings are made negative by the opposite-polarity PWM pulse. During this time, the roles of D_1 and D_4 are reversed, as are those of D_3 and D_2.

L_1 and L_2 are central to the buck-like operation of the forward converter. They are not just there for extra filtering.

24.10.7 Boost Converters

A buck converter can only create an output with a lower voltage than the dc source. A boost converter can create an output voltage greater than that of the dc source. A simple boost converter is shown in Fig. 24-72. The dc source is connected to one end of inductor L_1 and the other end of L_1 is connected to ground through switch S_1. L_1 is also connected to diode D_1, which charges load filter capacitor C_1 during the second half-cycle of operation. S_1 is turned on during the first half-cycle of operation, causing current to flow from the source through L_1 to ground. The current rises with time until S_1 is turned off. During this time energy is stored in the inductor.

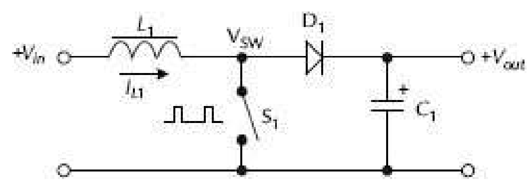


Figure 24-72. Boost converter.

When S_1 is opened during the second half cycle, the inductor seeks to keep the current flowing by creating a flyback voltage that forward biases D_1. This allows the energy to be dumped into the load. The current in the inductor will then decrease with time as the energy is transferred to the load. In many designs the inductor current will reach zero before the beginning of the next cycle when S_1 is once again turned on. This is referred to as *discontinuous* operation because there is a time interval when the inductor current is zero.

Under different operating conditions, such as with a larger value of inductance, current in L_1 will not decrease to zero by the end of the cycle. In this case, current is flowing all of the time during both half cycles. This is referred to as the *continuous* mode of operation and has somewhat different behavior as described in reference 34.

The amount of energy transferred from the source to the load depends on the duty cycle of the *on* time of S_1. A PWM controller is used to control the *on* time and thus regulate the output voltage against input voltage and load current changes. During the second half cycle the input source is also supplying the current that is flowing through the inductor into the load. This means that the energy flowing into the load is not just from the energy stored in the inductor, but also that flowing from the source into the inductor as well.

24.10.8 Flyback Converters

If a second inductor L_2 is wound on the same core as L_1 in the boost converter, D_1 and C_1 can be fed from L_2. This is a *flyback* converter. The energy stored in the core by L_1 during the first half cycle is given up to the load by means of the L_2 winding when S_1 is opened. This is much like the operation of a boost converter, but with isolation. It is tempting to refer to the combination of L_1 and L_2 on the same core as a transformer, but it is more technically correct to view them as two inductors sharing the same core.[34] In transformer operation, current is usually flowing in both the primary and secondary at the same time. This is what happens in the *forward* converter. In the dual-inductor flyback operation, current flows only in one inductor at a time. The flyback voltages appearing on L_1 and L_2 are still related by the turns ratio of L_1 and L_2. The Kettering ignition in automobiles is basically a flyback converter.

24.10.9 Power Factor Correction

Power factor describes the degree to which power is transferred for a given combination of delivered voltage and amperage. A resistive load has 100% of the apparent power (VA) delivered to it as real power, and therefore has the ideal power factor of unity. The power factor is unity because the voltage and current wave-shapes are identical and in phase. Any departure from the wave-shapes being identical and in phase results in a power factor of less than unity. For example, in the extreme case of a purely inductive load, voltage and current are 90° out of phase and no power is delivered to the load, in spite of the fact that significant voltage is present and significant current flows (VA). Nonlinear loads, such as rectifiers with capacitor input filters, also represent a load with a poor power factor. Such inefficiency of power transfer is bad for the utility and sometimes bad for the user.

For linear loads, like inductive loads encountered in utility power distribution, power factor correction (PFC) is achieved passively with inductors and capacitors that shift the phase of the load current to be nearly the same as that of the voltage. Such passive power factor correction does not work for nonlinear loads. The highly distorted, high amplitude input current pulses of the front-end rectifier supply of a conventional switching power supply cause a great departure from an ideal power factor of unity. Once again, the definition of unity power factor is when the load acts like a pure resistance.

24.10.9.1 Active Power Factor Correction

The objective of active power factor correction is to alter the input current waveform to be one that has the same waveshape and is in-phase with the mains voltage.[33,34,35] Note that the mains voltage waveform is not always truly sinusoidal, sometimes due to the presence of other nonlinear loads. For example, other ordinary rectifier loads may tend to flatten one or both tops of the mains voltage sinusoid. The input current should follow the shape of these waveforms as well, just as a resistor's current would.

Active PFC circuits use a bridge rectifier on the mains, but no filter capacitor, as shown in Fig. 24-73. Instead the full-wave-rectified voltage, called a *haversine*, feeds a dynamic boost converter. The boost converter creates a high-voltage dc intermediate bus supply of about 385Vdc that feeds the subsequent DC-DC converter. The filter capacitor, C_1, is moved to after the boost converter. The boost converter can operate over a wide input voltage range while delivering the 385Vdc output. This allows the power supply to operate over the full universal input voltage range of 88–264Vac without a mains voltage switch.

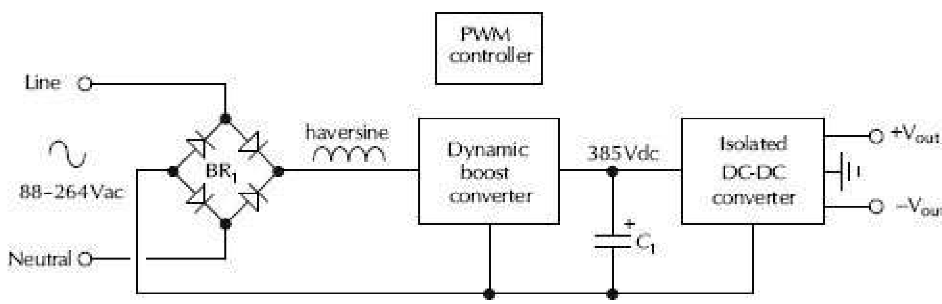


Figure 24-73. Active power factor correction.

The MOSFET switch in the boost converter is controlled in such a way that the average input current has the same waveform and phase as the input voltage. The average input current is adjusted by the controller to regulate the resulting boost voltage. This is accomplished with an IC specifically designed for the PFC function.[36] The controller IC uses feedback techniques to match the current waveform to the mains voltage waveform. The resulting boosted intermediate bus voltage must always be larger than the rectified mains voltage, since a boost converter cannot output a voltage lower than its input voltage. Inrush control can be easily implemented with the IC controller. The regulated intermediate bus voltage permits the DC-DC converter to operate under its optimum conditions, even under substantial line voltage variations.

24.10.9.2 The Dynamic Boost Converter

The boost converter can take virtually any input voltage and boost it to a higher voltage, where the energy is then dumped into a storage capacitor. For the PFC application, it must dynamically adjust its boost ratio as the instantaneous voltage of the haversine waveform goes from its valley to its peak. The switching frequency is much higher than the mains frequency, so in a relative sense the line voltage waveform at the input to the PFC is changing very slowly, and the switching converter has plenty of time to adapt its operating parameters to the “slowly-changing” haversine input voltage waveform.

Near the bottom of the input voltage waveform, the boost ratio of the PFC boost converter must be large to achieve the higher regulated output voltage from the small input voltage of the haversine. At the same time, the input current should be small. We note that the power consumed by a real resistance from its source is proportional to the square of the voltage, so when the haversine input waveform is at 10% of its peak value, for example, the instantaneous power being drawn from the full-wave-rectified source by the PFC converter will be only 1% of the maximum it will draw at the peak of the waveform. Fig. 24-74 illustrates the boost converter input current with ripple as compared to the desired mean haversine current waveform. The boost converter should operate in the continuous conduction mode (CCM), otherwise substantial amounts of high-frequency ripple current will be introduced into the mains supply.

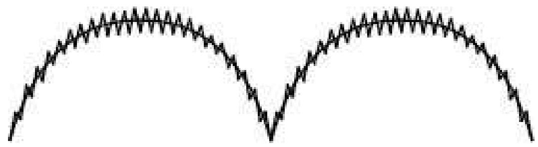


Figure 24-74. Power factor correction waveform.

24.10.9.3 The PFC System

The PFC control circuit must monitor both the PFC output voltage and its input current so as to provide a regulated output dc voltage while drawing input current that is proportional to the input haversine voltage waveform. To achieve the latter, it must also monitor its full-wave-rectified input voltage.

Fig. 24-75 shows a very simplified PFC circuit, including the mains bridge rectifier, the boost converter and the PFC controller chip. R_3 senses the input current and allows the PFC controller to match the input current waveform to the input voltage haversine waveform reference presented at the V_{in} pin. R_4 and R_5 provide a scaled version of the output voltage for output regulation by feedback. Note importantly that the controller chip has two feedback jobs to do, one for input current and one for output voltage. The controller chip is powered from a small auxiliary supply.

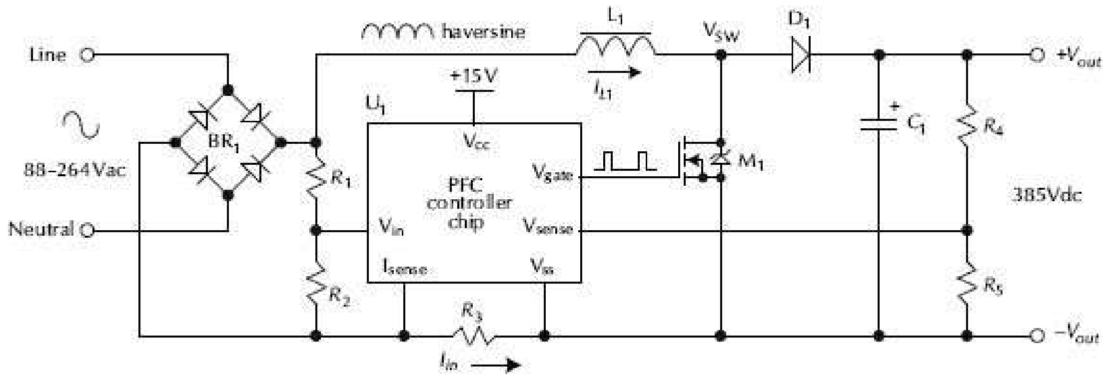


Figure 24-75. Power factor controller.

The PFC circuit is easily designed to provide inrush control, since the amount of current it draws from the mains is determined by its switching duty cycle, which can be brought up slowly. The same feedback loop that regulates the bus voltage can be made to increase that bus voltage gradually.

Bear in mind that the subsequent DC-DC converter will draw more input current when its input voltage is low, in order that it can deliver the same needed output power. This represents a negative resistance effect; as the input voltage falls, the current flow increases. Therefore, when the intermediate bus voltage is being brought up gradually, the switching converter's turn-on behavior must be controlled in some way—for example, the DC-DC converter should not be enabled until the bus voltage rises to its nominal regulated value.

Power factor correction is not employed merely to allow the amplifier to act as a "good neighbor" or to satisfy utility standards. It also enables the amplifier to extract the maximum amount of energy out of a mains supply with finite impedance, since power is being extracted over the full mains voltage cycle, not just in high-current pulses of short duration.

24.11 Professional Power Amplifiers

Professional power amplifiers are used to drive the loudspeakers in professional sound reinforcement systems. In professional audio amplifiers, it's all about power, reliability, size, weight, convenience of set-up and reduction of system wiring.

In the 1970s they were not unlike higher-powered consumer audio amplifiers that were made more rugged. They also often included balanced inputs, power monitoring displays and clipping indicators. They were often operated in bridged mode to achieve higher power with given supply voltages. The march to higher power levels in the 1980s saw the introduction of class G and class H amplifiers. These designs reduced heat generation and allowed more power to be made available in a given amount of precious rack space. The Crest 8001 was a good example of such high-power amplifiers of the time. It provided over 2000W into an 8Ω load in bridged mode in a 3 rack-space design.

In more recent years there has been a strong migration to the use of class D amplifiers in these applications. The need for higher power, less heat generation, weight, and rack space for a given power has driven this trend. These newer amplifiers usually employ switching power supplies (SMPS) to greatly reduce the size and weight. Many newer amplifiers incorporate power factor correction (PFC) to make the most efficient use of the mains power and obey governmental directives. Modern SMPs also enable great agility in powering for many different mains standards, often over a range of 88 to 264Vac, often without the need for a voltage switch. By the same token, such amplifiers can deliver their full power when the mains voltage sags and they do not need to be over-designed to handle worst-case high mains voltages.

Balanced and single-ended inputs have always been a must for professional power amplifiers, but many have included AES3 (AES/EBU) digital audio transport standard inputs for quite some time. These digital inputs are available with electrical and optical interfaces. Beginning around 1996, digital audio inputs were also made available over proprietary Ethernet interconnect.

24.11.1 Microcomputers

The introduction of microprocessors and DSP changed the landscape. Think of the modern consumer AV receiver and all of the control processing and DSP in it. Most now use digital innards and class D power amplifiers. Many pro amplifiers now also employ such technology. Microprocessors began to be incorporated in professional power amplifiers in the 1980s.

Microcomputers in many amplifiers have also permitted the implementation of advanced protection and monitoring schemes. The use of FLASH memory now allows amplifiers to retain settings and implement different software modes. In many cases the microcomputer implements power-on self-test (POST) and diagnostic aids. Some amplifiers can implement a USB port for local access of diagnostics and even for software upgrades in the field. The modern professional power amplifier is truly a system with many advanced and intelligent features, as illustrated by the block diagram in Fig. 24-76.

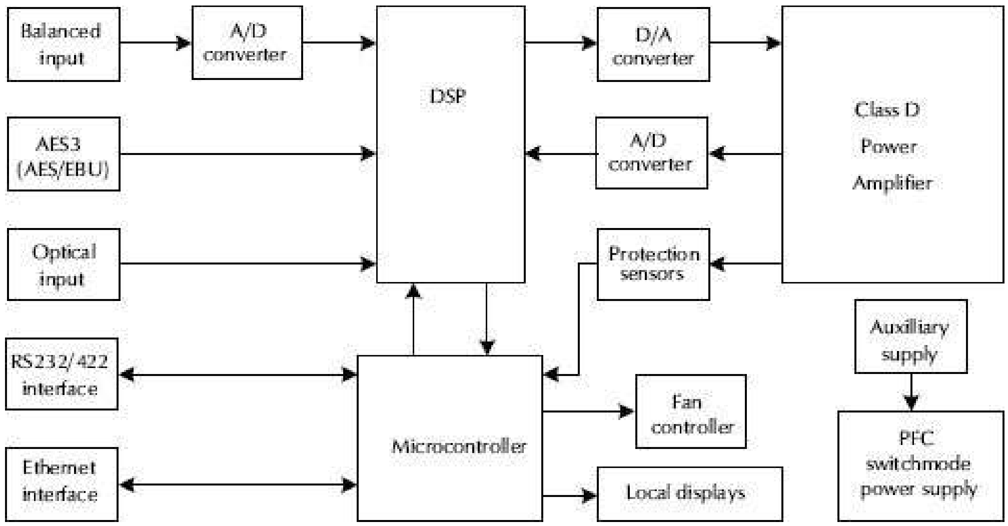


Figure 24-76. Professional amplifier block diagram.

24.11.2 Networked Control and Monitoring

Remote control and monitoring has long been important with professional audio amplifiers, especially in large venues. In the 1980s Crown introduced the IQ system, where RS232/422 serial interfaces made it possible to connect many amplifiers in a daisy chain that was connected to a host computer for control and monitoring of many amplifiers.

In 1993 a major step forward was made by Peavy with the introduction of the MediaMatrix® networking system based on Ethernet connectivity. This was far superior to interconnecting power amplifiers with serial ports.[37]

24.11.3 Digital Signal Processing

The availability of economical digital signal processing (DSP), A/D conversion and D/A conversion in the early 1990s made it inevitable that DSP would be introduced into professional power amplifiers and play a major role in reducing system costs while greatly increasing system flexibility and capability. In many cases, DSP has allowed the power amplifier to swallow many functions that were formerly performed by external boxes. On of the most important areas in which DSP allows system integration in one place is the group of functions called loudspeaker management.[37,38,39] These functions include:

- Filtering and equalization.
- Delay for time alignment.
- Compression and limiting.
- Electronic crossovers.
- Protection.
- Load sensing.

In conjunction with the microcomputer system, different DSP functionality, processing programs and settings can be conveniently selected. DSP also makes it practical to include some system testing features in the amplifier, such as the generation of sine wave tones, pink noise, white noise, MLS, etc. DSP also makes it possible to implement more sophisticated protection and monitoring systems for the amplifier and connected loudspeaker. It is important that the DSP in power amplifiers introduce only a very small amount of latency, preferably less than 2 μs.

24.11.4 Audio Networking

The combination of digital audio and Ethernet has made audio networking possible. For power amplifiers, this means the digital delivery of the audio signal and control over the Ethernet infrastructure in an integrated way. This approach began in earnest in 1996 when CobraNet® was introduced by Peak Audio.[40,41] This

Audio over Ethernet system (AoE) has enjoyed considerable success, but does not allow one to use much off-shelf Ethernet hardware and software. It also suffers from latency on the order of as much as 5 ms.[42,43] While such latency is acceptable for many pro sound applications it is a significant concern for live sound applications. Live sound is the most stringent networked audio application because of its need for very low latency.

There have been numerous other system approaches to networked audio that were based on AoE or Audio over Internet Protocol (AoIP), such as RAVENNA[44] and Livewire.[45] Another noteworthy one is Audinate's Dante® system.[42,43] This system has largely solved the latency problem while enabling the full use of Ethernet off-shelf hardware and routing capabilities.

More recently the Audio Engineering Society has issued inter-operability standards for AoIP systems and technology.[46] The AES67-2013 standard was published in September 2013. The AES67 standard is the result of an industry collaborative effort, and does not seek to obsolete existing AoIP approaches or hardware. Much of the compliance with AES67 can be achieved with software changes.

Of course, AoIP pertains to the networked interconnection of all professional audio equipment, not just power amplifiers. AoIP is a rapidly-moving technology that has already revolutionized the professional sound industry.

References

1. Bob Cordell, *Designing Audio Power Amplifiers*, McGraw-Hill 2010.

2. Giuseppe Massobrio and Paolo Antognetti, *Semiconductor Device Modeling with SPICE* (2nd ed), McGraw-Hill, 1993.

3. Harold Black, U.S. patent 2,102,671, issued 1937.

4. Walt Jung, *Audio IC Op-Amp Applications*, Howard W. Sams & Co. 1987.

5. Walt Jung, *Op Amp Applications Handbook*, Analog Devices Inc., Newness 2005.

6. Burkhard Vogel, *The Sound of Silence*, Springer, 2011.

7. J. W. Haslett, F. N. Trofimenkoff, *Thermal Noise in Field-effect Devices*, PROC. IEE, Vol. 116, No. 11, November 1969.

8. Bob Cordell, Linear Systems LSK489 application note, www.linearsystems.com, February 2013.

9. Bob Cordell, *www.cordellaudio.com*.

10. Douglas Self, *Small Signal Audio Design*, Focal Press, 2009.

11. Peter J. Baxandall, *Audio Gain Controls*, Wireless World, November 1980.

12. TI/National LM4562 datasheet, October 2007.

13. Bob Cordell, *VinylTrak – A full-featured MM/MC phono preamp*, Linear Audio, Vol. 4, September 2010, www.linearaudio.net.

14. THAT Corporation, *InGenious®High CMRR Balanced Input Line Receiver IC*, 1200 datasheet, www.thatcorp.com.

15. THAT Corporation, *OutSmarts® Balanced Line Driver IC*, 1646 datasheet, www.thatcorp.com.

16. THAT Corporation, *Blackmer® Pre-trimmed IC Voltage Controlled Amplifier, 2180 and Dual Pre-trimmed 2162 datasheets*, www.thatcorp.com.

17. THAT Corporation, *VCAs in a Pan Potentiometer Application*, Design Note 120.

18. Douglas Self, Audio Power Amplifier Design Handbook, 5th ed., Focal Press, 2009

19. Ben Duncan, *High Performance Audio Power Amplifiers*, Newnes, 1997.

20. G. Randy Slone, *High Power Audio Amplifier Construction Manual*, McGraw-Hill, 1999.

21. Jun Honda and Jonathan Adams, *Class D Audio Amplifier Basics*, International Rectifier Application Note AN-1071, February 2005.

22. Karsten Neilsen, *A Review and Comparison of Pulse-Width Modulation (PWM) Methods for Analog and Digital Input Switching Power Amplifiers*, paper #4446, 102nd AES Convention, March 1997.

23. Eric Gaalaas, *Class D Audio Amplifiers: What, Why, and How*, Analog Dialog 40-06, June 2006.

24. International Rectifier, *High Power Class D Audio Power Amplifier using IR2011S*, application note IRAUDAMP1, www.irf.com, 2005.

25. Jun Honda and Jonathan Adams, *Class D Audio Amplifier Basics*, International Rectifier Application Note AN-1071, February 2005.

26. Marshall Leach, Jr., *Introduction to Electroacoustics and Audio*

Amplifier Design, Second Edition, W. Kendall/Hunt, 2001.

27. U.S. Patent #7,113,038, “*Power Amplifier*” (Bruno Putzeys), September 26, 2006.

28. Bruno Putzeys, *Simple Self-Oscillating Class D Amplifier with Full Output Filter Control*, Presented at the 118th AES Convention, May 2005.

29. R. Schreier, and G., Temes, *Understanding Delta-Sigma Data Converters*, Wiley-IEEE Press, 2004.

30. U.S. Patent #5,777,512, *Method and Apparatus for Oversampled, Noise-Shaping, Mixed-Signal Processing*, July 7, 1998.

31. M. J. Hawksford, *Oversampling and Noise Shaping for Digital to Analogue Conversion, Institute of Acoustics, Reproduced Sound 3*, pp. 151-175, 1987.

32. Audio Engineering Society, *AES Standard Method for Digital Audio Engineering – Measurement of Digital Audio Equipment*, 1998.

33. A. Pressman, K. Billings, and T. Morey, *Switching Power Supply Design*, Third Edition, McGraw-Hill, 2009.

34. Keith Billings, *Switchmode Power Supply Handbook*, McGraw-Hill, 2010.

35. ON Semiconductor, *Power Factor Correction (PFC) Handbook*, 2011, www.onsemi.com.

36. Fairchild Semiconductor, *ML4826 PFC Controller datasheet*, 2001.

37. Don Davis and Eugene Patronis, Jr., Sound System Engineering, Focal Press, 2006.

38. Crest Audio (Peavey Commercial Audio), *Pro-LITE Series amplifiers*, www.peaveycommercialaudio.com.

39. QSC Audio Products, *CXD Series amplifiers*, www.qscaudio.com.

40. Yamaha System Solutions, An Introduction to Networked Audio, White Paper, www.cobranet.info

41. Crown by Harman, CobraNet Primer, White Paper, www.crownaudio.com.

42. Joao Martins, *Audinate Dante – Making Digital Audio Networking Easy*, Audioxpress, February 2014, www.audioxpress.com.

43. Audinate, *Audio Networks Past, Present and Future (CobraNet and Dante), White Paper*, www.audinate.com.

44. ALC NetworX, *RAVENNA*, White Paper/brochure, Ravenna@alcnetworx.de.

45. Axia Audio, *A Closer Look at Livewire*, informational web page, www.axiaaudio.com/livewire.

46. Audio Engineering Society, *AES Standard for Audio Applications of Networks – High Performance Streaming Audio-over-IP (AoIP) Interoperability*, AES67-2013.

Chapter 25

Preamplifiers and Mixers

by Ray Rayburn and Bill Whitlock

25.1 Microphone Preamplifier Fundamentals

Microphones are transducers that typically have very low output signal levels. A voltage gain of 1000 (60dB) or more may be required to bring the signals up to standard line levels, hence the name *preamplifier*. Amplifiers for such low-level signals are prone to problems unique to high-gain, low-noise electronic circuits. Microphone preamplifiers are available as stand-alone devices or as part of simple mixers or complex recording consoles. In this section, we will limit our discussion to preamplifiers and mixers intended for use with professional microphones that have balanced, low-impedance outputs.

25.1.1 The Microphone as a Signal Source

As discussed in Chapter 20 *Microphones* may vary considerably in output impedance, output level or sensitivity, and self-noise. For professional microphones, impedance has a rated or *nominal* value of 150Ω (U.S. standard) or 200 Ω (European standard). Dynamic microphones, like loudspeakers, have an *actual* impedance that varies with frequency. Note the similarity of the impedance plot of Fig. 25-1 to that of a loudspeaker. A single figure representing the impedance of such devices is usually taken as the first *minimum* that occurs after the first *maximum* as frequency is increased from a low-frequency limit. The first maximum is usually cone or diaphragm resonance. For microphones, this impedance would be measured between the signal the output pins (2 and 3 for an XLR) and is variously referred to as *output, source, internal, signal*, or *differential* impedance. Dynamic microphones and condenser microphones with transformer coupled outputs are broadly classified as *floating balanced* sources. *Floating* refers to the fact that the common-mode impedances, i.e., those from output (pins 2 and 3) to case and shield (pin 1), are very much higher than the signal or differential impedance. Many modern condenser microphones have transformerless output circuits, which while balanced, are not floating. In other words the common mode output impedances are low. As emphasized in Chapter 36 *Grounding and Interfacing*, balanced refers to the matching of these common-mode impedances.

25.1.1.1 Electrical Model of the Microphone

Since the Shure SM57 dynamic microphone is so popular, it will be used as an example. Fig. 25-2 shows its internal schematic, the electrical equivalent circuit of the capsule and transformer, and the combined equivalent circuit. The equivalent circuits do not model the diaphragm resonance at approximately 150Hz. Note the pair of 17pF capacitances to the case. These are parasitic or stray capacitances due to the construction of the microphone and not physical components. They determine the common-mode output impedances at high frequencies that play a role in noise rejection or CMRR when the microphone is connected to a preamplifier and cable. The actual measured output impedance of an SM57 is shown in Fig. 25-1. The Shure data sheet accurately specifies the actual impedance as 310 Ω. The equivalent circuit of Fig. 25-2 models the impedance behavior above 1 kHz as well as the equivalent noise resistance. Condenser microphones generally have both lower output impedances and less variation with frequency than dynamic types.

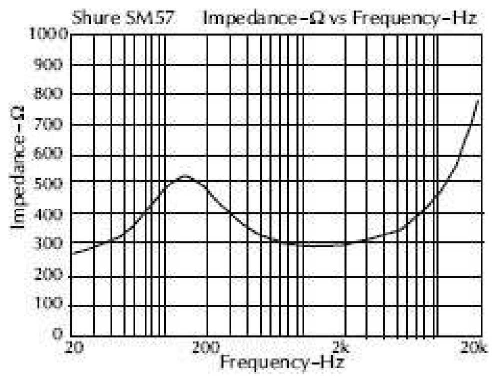


Figure 25-1. Measured output impedance of the Shure SM57.

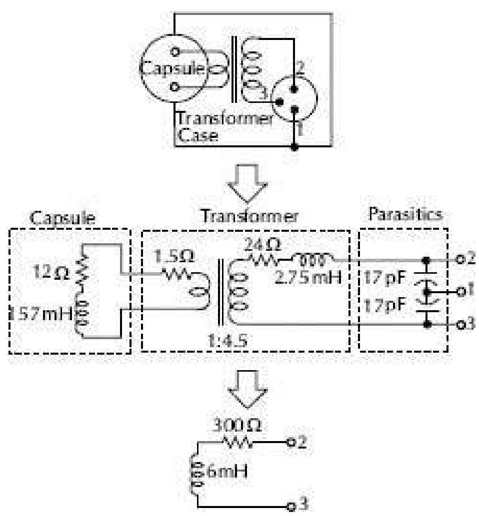


Figure 25-2. Shure SM57 schematic and equivalent circuits.

25.1.1.2 Interactions with Preamplifier and Cable

A microphone preamplifier is normally designed to recover as much of the available microphone output *voltage* as possible. Since the noise floor of the pre amplifier is nearly constant, signal-to-noise performance is improved by making its input voltage as large as possible. It is very important to understand that the fraction of available microphone voltage actually delivered to the preamplifier depends on both the output impedance of the microphone and the input impedance of the preamplifier.

As shown in Fig. 25-3, these two impedances effectively form a voltage divider. The voltage lost in the output impedance Z_S of the microphone depends on the input impedance Z_L of the preamp. *Loading loss*, usually expressed in dB, compares the output voltage with some specified load to the output voltage under open circuit or unloaded conditions. For example, a 150Ω impedance (actual) microphone will deliver 91% of its unloaded voltage when loaded by a preamplifier having a 1.5kΩ input impedance. The loading loss is then 20 × log0.91, which is 0.8dB. Generally, loading loss is negligible (under 1 dB) if load impedance is ten or more times the source impedance. Therefore, as discussed in Chapter 20 *Microphones*, it is neither desirable nor necessary to "match" the impedances of the preamplifier and microphone. If impedances are matched, half the available output voltage from the microphone is lost, degrading signal-to-noise ratio by 6 dB. Although impedance matching transfers maximum power, this is *not* what we want.

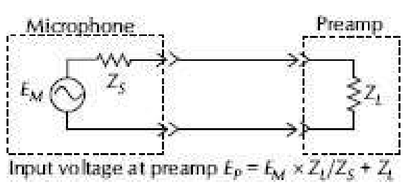


Figure 25-3. Microphone and preamplifier voltage divider.

When a microphone is connected to a cable and a preamplifier, a passive two-pole (12dB/octave) low-pass LC filter is formed as shown in Fig. 25-4. The behavior of LC filters as they approach their cutoff or resonant frequency is controlled by resistive elements in the filter. This resistive damping is largely provided by the input resistance R_L of the preamplifier. Fig. 25-5 shows the deviation in frequency response of a Shure SM57 microphone loaded by the 2.5nF capacitance of 75 ft of typical microphone cable and three different values of preamplifier input resistance. The upper curves, 10kΩ and 3kΩ are typical of preamps that don't use an input transformer. Note the high-frequency response peaking caused by insufficient damping. The lower curve, 1.5kΩ, is typical of a preamp using an input transformer.

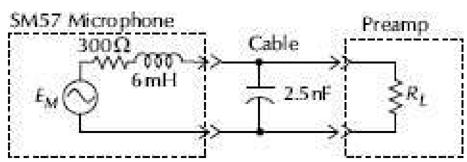


Figure 25-4. Low-pass filter formed by microphone, cable, and preamplifier.

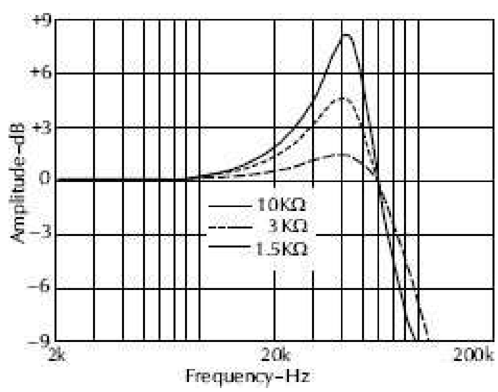


Figure 25-5. Effect of load resistance on damping.

Capacitance of shielded twisted pair cable is usually specified as that from one conductor to the other conductor and the shield. Belden 8451, for example, is listed at 67pF/ft. However, the differential signal is affected by the capacitance between the conductors, which is about half that, or 34pF/ft. With dynamic microphones, high cable capacitance causes high-frequency roll-off. For the SM57 microphone, about a 500ft of this cable (about 17nF) will limit high-frequency bandwidth to about 15kHz. Because condenser microphones use internal amplifiers to drive the output cable, high cable capacitance can cause distortion. If the amplifier has limited output current, it will distort or clip high-level, high-frequency (i.e., high slew rate) signals such as vocal sibilance or a cymbal crash. "Star-quad" cable, although it offers increased freedom from magnetic pickup problems, has about twice the capacitance per foot of standard cable. This fact must be seriously considered for long cables.

Keep in mind, however, that other types (or even models) of microphones may behave quite differently, depending on their exact equivalent circuit. For example, some condenser types have low (around 30 Ω) and almost purely resistive output impedances while some dynamic types can have actual midband impedances over 600 Ω.

To a greater or lesser degree, the frequency response of any microphone will be affected by the load capacitance of the connected cable and preamplifier as well as the input impedance characteristics of the preamplifier. Dynamic microphones, which have significant changes in their output impedance with frequency, will see corresponding changes in their frequency response when driving lower input impedances. Perhaps this is why the selection of microphones and preamplifiers is such a subjective issue.

25.1.2 Some Considerations in Practical Preamplifiers

Because many aspects of preamplifier circuit design and the tradeoffs involved are discussed in Chapter 29 *Consoles and DAWs*, we will discuss only a few topics here.

25.1.2.1 Gain and Headroom

Microphone preamplifiers commonly have maximum voltage gains of about 60dB to 80dB and minimum gains from 0 dB to 12 dB. A typical microphone, such as the Shure SM57, will have an output of 1.9mV or −52dBu with a 94dB SPL acoustic input. For a very high acoustic input of 134dB SPL, its output would be 190mV or −12dBu. But a high-sensitivity microphone such as the Sennheiser MKH-40 will have an output of 25mV or −30dBu at 9dB SPL and 2.5V or +10dBu at 134dB SPL. Such high input levels can actually require the preamplifier to have a loss (i.e., negative gain) to produce usable line level output. Such high input levels can also overload the preamp. Both problems are most commonly avoided with an input attenuator or pad, typically of 20dB. It is crucial that pads not lower the input impedance of the preamplifier when inserted. Microphone pads should never be designed with 150 or 200 Ω inputs (typical H pads) but instead with 1500 Ω or higher input impedance and low output impedance, using U pad technology. See Chapter 15 *Audio Transformers* for a discussion of the distortion and level handling characteristics of audio transformers.

25.1.2.2 Input Impedance

As shown in Fig. 25-5, some input transformers have input impedances that load the microphone and, as discussed in the preceding section, alter the response of the system at frequency extremes. However, well-designed transformers such as the Jensen JT-16B have substantially flat input impedance as shown in Fig. 25-6.

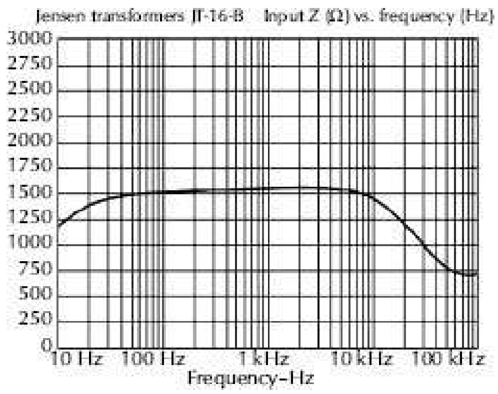


Figure 25-6. Input impedance of a Jensen JT-16B input transformer.

High preamplifier input impedance will improve the performance of many condenser microphones. For example the maximum *SPL* of a Shure KSM137 microphone is 145 dB SPL with a 5000Ω load, but reduces to 139dB with a 2500 Ω load, and to 134dB with a 1000 Ω load.

Likewise some condenser microphones suffer from dramatically reduced maximum *SPL* before distortion when the preamp input impedance drops below the microphone's specified minimum load impedance. One popular condenser microphone loses 12dB of maximum *SPL* capability when the load goes from 1000Ω to 750Ω.

25.1.2.3 Noise

The random motion of electrons in electrical conductors creates a voltage variously called *thermal noise, white noise*, or *Johnson noise* after its first observation by J. B. Johnson of Bell Labs in 1927. Thermal noise voltage is proportional to both temperature and the resistance of the conductor and is calculated as follows:[1]

$$E_t = \sqrt{4kTR\Delta f} \qquad (25\text{-}1)$$

where,
E_t is the thermal noise in V_{rms},
k is Boltzmann's constant or 1.38×10^{-23}Ws/K,
T is the temperature of the conductor in K,
R is the resistance of the conductor in Ω,
Δf is the noise bandwidth in Hz.

At a room temperature of 300K (80°F or 27°C), $4kT = 1.66 \times 10^{-20}$. For noise in the audio band of 20Hz to 20kHz, bandwidth is 19.98kHz. It's important to note that noise bandwidth here refers to a rectangular "brick wall" response, not the more conventional measure at the −3dB points. For a 150 Ω resistance under these conditions, noise is

$$\begin{aligned} 223\,\text{nVrms} &= -133.0\,\text{dBV} \\ &= -130.8\,\text{dBu} \end{aligned}$$

For a 200Ω resistance under the same conditions, noise is

$$\begin{aligned} 258\,\text{nVrms} &= -131.8\,\text{dBV} \\ &= -129.5\,\text{dBu} \end{aligned}$$

Here we use the nominal impedance of an idealized microphone simply to allow a simple but fair comparison of preamplifier noise performance.

Regardless of whether the conductor is copper wire, silver wire, an expensive metal-film resistor, or a cheap carbon resistor, the thermal noise is exactly the same! *Excess* noise refers to additional noise generated when dc flows in the resistor. Excess noise varies markedly with resistor material and construction. Note that only the resistive portion of impedance generates noise—pure inductors and capacitors do not generate thermal noise. Therefore, in our Shure SM57 circuit model of Fig. 25-1, thermal noise is generated by the 300Ω resistance but not by the 6mH inductance.

In a practical microphone preamp, we are usually concerned with the signal-to-noise ratio at the output. Although there may be many sources of internal noise in the preamplifier and its gain may be varied over a wide range, for simplicity noise is usually stated in terms of *EIN* or *equivalent input noise*. This simplification works because, in a good design, the dominant noise source is the first amplification stage and subsequent stages contribute no significant noise.

As shown in Fig. 25-7, the *EIN* has three components:

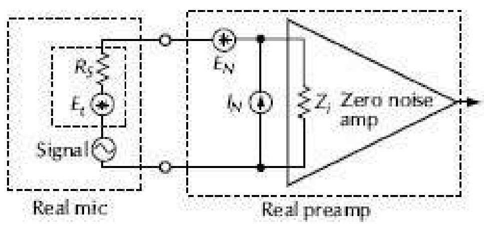


Figure 25-7. Contributions to equivalent input noise.

1. E_t: the thermal noise of the source resistance.
2. E_n: the voltage noise of the amplifier.
3. I_n: the current noise of the amplifier.

When noise voltages are produced independently and there is no relationship between their instantaneous amplitudes or phases, they are said to be uncorrelated. Total noise power is the sum of individual noise powers. Therefore, the resultant voltage is the square root of the sum of the squares of the individual voltages. For example, adding two uncorrelated 1 V noises will result in a 1.4 V noise because

$$E = \sqrt{1^2 + 1^2} = \sqrt{2} = 1.414 \tag{25-2}$$

When adding two noises, unless the second is a third or more of the first (less than 10dB difference), it will have little effect on the total. For any amplifier, minimum total noise is added when the source resistance is such that I_N flowing through the source creates a noise voltage equal to E_N. This source resistance is called the *optimum* source resistance for that particular amplifier. Perhaps the most useful function of an input transformer in a microphone preamplifier is to convert, as explained in Chapter 15 *Audio Transformers*, the impedance of the microphone to this optimum value in order to maximize *SNR*.

Measurement of noise is fertile ground for technical misrepresentation. Some rather unbelievable *EIN* numbers have appeared over the years. Most were based on measurements taken with the preamplifier input shorted, which ignores the noise contributions of both R_S (source resistance) and I_N (amplifier current noise), leaving only E_N (amplifier voltage noise). Bias current noise generates additional voltage noise when it flows in the source impedance (not just resistance). In this case the inductance of our SM57 model will indirectly contribute real-world noise. To have any meaning at all, *EIN* must specify the source impedance. With a 150Ω source resistance,

$$\begin{aligned} EIN &= 223\ \text{nVrms} \\ &= -133.0\ \text{dBV} \\ &= -130.8\ \text{dBu} \end{aligned}$$

for an ideal noiseless amplifier. If the preamplifier noise is equal to that of the source, *EIN* will be 3dB higher or −130.0dBV = −127.8dBu. *Noise figure*, or NF, is a measure of *SNR* degradation attributed to the amplifier—in this case 3 dB. From an engineering point of view there is little point in attempting to achieve NF below 3dB.[2]

Note that the thermal noise Eq. 25-1 also includes a term for bandwidth. Noise specifications such as *EIN* frequently appear in data sheets without a specified noise bandwidth. All other things equal, noise increases as the square root of bandwidth. Therefore, there is 1.25 dB less noise in a 15 kHz bandwidth, and 3 dB less noise in a 10kHz bandwidth, than in a 20kHz bandwidth. Likewise, while measurements such as A-weighted noise are both legitimate and useful, they cannot be directly compared to unweighted measurements. When comparing noise specifications, be sure it's an "apples to apples" comparison.

25.1.2.4 Bandwidth and Phase Distortion

Performance in the *time domain*, or waveform fidelity, is critically important to accurate music reproduction. Accurate time domain performance, sometimes called *transient response*, requires low phase distortion. Changes in amplitude response vs. frequency also create changes in the phase response. If the response is minimum phase (most analog circuitry has a minimum phase characteristic), then the amplitude defines the phase response and vise-versa. Pure time delays exhibit a linear phase versus frequency characteristic. True phase distortions are expressed as DLP or *deviations* from this linear phase relationship. Phase *shift* is not necessarily phase distortion.[3] In order to achieve a DLP of 5° or less from 20Hz to

20kHz, frequency response must extend from 0.87Hz to 35kHz, assuming 6dB per octave (first order) filter responses. If the low-pass filter is a second order Bessel, the cutoff frequency can be as low as 25 kHz.[4] Notice that extreme high-frequency response is not required, but extended low-frequency response is!

Phase distortion not only alters musical timbre, but it has potentially serious system headroom implications as well. Even though frequency response may be flat, peak signal *amplitudes* can increase up to 15 dB after passing through a network with high phase distortion. This can be a serious problem in digital recording systems. Even ultrasonic phase distortions caused by undamped resonances can excite complex *audible* cross-modulation products in subsequent nonlinear (any real world) amplifier stages.[5]

Low-frequency phase distortions are often described as muddy bass and high-frequency phase distortions as harshness or midrange smear. The complex cross-modulation products are usually described as dirty sounding and often are the cause of listener fatigue.

25.1.2.5 Common-Mode Rejection, Phantom Power, and RF Immunity

Common-mode rejection, as discussed in Chapter 15 *Audio Transformers*, is not just a function of the amplifier input circuitry. It depends on the impedance balance achieved by the combination of the microphone's output circuitry, cable, and the preamp's input circuitry. Common-mode rejection ratio (CMRR) is seldom an issue with dynamic microphones because, as shown in Fig. 25-1, the common-mode impedances are small parasitic capacitances. However, when phantom power is involved, very high CMRR can be difficult to achieve. The circuitry in the microphone that extracts phantom power from the two signal lines, as shown in the examples in Chapter 20 *Microphones*, can unbalance their line impedances to ground. The resistors R_2 and R_3 that supply phantom power, shown in the preamplifier of Fig. 25-8, must also be tightly matched to achieve high CMRR. For example, CMRR may be limited to 93 dB if ±0.1% resistors are used and may be limited to 73dB if ±1% resistors are used. For comparison, the JT-16B transformer used in Fig. 25-9 achieves a CMRR of 117dB when phantom power resistors are absent.

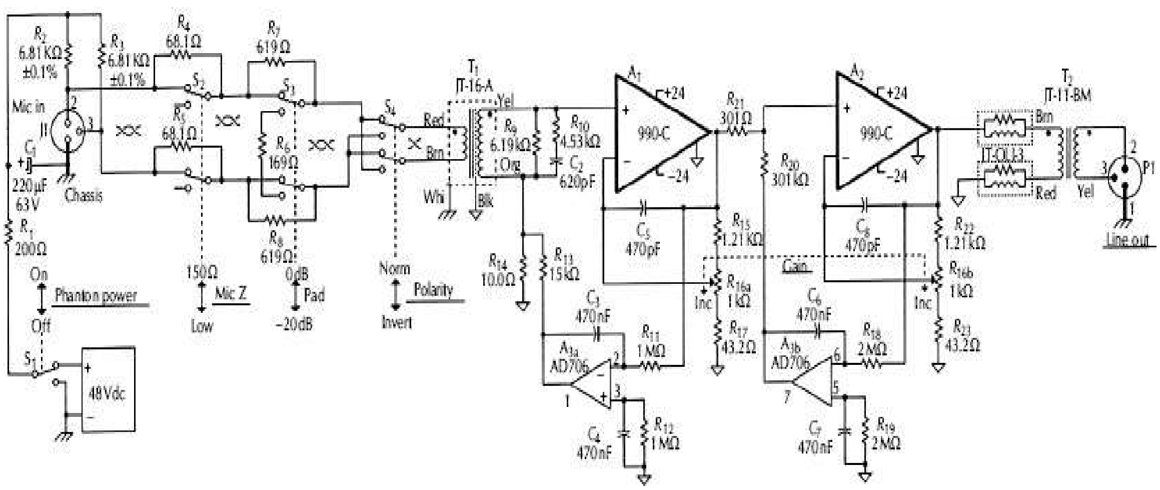


Figure 25-8. Signal path schematic of Jensen Twin-Servo 990 microphone preamplifier. Courtesy Jensen Transformers, Inc.

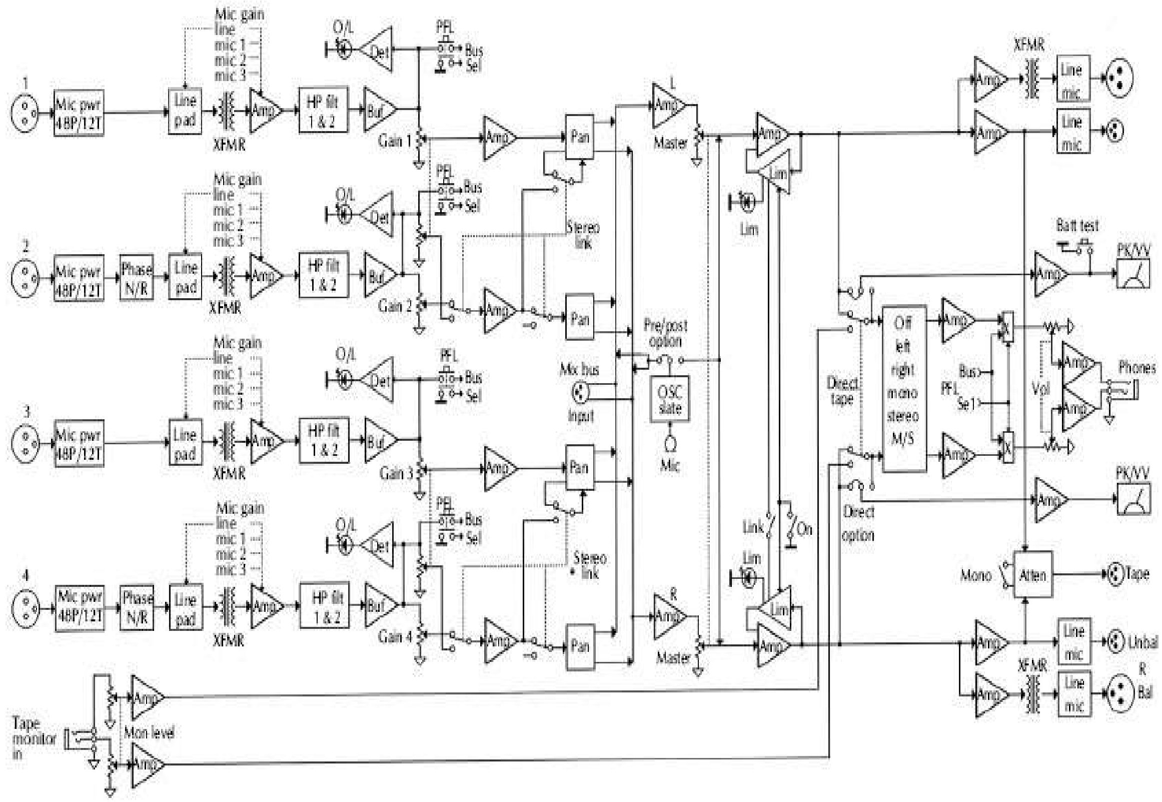


Figure 25-9. Block diagram of the Cooper Sound CS 104.

Connecting or disconnecting phantom power microphones while phantom power is supplied can generate large transient voltages. If the preamp is not carefully designed these transients can degrade the noise performance of the preamplifier input. The resulting noise can also potentially damage hearing and loudspeakers. Therefore best practice is to turn off phantom power before connecting or disconnecting phantom powered microphones.

Sometimes, as in Fig. 25-10, phantom power is supplied through a center-tap on a microphone input transformer. This presents a transformer design problem that can be even more difficult—simultaneously matching both the number of turns and the dc winding resistance on each side of the center tap. The current edition of the IEC 61938 Standard no longer allows phantom power via a center tap on a preamplifier input transformer.

RF interference, usually in the form of common-mode voltage, is another potential problem for microphone preamplifiers because it is likely to be demodulated in amplifier circuitry. In transformer-less circuits, suppression measures usually consist of capacitors from each input to ground and sometimes series resistors, chokes, or ferrite beads. Unless the capacitors are carefully matched, they will *unbalance* the common-mode input impedances and degrade CMRR. Because they also lower common-mode input impedances, they can make the circuit more sensitive to normal impedance imbalances in the microphone. Benchmark Media uses a common mode choke between the preamplifier input and the input bypass capacitors which increases the common-mode input impedance at higher frequencies, and reduces these problems. These tradeoffs can be largely avoided by using a Faraday-shielded input transformer that has inherent RF suppression characteristics.

A good microphone preamplifier should also be free of the so-called pin 1 problem and should confirm to the AES48, AES54-1, AES54-2, and AES54-3 Standard and the June 1995 special issue of the AES Journal "Shields and Grounds." The microphone cable should be free of shield-current-induced noise (SCIN), which can be a serious problem with foil shield and drain wire construction. Both of these problems are discussed in Chapter 18 *Transmission Techniques: Wire and Cable*.

25.2 Real-World Preamp and Mixer Designs

25.2.1 Transformers

Manufacturers of microphone preamplifiers have a natural desire to differentiate their product from all others. One of the major divisive issues is the use of audio transformers. According to the antitransformer camp, all audio transformers have inherent limitations such as limited bandwidth, high distortion, mediocre transient response, and excessive phase distortion. Unfortunately, many such transformers do exist and not all of them are cheap. The makers of such transformers are simply ignorant of sonic clarity issues, have a poor understanding of the engineering tradeoffs involved, or are willing to take manufacturing shortcuts that compromise performance to meet a price.

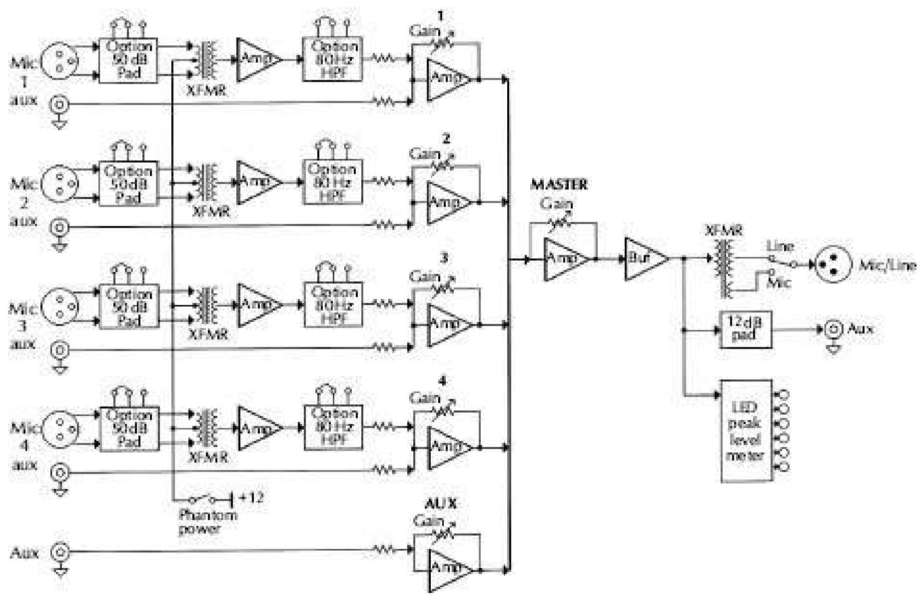


Figure 25-10. Block diagram of a Shure SCM268.

As stated earlier, bandwidth and phase distortion are intimately linked in any electronic device. A very high level of performance can be reached with proper transformer design. Consider the Jensen JT-16B microphone input transformer. Its frequency response is −3dB at 0.45 Hz and 220kHz and −0.06dB from 20Hz–20kHz, with a second order Bessel high-frequency roll-off characteristic. Low frequency roll-off is less than 6dB per octave owing to properties of the core material, which further improves phase performance. Its deviation from linear phase is under 2° from 20Hz–20kHz, giving it truly excellent waveform fidelity and square-wave response.

As discussed in detail in Chapter 15 *Audio Transformers*, audio transformer distortion is quite different from electronic distortion in ways that make it unusually benign. First, transformer distortion is frequency and level dependent. Significant distortion occurs only at low frequencies and high signal levels, typically dropping to under 0.001% above a few hundred hertz. Second, the distortion is nearly pure third harmonic and is *not* accompanied by the high levels of much more irritating intermodulation distortion that occurs in electronics.

A high degree of RF attenuation, both normal mode and commonmode, is also inherent in transformers that contain Faraday shields. For example, in Jensen designs, common-mode attenuation is typically over 30dB from 200kHz–10MHz. And, as discussed in Chapter 15 *Audio Transformers*, transformers enjoy a great CMRR advantage over most electronically balanced input stages because they are relatively insensitive to the impedance imbalances that normally exist in real-world signal sources. If well designed and properly applied, audio transformers qualify as true high-fidelity devices. They are passive, robust, and stable and have significant advantages, especially in electrically hostile environments.

25.2.2 Class A Circuitry

Another divisive issue among preamplifier manufacturers involves class A circuitry. Although it has certain advantages, it is not necessarily inherently superior to much more widely used class AB designs. Class A operation occurs when the active device (or devices in the case of a push-pull output stage) conducts current during the entire 360° signal cycle. Class AB occurs when each device conducts for more than 180° but less than 360°. In class B operation, each device conducts for exactly 180°. In class C, conduction is less than 180° and this is generally done only in RF circuits or where intentional distortion is desired.

Most op-amp output stages operate class AB to reduce crossover distortion of small signals. Practical active devices are generally unable to behave linearly near zero current (cutoff) as is required for low-distortion pure class B operation. A small idling or quiescent current flows in both devices at zero signal and operation remains class A (both devices conducting for full signal cycle) up to some signal level, at which point one device begins to be cut off for part of the cycle, producing class AB operation.

For example, in the Jensen-Hardy 990 amplifier module used in the circuit of Fig. 25-8, the output stage quiescent current is about 15 mA. Therefore, amplifier operation is class A until peak output current (plus or minus) reaches about 15 mA. Peak output current, of course, depends on peak signal level and load impedance. For example, the output voltage clips at about 24 Vpeak, so any load impedance higher than about 1.6 kΩ results in class A operation at all times. Likewise, with a 600 Ω load, operation is class A until output signal level reaches ±9V peak. Above that peak level, operation becomes class AB. The “front end” circuitry of the 990, like most operational amplifiers, always operates class A unless the output is clipped.

Even in a well-designed class AB circuit, there is a slight nonlinearity as the signal transitions from positive to negative and back. At high signal levels this non-linearity is insignificant, but at low signal levels becomes of greater significance. One solution, popularized by Rupert Neve, is to bias the circuitry so the AB transition does not happen at the signal zero level crossing, but instead at a higher signal level. With this approach the lowest audio signals only pass through one of the output devices, and these low level signals are amplified essentially class A.

The line between class A and AB operation is very distinct: operation is no longer class A as soon as current in any active device (vacuum tube or transistor) becomes zero. The main advantage of class A circuit designs is that the curvature of the nonlinearity plot is likely to be smoother (i.e., free of a sharp discontinuity at crossover) so that there will be fewer problems related to negative feedback, slew rate, and gain-bandwidth limitations.

25.2.3 Shure SCM268 Four-Channel Mixer

The Shure SCM268 is an example of a compact, simple mixer with basic features. Notable features include transformers on balanced inputs and outputs, mic or line level output, phantom power, and optional pads allowing for balanced line level inputs, Fig. 25-11. A functional block diagram is shown in Fig. 25-10.

Figure 25-11. Shure SCM268 four-channel microphone mixer. Courtesy Shure Incorporated.

25.2.4 Cooper Sound CS 104 Four-Channel ENG Mixer

The Cooper Sound Systems CS 104 is an example of a portable, battery-powered mixer with a number of sophisticated features, Fig. 25-12. Notable features include stereo mixing, pan pots and channel linking, transformers on main inputs and outputs, built-in stereo limiter, input overload indicators, selectable high-pass filters, prefade listen, tape monitor, and built-in tone and slate functions. Fig. 25-9 is its functional block diagram.

Figure 25-12. Cooper Sound Systems CS 104 battery-powered portable ENG/EFP four channel audio mixer.

25.2.5 Jensen-Hardy Twin-Servo® 990 Microphone Preamplifier

The Jensen-Hardy Twin-Servo® 990 Microphone Preamplifier is an example of a high-performance design, Fig. 25-13. It features patented discrete component 990 amplifier modules that combine low-input noise, high-output voltage and current, low distortion, and high gain-bandwidth performance that is unavailable with integrated circuits. It uses two cascaded variable gain stages per channel to maintain high bandwidth and low distortion overall. Unlike most designs, this topology also keeps EIN very low at the lowest gain settings. Extended low-frequency response is preserved by using dc servo feedback circuitry to eliminate coupling capacitors and their attendant problems, Fig. 25-8.

Figure 25-13. Jensen Twin-Servo® 990 four channel microphone preamplifier. Courtesy Jensen Transformers, Inc.

25.2.6 Rane TTM56S

The Rane TTM56S is a DJ mixer that uses a unique magnetic fader. A contactless design means no noise or dust that may enter via the slot in the panel, Fig. 25-14.

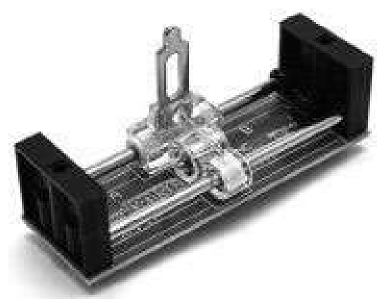


Figure 25-14. Rane magnetic fader. Courtesy Rane.

25.3 Automatic Microphone Mixers

Automatic microphone mixers, also known as *voice-activated mixers* or *sound-activated mixers*, have become a necessary part of sound systems designed for speech. All automatic microphone mixers have a fundamental function: to attenuate (reduce in level) any microphone that is not being spoken into by a talker, and conversely, to rapidly activate any microphone that is being spoken into by a talker. An automatic microphone mixer should be considered when the number of microphones required for the sound system is four or greater, or when more than a single microphone might pick up the same voice. For example a talker wearing a lapel or earset microphone who then approaches a podium with a gooseneck mic, or two actors wearing lapel or earset microphones who approach each other on stage.

When used in a sound reinforcement system, an automatic microphone mixer provides a significant increase in gain before feedback when multiple microphones must be used without a sound engineer. It also improves the quality of the sound system output by reducing the amount of extraneous room sound being picked up, and by reducing comb filtering. In addition, it automatically adjusts system gain to compensate for the number of microphones in use at any instant. Thus, an automatic microphone mixer attempts to provide the same system control that might be produced by a skilled human sound engineer who can react instantly as multiple persons talk.

As automatic microphone mixers are optimized for speech applications, their use in musical applications is not recommended. Mixing microphones for music is as much art as science, and therefore the artistic judgment of a human sound engineer is much preferred to the electronic decision process of an automatic microphone mixer.

In summary, when used in a speech sound system with multiple microphones, the ideal automatic microphone mixer assures that the number of active microphones at any moment equals the number of active talkers at the same moment. All unused microphones at that moment are attenuated.

25.3.1 The Audio Problems Caused by Multiple Open Microphones

High-quality audio becomes progressively more difficult to achieve as the number of open microphones increases. All audio systems face the same problems whenever multiple open microphones are needed. These problems are:

1. Build-up of background noise and reverberation.
2. Reduced gain before feedback.
3. Comb filtering.

These problems can plague boardrooms, city council chambers, conference centers, houses of worship, teleconferencing rooms, radio talk shows—anywhere multiple microphones are used. Since audio quality rapidly deteriorates as the number of open microphones increases, the solution is to keep the minimum number of microphones open that will handle the audio. An automatic microphone mixer keeps all unused microphone input channels attenuated, and activates any microphone spoken into within milliseconds.

25.3.1.1 Buildup of Background Noise and Reverberation

The first problem of multiple open microphones is the buildup of background noise and reverberation. This buildup can adversely affect the quality of recordings or broadcasts originating from the audio system. Consider the case of a city council with eight members and eight microphones. For this example, only one member is talking. If all eight microphones are open when only one microphone is needed, the audio output will contain the background noise and reverberation of all eight microphones. This means the audio signal will contain substantially more background noise and reverberation than if only the talker's microphone was open. This buildup of background noise and reverberation greatly deteriorates the audio quality. Speech clarity and intelligibility always suffer as background noise and reverberation increase.

As the number of open microphones increases, the background noise and reverberation in the audio output also increase. In our city council example, the audio output from eight open microphones would contain 9 dB more background noise and reverberation than a single open microphone. To the human ear, the noise would sound almost twice as loud when all eight microphones were open.

To minimize background noise and reverberation buildup, an automatic microphone mixer activates only the microphone(s) being addressed and employs a Number of Open Microphones Attenuator (NOMA) circuit. NOMA systematically decreases the master gain whenever the number of open microphones increases. Without NOMA, the audio system would produce objectionable noise modulation (pumping and breathing) as background noise and reverberation increase and decrease with the number of open microphones. With a properly designed automatic microphone mixer, background noise and reverberation remain constant no matter how many or few microphones are activated.

25.3.1.2 Reduced Gain Before Feedback

The second problem of multiple open microphones is reduced Gain Before Feedback (GBF). Acoustic feedback ("howling") can be a problem anytime a sound reinforcement (PA) system is used. To avoid feedback, PA systems are operated below the point where the system becomes unstable and starts to howl. However, this feedback stability margin is reduced each time another microphone is opened. Have one too many open microphones and the result is feedback. When the GBF reaches 0dB the sound system goes into sustained feedback, but even a GBF of −10dB can cause audible ringing.

The automatic microphone mixer solution is to keep unused microphones turned off and utilize NOMA. As more microphones are activated, the overall gain will remain constant thanks to the NOMA circuit. An automatic microphone mixer assures that as long as the GBF of all microphones are the same if the audio system does not feedback when any one microphone is open, the system will remain feedback free even if all the microphones are open.

25.3.1.3 Comb Filtering

The third problem of multiple open microphones is comb filtering. Comb filtering occurs when open microphones of the same gain at different distances from a talker are mixed together and the level at each microphone is close to the same, Fig. 25-15. If the level difference as mixed is greater than 10dB then audible comb filtering is not usually a problem. Since sound travels at a finite speed, the talker's voice arrives at the microphones at different times. When combined in a mixer, these out-of-step microphone signals produce a combined frequency response very different from the frequency response of a single microphone. (A frequency response chart of the out-of-step signals looks like the teeth of a hair comb, thus the name.) The aural result of comb filtering is an audio signal that sounds hollow, diffuse, and thin. Even worse, if the distances are not constant such as when the talker moves, the frequencies at which the notches occur also move creating a very noticeable swishing sound.

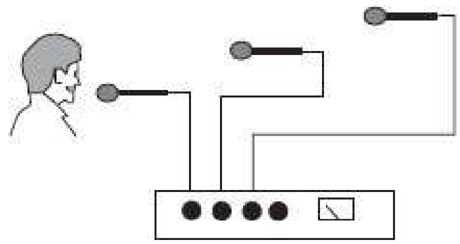


Figure 25-15. Comb filtering occurs whenever open microphones at different distances from a talker are mixed together.

The solution to comb filtering also is keeping the number of open microphones to an absolute minimum. When two or more microphones receive close to the same sound level, an automatic mixer can select the microphone with the strongest signal and attenuate the other microphones. This reduces the spatial area where audible comb filtering occurs and improves the sound quality.

25.3.1.4 Summary

1. Keeping the number of open microphones to a minimum always improves overall audio quality.
2. The primary function of an automatic microphone mixer is to keep unused microphone input channels attenuated (turned down or off) and to instantaneously activate microphones when needed.
3. Buildup of background and reverberant noise, reduced gain before feedback, and comb filtering can all be controlled by using an automatic microphone mixer.
4. An automatic mixer acts like a skilled operator would if he had enough hands to operate all level controls and never was distracted.

25.3.2 Design Objectives for Automatic Microphone Mixers

As shown in Fig. 25-16, a conventional microphone mixer in a sound system amplifies the signal from each microphone and combines these amplified signals together to produce a single output. This output feeds a power amplifier and then one or more loudspeakers. Each doubling of the number of open microphones feeding into a sound system reduces the available gain before feedback by 3 dB. This fact surprises the layman who often believes that more microphones equate to the sound system being louder, not softer. A sound system with numerous microphones easily becomes ineffective if a sound engineer is not present to control levels and switch off unused microphones. Since gain before feedback can often be marginal because of the acoustical characteristics of a room, an automatic microphone mixer may be the only way to provide adequately loud program levels to the audience with an unattended sound system.

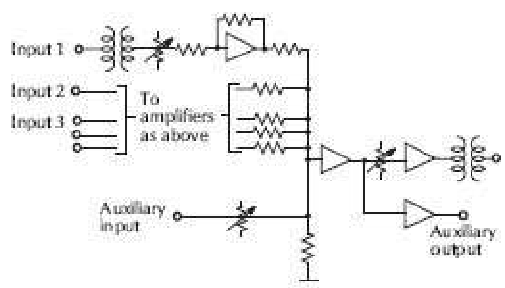


Figure 25-16. Simplified diagram of a microphone mixer.

Examples of Design Objectives for an Automatic Microphone Mixer

1. Keeps the sound system gain below the threshold of feedback

instability.

2. Requires no operator or sound technician at the controls.
3. Does not introduce spurious, undesirable noise or distortion of the program signals.
4. Can be installed as easily as a conventional mixer.
5. Responds only to the desired speech input signals and is relatively unaffected by extraneous background noise signals.
6. Activates input channels fast enough that no audible loss of speech signals occurs.
7. Allows more than one talker on the system when required by the discussion content while still maintaining control of the overall sound system gain.
8. Adjusts the system gain to compensate for a range of talker input levels.
9. Provides system status outputs for peripheral equipment control and can interface with external control systems for advanced system design if required.

The automatic microphone mixer operation should provide relatively easy and very rapid input activation. Desired speech from a talker should cause immediate activation of the appropriate input channel, which may not always happen if the design of an automatic microphone mixer is poor. Also, random false activation of microphones remote from the talker can occur with some automatic microphone mixer designs. However, this false activation is typically not troublesome as the false signals are normally much lower in level than the desired talker signal. The automatic microphone mixer is doing its job if all talkers are clearly heard by the audience when they speak, and the sound system remains below the point of feedback.

An automatic microphone mixer cannot improve the performance of microphones. Its primary benefit comes from limiting the number of microphone signals fed to the mixer output. A side benefit is often the apparent increase of critical distance in a multiple microphone system. (Critical distance can be defined as a point in the room where the direct signal of the talker equals the reflected signal of the talker, i.e., 50% direct signal and 50% reverberant signal.) Because unused microphones remote from the talker are attenuated, room reverberation and ambience that would otherwise be amplified are reduced.

25.3.3 Controls and Features of Automatic Microphone Mixers

Automatic microphone mixers have many of the same controls and features of manual microphone mixers. Examples are:

- Level control for each input channel.
- Master level control for each output channel.
- Input signal attenuation ("trim").
- Phantom power.
- Two or three band equalization for each input channel.
- Output level metering.
- Output signal level limiter.
- Nonautomatic auxiliary inputs.
- Headphone output with level control.

These controls and features may be configured in hardware–e.g., switches, potentiometers, LED strings—or they may be configured in software. In either case, the function of the control or feature remains the same.

25.3.3.1 Controls and Features Unique to Automatic Microphone Mixers

As automatic microphone mixers typically perform more functions than a manual microphone mixer, there are controls and features that are unique to automatic microphone mixers.

Input Channel Threshold (Gated style automatic mixers only). Determines at what signal level a gated automatic microphone mixer input passes the incoming microphone signal to the mixer's output. Gain sharing automatic mixers do not have or need this control.

Input Channel On Indicator (Gated style automatic mixers). Illuminates to indicate that an input channel has gated open and is passing the microphone signal onto the mixer output. Gain sharing automatic mixers many provide a similar indicator by a different means.

Direct Output for Each Input Channel. Provides an isolated output for each input channel that may be unaffected by the automatic microphone mixer action. Individual channel outputs of gain sharing automatic mixers have the automatic mixing action already applied.

Last Microphone Lock On (Gated style automatic mixers only). Keeps on the most recently activated input channel on a gated automatic microphone mixer until another input channel is activated. This maintains room ambience when the automatic microphone mixer is used to provide a broadcast feed, a recording feed, or a feed to an assistive hearing system.

Hold Time (Gated style automatic mixers only). Keeps an activated input channel on a gated automatic microphone mixer on for a period of time after speech has ceased. This feature bridges the natural gaps that occur in speech patterns.

Input Attenuation. Determines how much gain reduction is applied to an input channel of a gated automatic microphone mixer when the channel is not activated. Typical range of adjustment is 3dB to 70dB of attenuation, with 15 dB being a common value. Most automatic mixers have two gain states for each input, On or Off. Some automatic mixers have a three state design with Active, Rest, and Off as the three states. In such a design the Rest state attenuation might be around 12dB, and the Off state attenuation might be around 30dB.

Decay Time (Gated style automatic mixers only). Establishes the time required for an input of a gated automatic microphone mixer to be lowered from the activated state to the attenuated state. Decay time is always in addition to the hold time.

Manual/Auto Select. Allows the automatic microphone mixer to operate in a nonautomatic (manual) mode.

25.3.3.2 External Control Capability and Status Indication of Automatic Microphone Mixers

Most automatic microphone mixers include the ability to be controlled by external switches, potentiometers, touch screens, personal computers, and other types of control devices. These devices are typically connected to the automatic microphone mixer via screw terminals or multipin connectors on the mixer’s rear panel on analog mixers, while digital automatic mixers typically use a data port or network connection. The controllable functions and the communication protocol depends upon the manufacturer and model of the automatic microphone mixer. Examples of automatic microphone mixer functions that can be externally controlled follow.

Gain of an Input Channel or the Master Output. In a courtroom, the court clerk could control the volume level of the witness microphone or the entire sound system using a potentiometer located at a distance from the automatic microphone mixer.

Individual Microphone On/Off Switch. Allows a user to disable their microphone if they have no intention of talking soon. This helps prevent side conversations from being amplified. This switch might be in the form of a push-on push-off push button near the microphone that illuminates when on.

Mute an Input Channel. In legislative facilities, a member could have a privacy or "cough" switch. Such a switch might be located where it would be easy for an aide to press if they needed to have a private word with the member.

Global Mute of All Input Channels. In a government hearing room, the presiding member could mute all inputs to regain control of a meeting. This switch might also force his mic on while it mutes all others.

Ambient Microphone. Systems that have a broadcast feed might automatically activate an ambient pickup microphone when none of the regular system microphones is in use to prevent the appearance of “dead air” which broadcasters hate. Such an ambient microphone if provided should only drive the broadcast output not the sound reinforcement output of the automatic mixer,

Routing of Input Channels to Different Outputs. In a hotel meeting facility with movable dividing walls, input channels could be sent to different banks of loudspeakers depending on the room configuration.

Status Terminal. As input channels are activated and attenuated by the automatic mixing process, it is valuable to have a status terminal that indicates if a particular input channel is activated or not. This status terminal, also known as a *gate terminal*, can be thought of as an electronic switch that changes from open to closed based on the activity of the input channel. Examples of use for a status terminal:

- **Control of an LED or lamp to indicate input channel activity**. In a city council chamber, a council member could have a tally light located near the microphone indicating when the microphone’s input channel is activated by the automatic microphone mixer.
- **Control of a relay used to attenuate the nearest loudspeaker**. As the typical feedback path in a sound system is between a microphone and the nearest loudspeaker, attenuating the closest loudspeaker when a microphone is active could improve gain before feedback. A better solution, however, is to utilize a mix-minus or matrix-mix approach where each microphone never feeds the nearby loudspeakers. By not muting loudspeakers, you can still hear distant talkers even while you are

talking.

- **Control of a video switcher connected to multiple cameras**. In a courtroom, the proceedings could be videotaped by using cameras that follow the activation of input channels by the automatic microphone mixer.
- **Mute other input channels**. In a hotel meeting facility, one input channel could override all others in case of an emergency announcement.

Combining the externally controlled functions with the status terminals provides hundreds of unique system configurations. Most manufacturers of automatic microphone mixers have documentation of such configurations, often printed in product installation manuals and available on the manufacturer's web site. As previously noted, the communication protocol used to interpret the status terminals and control the mixer functions depends upon the manufacturer and model of the automatic microphone mixer.

25.3.3.3 Examples of Communication Protocols Used in Automatic Microphone Mixers

Contact Closure Protocol. The most basic of communication protocols, contact closure is provided by a simple single pole/single throw (SPST) switch or relay. The switch is connected to two terminals on the mixer that control a certain function—e.g., mute of an input channel. When the switch is closed, the input channel is muted. When the switch is open, the input channel is unmuted and can be activated.

Resistance Change or Voltage Change Protocol. Used primarily to control signal levels via a VCA (voltage-controlled amplifier), this protocol requires that defined changes in resistance or voltage be applied to the mixer's control terminals. In response, the VCA in the automatic microphone mixer will change the level of the audio signal.

TTL (Transistor-Transistor Logic). An electronic protocol established in the 1960s, TTL is simple to use. A control terminal on the automatic microphone mixer has one of two states: logic high (+5Vdc) or logic low (0Vdc). A status terminal could be logic high when a mixer input channel is attenuated and logic low when the mixer input channel is activated. This change of voltage informs an external control device that there is a change in the input channel status and some predetermined action should take place, e.g., illuminate an LED or switch on a camera.

RS-232. Used for communication with a computer, RS-232 is another common electronic protocol. RS-232 is most often used when proprietary control software is supplied with the automatic microphone mixer or when the mixer is connected to a control system such as those manufactured by Crestron or AMX.

RS-422. Basically a balanced line version of RS-232, RS-422 is designed for situations where an extremely long cable run must be used to connect the automatic microphone mixer to the external control device.

Ethernet. The most widely used protocol for computer networks. DSP based automatic microphone mixers will often provide the ability to remote control and monitor the automatic mixer using this protocol.

25.3.3.4 Number of Open Microphones Attenuation (NOMA)

NOMA is a function shared by all well-designed automatic microphone mixers. It is a simple method of ensuring system stability by automatically reducing the mixer output gain in proportion to the number of activated input channels. NOMA offsets the increase of gain that occurs as more microphones are activated. The attenuation in decibels should vary as:

$$\text{Attenuation in dB} = 10\log N \qquad (25\text{-}3)$$

where,
N is the number of activated microphones.

While NOMA helps maintain stable gain in a sound system as the number of activated microphones varies, it does not limit the number of microphones that can be activated.

25.3.3.5 Restricting the Number of Open Microphones (NOM)

Recent developments in automatic microphone mixer design have led to a feature best described as a *NOM restrictor*. This feature restricts the number of active input channels to a predetermined amount. For example, in a large legislative system with 100 microphones, it makes little sense to allow all 100 microphones to be active at any instant, even if all 100 legislators are talking.

Restricting the NOM to 5 microphones of the 100 allows spirited debate while not subjecting the audience to the cacophony of 100 open microphones. One possible difficulty is deciding which five of the 100 get to be active. The need for this sort of function depends greatly on the protocols of the group using the system.

25.3.3.6 Input Channel Attenuation

Automatic microphone mixers use some form of input channel attenuation to turn off unused microphones. The activation of an input channel becomes audibly apparent if the level change from the off state to the on state is too great or takes too long. Practical experience has shown that a 15 dB change from off to on is a good compromise for a two state automatic mixer. It is possible to use three attenuation states, on, rest, and off. If no one is talking the attenuations are set to a middle or rest state value. When speech is detected at one or more inputs, they ramp from the rest state up to the on state, and the unused inputs ramp from the rest state down to the off state. This minimizes both the amount of attenuation change, and the time it takes for the change to happen. Typical attenuation values for a three state automatic mixer might be; on 0dB, rest −12dB, off −30dB. As the number of microphones in the system increases, more input channel attenuation may be required for system gain stability. Adjustment of input channel attenuation is available on most automatic microphone mixers. This adjustment can be on an input- by-input basis or for all inputs at once. The relationship between gain before feedback, input channel attenuation, and the number of microphones is calculated by

$$\Delta G = 10\log\frac{N}{1+(N-1)10^{A/10}} \qquad (25\text{-}4)$$

where,

ΔG is the gain improvement in dB with only one microphone activated,

N is the total number of microphones,

A is the attenuation for all input channels in dB.

[Fig. 25-17](#) shows the relationship in graphical form. Note the asymptotic maximum value of gain improvement with infinite attenuation—i.e., all but one channel turned off. Also note that input channel attenuation greater than 30dB offers little improvement for systems with up to 256 microphones.

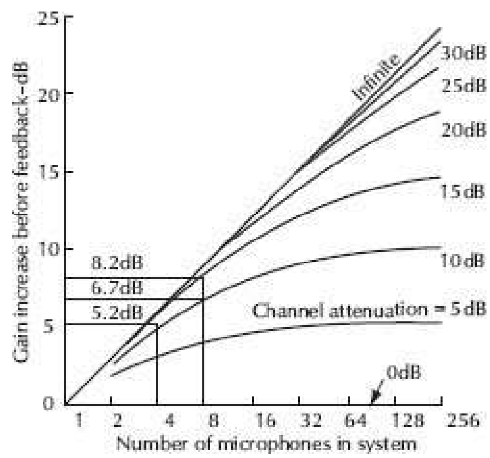


Figure 25-17. Gain improvement with different channel-off attenuations in a mixer that has a number of microphones and only one channel on.

25.3.3.7 Automatic Gain Control

Automatic gain control (AGC) of an input or output is a feature of a few automatic microphone mixers. A sound engineer rides gain to bring up weak signals or reduce overly loud signals and attempts to do this without destroying the inherent dynamic range of speech. An AGC in an automatic microphone mixer is typically designed to reduce gain only should the input signal level increase, or if gain increase is allowed it is limited in amount to prevent a quiet talker from sending the system into feedback. The AGC is adjusted so that the quietest talker has maximum gain (without feedback). All louder talkers will force the AGC to bring down the overall level.

The IRP Level-Matic circuit is an example. It automatically adjusts the master gain to maintain a uniform output level for input signal variations up to 10dB. A loud talker causes the gain to steadily decrease. When the talker stops, the gain holds as established by his or her average talking level. Such a gain hold when there is no speech is critical to prevent the AGC from bring up the background noise when no one is speaking. If a quiet talker then speaks, the gain steadily increases to a new value set by his or her average speaking level.

Some automatic mixers provide an AGC on each input rather than just on the overall mixed output. This helps when some talkers are quiet while others are loud. Each input gets the optimum gain for the talker at that input. If an AGC per input is used, it is important that the AGC follow the automatic mixer gain adjustment, so the automatic mixer function is not biased by the gain applied by the AGC.

Some designs with an independent AGC on each input compute the system stability continuously based on the gains applied by each AGC, and set a limit on the maximum gain possible in any AGC dynamically to maintain system stability. If there is just a single quiet talker, it is possible to allow the AGC for that input to add considerable gain without danger of feedback. However, if there are multiple quiet talkers the maximum gain the AGCs can be allowed to apply will be significantly less.

Some AGC designs allow either an indefinite gain hold time when there is no speech, or allow the selection of a slow ramp back to nominal gain after a time of no speech. Such a gain reset is useful if different talkers use the same microphone.

Gain control is based on the loudness versus frequency and loudness versus time response of the ear. Gain adjustments are made at a constant dB per second rate to minimize the pumping and breathing effects of simple level compression circuits. Some AGC designs use filtering on the gain control side-chains to reduce the sensitivity to noise and sounds other than speech.

25.3.4 Types of Automatic Microphone Mixers

An automatic microphone mixer can have an analog circuit design, a digital circuit design, or a combination of the two. Though a digital design might offer more design flexibility due to software control, the digital automatic microphone mixer is not inherently better than an analog automatic microphone mixer. Be it analog or digital, an automatic microphone mixer will fall in one of the following functional groups:

1. Fixed threshold.
2. Variable threshold.
3. Gain sharing.
4. Direction sensitive.
5. Noise adaptive.
6. Multivariable dependent.

25.3.4.1 Fixed Threshold Automatic Microphone Mixers

A detector circuit in the automatic microphone mixer activates an input channel when a microphone signal is present and attenuates the input when the microphone signal ceases. This basic function is often called a *noise gate*. To activate the input, the signal must be larger than a threshold preset for the channel during installation. This method has several shortcomings. First, there is the dilemma of where to set the activation threshold. If it is set too low, it will respond falsely to room noise, reverberation, and room-reflected sound. If the threshold is set too high in an effort to avoid false activation, desired speech signals may be chopped or clipped. The threshold should be set high enough to avoid activation by random noises, but low enough to turn on with desired speech signals. These are frequently contradictory requirements, and compromise is generally not satisfactory.

A more serious problem is that any number of input channels may activate with a very loud talker. One solution is a first-on inhibiting circuit that permits only one input channel to be on at a time. One-on-at-a-time operation is generally unacceptable for conversational dialog because the hold time needed to cover speech pauses will keep the second talker off.

Fixed threshold automatic microphone mixers have fallen out of favor and are now rarely employed. Early examples of fixed threshold activation products include the Shure M625 Voicegate (1973), the Rauland 3535 (1978), the Edcor AM400 (1982), and the Bogen AMM-4 (1985).

25.3.4.2 Variable Threshold Automatic Microphone Mixers

One attempt at overcoming the problems of a fixed threshold is to set the activation threshold based on a signal from a remote microphone. This microphone would be located in an area that is not expected to produce desired program input and is presumed to provide a reference signal that depends on variations in room noise or reverberation. Any desired talker input must then exceed this level by some preset amount. It is assumed that the desired talker signal will be louder than the reference. However, this may not be true, especially when the reference signal from a randomly selected microphone location does not represent the ambient sound in the vicinity of the talker's microphone. This is the basis of a system described by Dugan in U.S. Patent 3,814,856. An alternative source of reference threshold may be derived from the sum of the outputs of all the microphones in the system.

The discontinued JBL 7510 automatic microphone mixer employed a variable threshold design to override a fixed threshold. This design assumed that if a common acoustical disturbance was sensed at several microphone input channels, an input channel should not be activated. Instead, the overall system threshold should be raised. A talker must then be loud enough at the microphone to override the new raised threshold. Both the fixed threshold and the contribution of the background threshold reference would be set at installation. Release time, input attenuation, and gain were also necessary adjustments for each input channel. Variations on this concept of variable threshold design have been used in automatic microphone mixers from Audio Technica, Biamp, IED, Ivie, Lectrosonics, Inc, and TOA.

The Biamp autoTwo Automatic Mixer, Fig. 25-18, includes adaptive threshold sensing to minimize false gate triggering, a speech frequency filter to minimize false gating due to noise, logic outputs from channels for switching external circuits, and 6dB of hysteresis to reduce gate fluttering when near threshold. The block diagram is shown in Fig. 25-19.

Figure 25-18. Biamp autoTwo Automatic Mixer. Courtesy Biamp

Systems.

25.3.4.3 Gain-Sharing Automatic Microphone Mixers

Dugan's U.S. Patent 3,992,584 describes such a system where the sum of the gains of all inputs is constant. The signal level at each input is compared with the sum of the levels of all inputs and the ratio between those levels is used to determine the gain applied to that input. Speech from two persons talking into separate microphones with levels differing by 3 dB (both appreciably above the background level) would appear at the output of the system with a 6dB difference. In other words, the signal from a microphone with the highest output is given the most gain, and a signal from a microphone with the smallest output is given the least gain. With this operational concept, NOMA is not needed in the output stage. Theoretically, the system is configured so that the total gain is constant at a level that safely avoids feedback oscillation.

A gain-sharing automatic microphone mixer works from the premise that the sum of the signal inputs from all microphones in the system must be below some maximum value that avoids feedback oscillation. The safe system gain is set relative to the sum of all microphone gains in the system. If one microphone has more signal than the average of all signals, then that microphone channel is given more gain and all the other channels less gain roughly in proportion to the relative increase of signal level. When there is no speech detected on any input, all the inputs go to a rest state gain, Fig. 25-20A. When a single input detects speech its gain goes up towards 0dB, and all the others go down towards an off state gain, Fig. 25-20B. When two inputs detect speech at the same level, they both rise towards a gain of –3 dB, while the others go down towards the off state gain, Fig. 25-20C. When four inputs detect speech at the same level, they rise towards a gain of –6dB, and the others go down towards the off state gain, Fig. 25-20D. Since the power gains of each input all sum to 0dB at all times (gain sharing) there is no need for NOMA to be applied on a summed output since the equivalent action is happening at the inputs.

It is important that the control side-chain be filtered to prevent noise from being mistaken as speech. Dugan used a high pass filter for this purpose, but a bandpass filter could also be used.

Analog and DSP automatic microphone mixers marketed by Dugan, Lectrosonics, Protech Audio, Altec Lansing, Rane, Peavey, and QSC have used level proportional control based on average input signal amplitudes, Fig. 25-21.

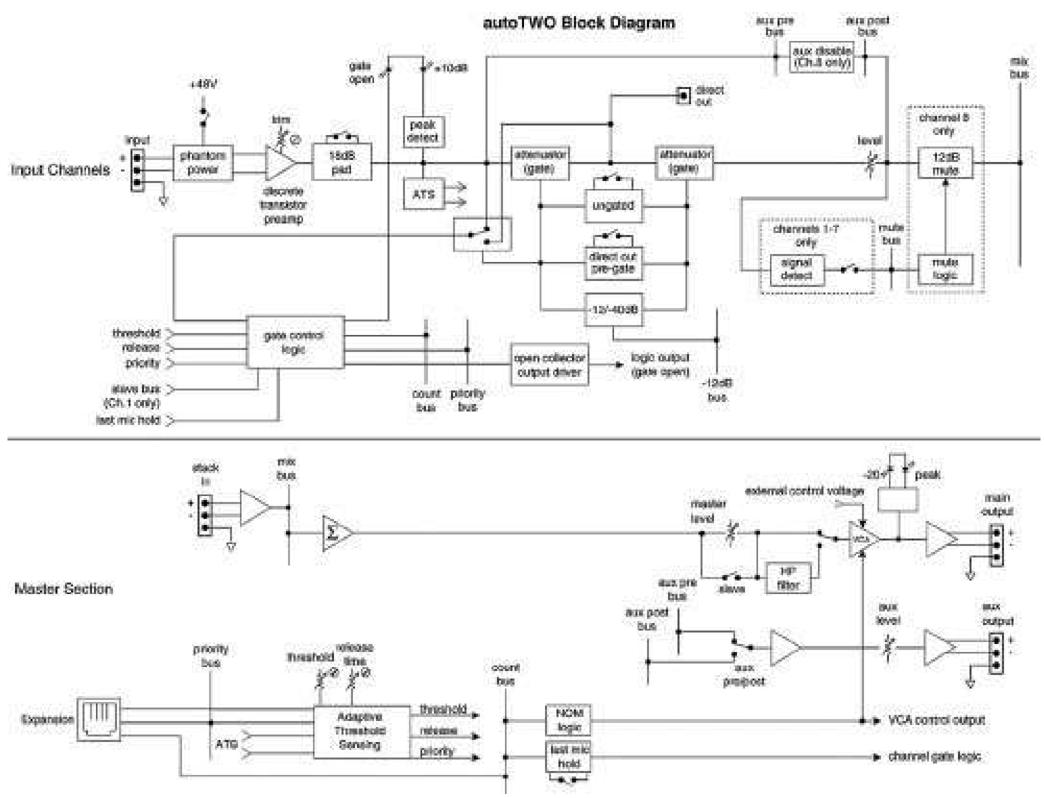
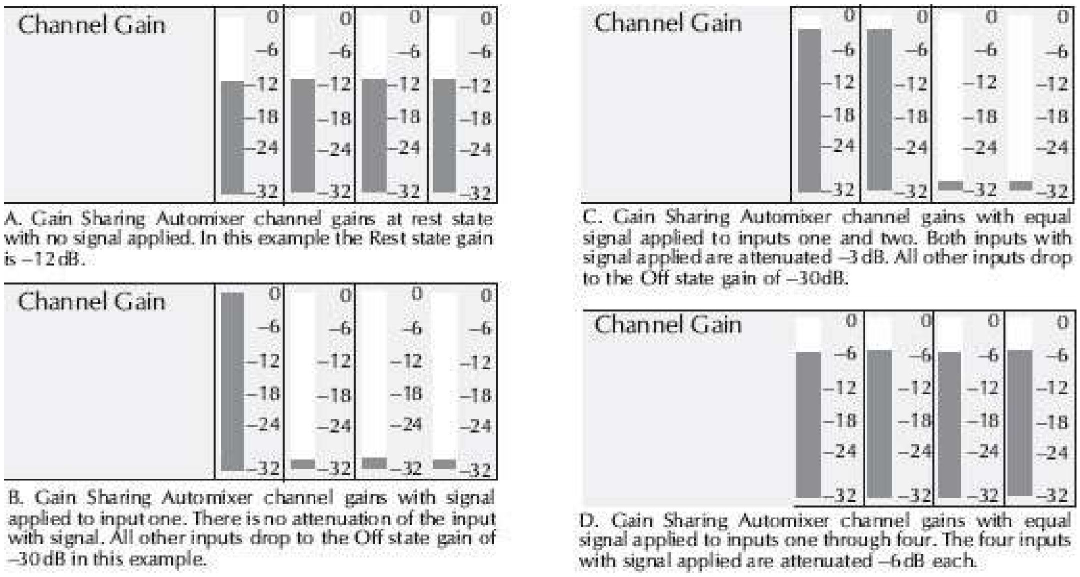


Figure 25-19. Block diagram of Biamp autoTwo Automatic Mixer. Courtesy Biamp Systems.



A. Gain Sharing Automixer channel gains at rest state with no signal applied. In this example the Rest state gain is –12 dB.

B. Gain Sharing Automixer channel gains with signal applied to input one. There is no attenuation of the input with signal. All other inputs drop to the Off state gain of –30 dB in this example.

C. Gain Sharing Automixer channel gains with equal signal applied to inputs one and two. Both inputs with signal applied are attenuated –3 dB. All other inputs drop to the Off state gain of –30dB.

D. Gain Sharing Automixer channel gains with equal signal applied to inputs one through four. The four inputs with signal applied are attenuated –6 dB each.

Figure 25-20. Gain sharing automixer channel gains.

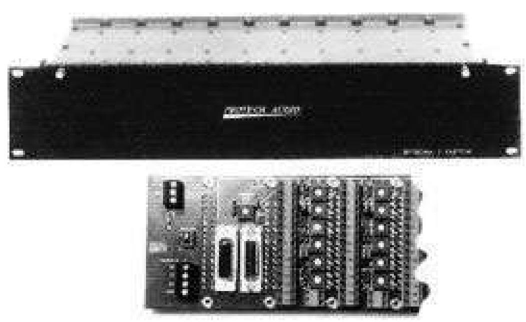


Figure 25-21. Protech Audio automatic mixer. Courtesy Dan Dugan.

25.3.4.4 Direction-Dependent Automatic Microphone Mixers

A direction dependent automatic microphone mixer responds to signals having acceptable levels within a predefined physical space in front of a microphone. By making the decision as to whether a channel should be on depends on the relative signal levels at two back-to-back cardioid microphone capsules in a single microphone housing. The Shure AMS automatic microphone system, now discontinued, responds in part to the location of the sound source. This mixer works only with its own unique two-capsule microphones.

When an AMS input channel is activated, the front facing microphone signal is transmitted to the mixer output. This mixer functions like a variable-threshold system with its threshold being a fixed level above the background ambient noise but with the threshold also being a function of the sound source location and its angular relationship to the microphone.

Any input channel may turn on when the signal level from the front microphone capsule is 9.5dB above the level from the rear capsule. Effectively, a signal to noise ratio of 5dB to 7dB is required for a channel to activate. Of course, a weaker sound source will not activate the channel. The level difference of 9.5 dB is derived from the criterion that a cardioid microphone response at 60° off-axis is typically one-third of its on-axis response. The activation angle of the mixer input channel is thus 120°. A sound source outside of the 120° angle will not activate an input channel no matter what the sound pressure level.

To keep the AMS microphones compatible with conventional shielded twisted pair cable while keeping the two microphone signals separated, an unbalanced signal path is used. This approach can be more susceptible to induced hum and noise pickup than a conventional balanced signal path. The use of current source preamplifiers in the microphone and unusually low impedance inputs in the mixer minimizes this potential problem.

It is recommended that an AMS microphone be installed within three feet of each talker, and the talker must be located within the

120° activation angle. Each AMS microphone should also be at least three feet from any wall behind it and at least one foot from objects behind it such as books, large ashtrays, or briefcases. This precaution is necessary to avoid unwanted reflection of the talker’s acoustic signal into the rear facing microphone capsule. Stray acoustic reflections can lead to unreliable input activation.

As the direction-dependent automatic microphone mixer process is covered under U.S. Patent 4,489,442, this type of automatic microphone mixer has been marketed only by Shure. In 2000, U.S. Patent 6,137,887 was issued to Shure for a new AMS design. Developed by Anderson, this patent adds a circuit that guarantees a single talker will activate only a single input channel, even if that talker is within the activation angles of multiple AMS microphones, Fig. 25-22.

Figure 25-22. Shure AMS8100 mixer. Courtesy Shure Incorporated.

25.3.4.5 Noise-Adaptive Threshold Automatic Microphone Mixers

This concept employs a dynamic threshold unique for each input channel. Using an inverse peak detector, each input channel sets its own minimal threshold that continually changes over several seconds based on variations in the microphone input signal.

Sound that is constant in frequency and amplitude, like a ventilation fan, will not activate an input but will add to the noise-adaptive threshold. Sound that is rapidly changing in frequency and amplitude, like speech, will activate an input. The mixer activates an input when two criteria are met:

1. The instantaneous input signal level from the talker is greater that the channel's noise-adaptive threshold.
2. The input channel has the maximum signal level for that talker.

Without this second criterion, a very loud talker might activate more than one input channel.

Note that this system deems any sound that is relatively constant in frequency and amplitude as nonspeech. Sustained musical notes may activate an input on attack, but after several seconds the sustained note will raise the threshold and the input will be attenuated. As previously stated, automatic microphone mixers are designed primarily for speech applications, not music.

Developed by Julstrom and covered by the U.S. Patent 4,658,425, the noise-adaptative threshold configuration has been used in analog and digital automatic microphone mixers manufactured by Shure, including the new SCM820 digital automixer, Fig. 25-23.

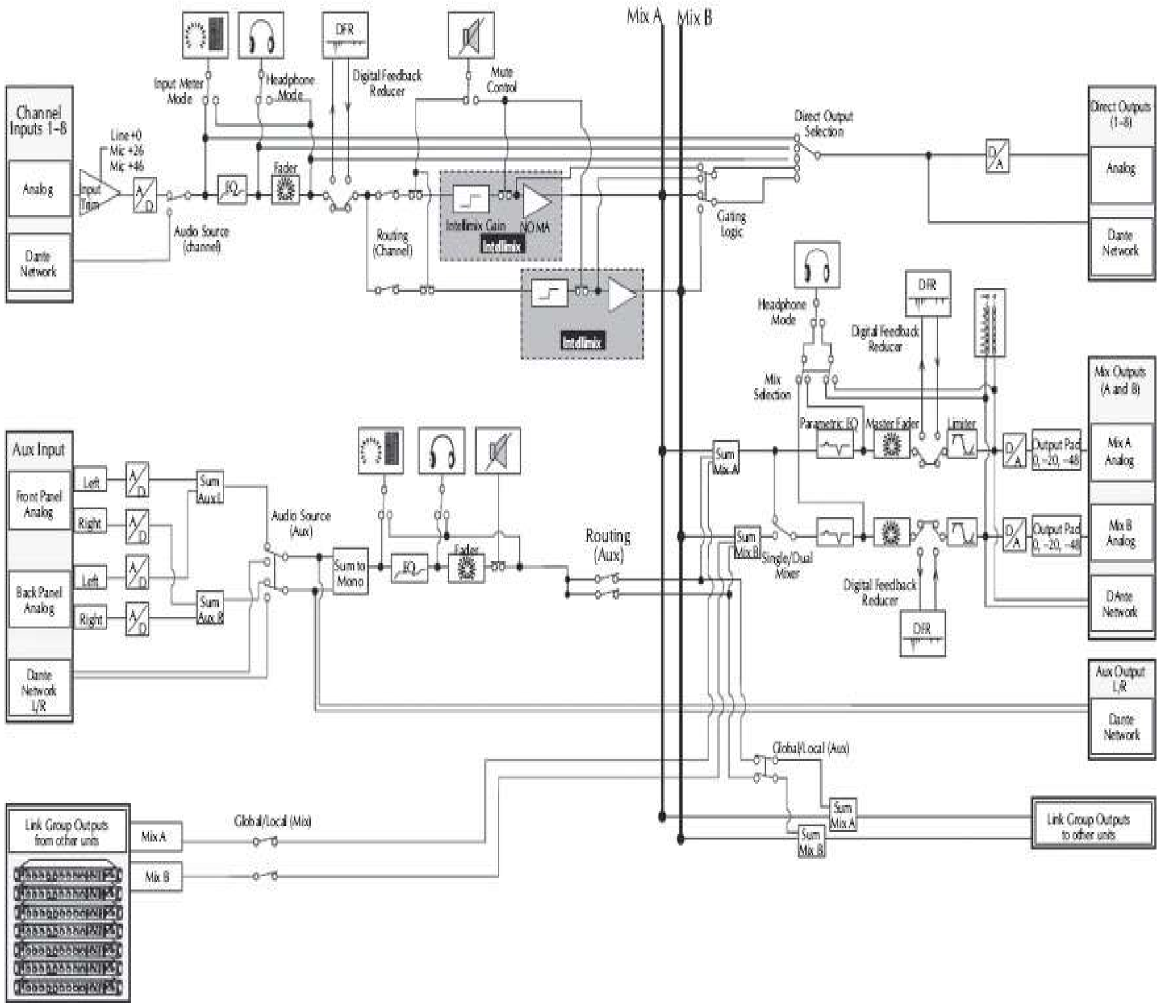


Figure 25-23. Block diagram of Shure SCM820 mixer. Courtesy Shure, Incorporated.

25.3.4.6 Multivariable-Dependent Automatic Microphone Mixers

The automatic microphone mixer methods described so far essentially use input signal amplitude as the activation variable. The relative timing of signals at each input is another variable that can be employed. A multivariable-dependent system makes its activation decision from both input signal amplitude and the time sequence of the input signals.

Peter's U.S. Patent 4,149,032 is such a design. The instantaneous positive signal amplitudes of all inputs are simultaneously compared to a threshold voltage (dc ramp) that falls 80dB in 10ms (or less) from a high value to a low value. Initially, all input channels are held in an attenuated state. The first input channel that has an instantaneous amplitude equal to the instantaneous value of the falling threshold is activated, while the other inputs remain attenuated. This activated channel remains so for 200ms.

Once an input channel is activated, the threshold voltage is reset to its high value and immediately starts to fall again in search of another input to activate. If all talkers are silent and an amplitude match is not found, the threshold search progresses the full 80dB in 10ms and then resets. However, this scenario is not typical. Most of the time, a signal on one of the inputs will produce a threshold amplitude match early in the search. In practice, the average input activation time is 3 or 4ms. Since the threshold resets every time an input is activated, the frequency of the threshold searches will be also every 3 or 4ms on average.

As mentioned, the input activation is maintained for 200ms. If on the second search the same input still has the largest signal amplitude, its activation status is renewed for another 200ms. If during a future threshold search, a different input channel has the higher amplitude, it is activated for 200ms. The first input activated times out and attenuates if not reactivated by a future search within the 200ms. As long as a talker keeps speaking, his input is continually renewed for 200ms intervals. This rapid response enables conversational dialog to be conducted and also permits easy activation of weaker sound sources during gaps in speech.

Since the activation gain of all input channels is the same, any signal source on an active channel has the same gain, and the relative levels of different talkers is preserved in the mixer output.

When multiple talkers vie for access to the system, the probability of all of them obtaining access decreases in proportion to the number. This effectively limits the maximum number of input channels that can be activated at any given time. For example, ten equally loud talkers will each be on 88% of the time. But as more than three or four persons talking at the same time is not intelligible, this limitation is normally of little consequence.

Also unique to the Peters design is the variable known as the *access ratio*. Simply put, the access ratio is the time an input is kept activated (200ms) compared to the decision time taken to activate an input (10ms). Access ratio may be readjusted to control the number of input channels that can activate at one time. Selective adjustment of the access ratio can also reduce missed beginnings of words. IRP Voice-Matic automatic microphone mixers used this design, and sometimes included their Level-Matic AGC on the output.

25.3.4.7 Automatic Microphone Mixers with Matrix Mixing to Multiple Outputs

Recent designs in automatic microphone mixers, and in particular those part of DSP based virtual Systems (see Chapter 41 *Virtual Systems*) have introduced matrix mixing to multiple outputs. This feature allows any input channel to be sent to any number of output channels and to be sent at different levels depending on the signal mix desired at the individual output. The Lectrosonics AM16/12 is a marriage of analog with digital. All control is accomplished via proprietary software that operates on a Windows-based computer. Software control allows a 16 in/12 out automatic microphone mixer with matrix mixing to fit in a two-rack space chassis. The software control also deters unauthorized readjustment as there are no knobs to twiddle. DSP based Virtual Systems such as the Peavey MediaMatrix and QSC Q-Sys offer the ability to configure several different styles of automixers with matrix mixer outputs.

Matrix mixing may be used for creating unique audio feeds for recording, teleconferencing, hearing assistance, language translation, etc. A courtroom is an example of a facility where all of these different audio systems might be required. Matrix mixing also provides the capability for providing an independent mix for the loudspeaker associated with each microphone. This allows a given microphone to be sent to all distant loudspeakers at full level, to nearby loudspeakers at reduced levels, with no signal at all sent to the nearest loudspeaker. This allows improved resistance to feedback and increased system stability while never muting a loudspeaker. Since loudspeakers are never muted, this allows the participants to clearly hear more distant talkers, even if they are talking themselves. Another way to look at this is matrix mixing with an independent output for each loudspeaker allows full duplex operation, as opposed to designs that mute loudspeakers for stability which provide half-duplex operation.

Some systems allow dynamic adjustment of the matrix mix, so that if a given talker is quiet and therefore the AGC has turned up his level, the spatial size of the zone where the loudspeakers receive reduced level from that microphone can be increased.

As matrix mixers get larger the number of controls and the analog circuitry or DSP resources required to implement the mixer get quite large. For example a system with 100 microphone inputs and 100 loudspeaker outputs requires a matrix mixer with 10,000 level controls. To reduce the number of controls and support circuitry required, the mix-minus approach was developed. Let's say that in our example each microphone would drive 90 loudspeakers at full level, and only drive 10 loudspeakers at reduced level. First one makes a mono mix of all the microphones, then for any given output the signal from the 10 nearest microphones is inverted in polarity and mixed with the mono sum. Thus only 11 inputs are required for each mixer output, which reduces the number of controls from 10,000 to 1,100. For the output nearest any given microphone the microphone's signal is inverted and mixed at full level with the mono sum. Since the sum already contains the signal from that microphone at full level, adding a polarity inverted copy of that signal will cancel the signal out in that output. For the other 9 inputs that are near that output somewhat less than the full inverted signal is mixed in, resulting in reduced level from those microphone inputs.

The biggest problem with implementing a mix-minus system in analog is ensuring the stability of all the gain controls in the system. Digital implementations solve the control stability issue but may add the issue of making sure that all the signals arrive at exactly the same time at the inputs of the mix-minus mixers.

With some modern DSP based Virtual Systems implementing huge matrix mixers may be easier than attempting a mix-minus approach.

The matrix mix concept improves gain before feedback. If a microphone signal does not appear in the closest loudspeaker, gain before feedback is better than if the microphone signal does appear in that loudspeaker. In a typical meeting room, talkers do not need to hear their own voice in the closest loudspeaker. They need to hear their other talkers located far away from their location. Matrix mixing or mix-minus provides this capability, Fig. 25-24.

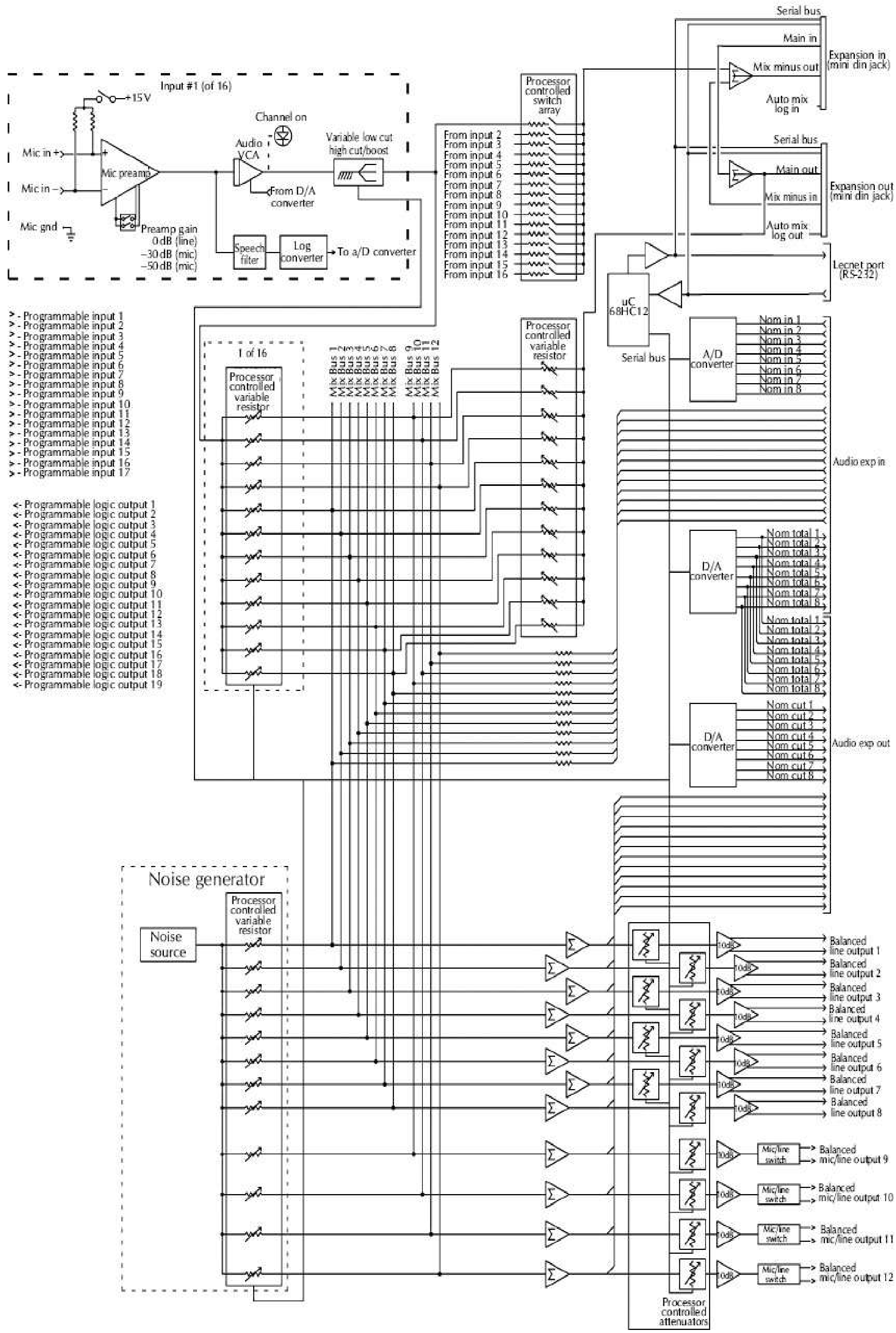

Input #1 (of 16)
+15V
Mic in +
Mic in –
Mic gnd
Mic preamp
Preamp gain
0 dB (line)
–30 dB (mic)
–50 dB (mic)
Channel on
Audio VCA
From D/A converter
Variable low cut high cut/boost
Speech filter
Log converter
To a/D converter
Processor controlled switch array
From input 2
From input 3
From input 4
From input 5
From input 6
From input 7
From input 8
From input 9
From input 10
From input 11
From input 12
From input 13
From input 14
From input 15
From input 16
Serial bus
Main in
Mix minus out
Auto mix log in
Expansion in (mini din jack)
Serial bus
Main out
Mix minus in
Auto mix log out
Expansion out (mini din jack)
Lecnet port (RS-232)
uC 68HC12
Serial bus
A/D converter
Nom in 1
Nom in 2
Nom in 3
Nom in 4
Nom in 5
Nom in 6
Nom in 7
Nom in 8
Audio exp in
D/A converter
Nom total 1
Nom total 2
Nom total 3
Nom total 4
Nom total 5
Nom total 6
Nom total 7
Nom total 8
Nom total 1
Nom total 2
Nom total 3
Nom total 4
Nom total 5
Nom total 6
Nom total 7
Nom total 8
D/A converter
Nom cut 1
Nom cut 2
Nom cut 3
Nom cut 4
Nom cut 5
Nom cut 6
Nom cut 7
Nom cut 8
Audio exp out
> - Programmable input 1
> - Programmable input 2
> - Programmable input 3
> - Programmable input 4
> - Programmable input 5
> - Programmable input 6
> - Programmable input 7
> - Programmable input 8
> - Programmable input 9
> - Programmable input 10
> - Programmable input 11
> - Programmable input 12
> - Programmable input 13
> - Programmable input 14
> - Programmable input 15
> - Programmable input 16
> - Programmable input 17
<- Programmable logic output 1
<- Programmable logic output 2
<- Programmable logic output 3
<- Programmable logic output 4
<- Programmable logic output 5
<- Programmable logic output 6
<- Programmable logic output 7
<- Programmable logic output 8
<- Programmable logic output 9
<- Programmable logic output 10
<- Programmable logic output 11
<- Programmable logic output 12
<- Programmable logic output 13
<- Programmable logic output 14
<- Programmable logic output 15
<- Programmable logic output 16
<- Programmable logic output 17
<- Programmable logic output 18
<- Programmable logic output 19
1 of 16
Processor controlled variable resistor
Mix Bus 1
Mix Bus 2
Mix Bus 3
Mix Bus 4
Mix Bus 5
Mix Bus 6
Mix Bus 7
Mix Bus 8
Mix Bus 9
Mix Bus 10
Mix Bus 11
Mix Bus 12
Processor controlled variable resistor
Noise generator
Noise source
Processor controlled variable resistor
Balanced line output 1
Balanced line output 2
Balanced line output 3
Balanced line output 4
Balanced line output 5
Balanced line output 6
Balanced line output 7
Balanced line output 8
Mic/line switch
Balanced mic/line output 9
Mic/line switch
Balanced mic/line output 10
Mic/line switch
Balanced mic/line output 11
Mic/line switch
Balanced mic/line output 12
Processor controlled attenuators

Figure 25-24. Block diagram of the Lectronics AM1612 mixer.

25.3.4.8 Automatic Mixing Controller

The Model E-1A Automatic Mixing Controller, Fig. 25-25, helps professional audio mixers handle multiple live microphones without having to continually ride their individual faders. This eight-channel signal processor patches into the input insert points of an audio mixing console. It detects which microphones are being used and makes fast, transparent cross-fades, freeing the mixer to focus on balance and sound quality instead of being chained to the faders. The Model E-1's voice-controlled crossfades track unscripted dialogue perfectly, eliminating cueing mistakes and late fade-ups while avoiding the choppy and distracting effects common to noise gates. Without the need for gating, a natural low-level room ambience is maintained.

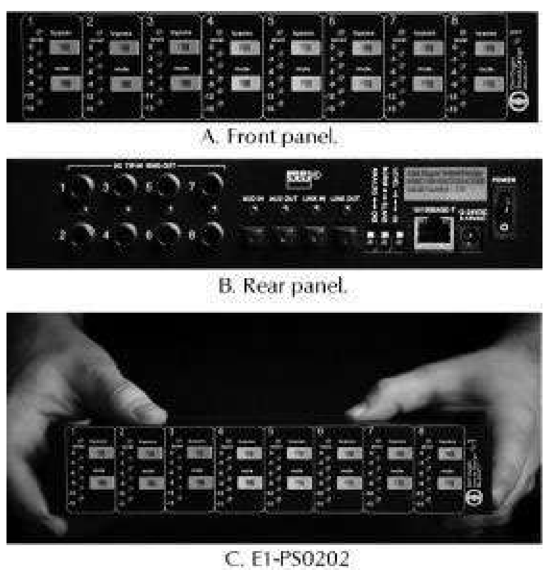


Figure 25-25. Dugan E-1A Automatic Mixing Controller. Courtesy Dan Dugan Sound Design.

Dugan automatic mixing controllers are used with multiple live mics and unscripted dialogue including talk shows, game shows, conference sound reinforcement, houses of worship, dramatic dialogue, wireless microphones in theaters, and teleconferencing. The Dugan controllers are typically connected in the insert points of the console's mic inputs, Fig. 25-26. Fig. 25-27 is the block diagram of the E-1A automatic mixing controller. Each unit handles up to eight channels, and the units can be linked together to accommodate a maximum of 128 microphone channels.

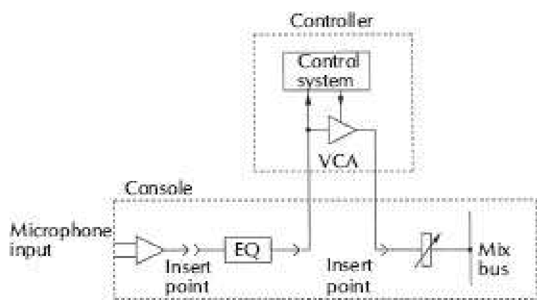


Figure 25-26. The automatic mixing controller is connected to the insert points of console input channels. Courtesy Dan Dugan Sound Design.

The Model E-1A is an eight-channel line-level or ADAT digital insert device in a half-rack, one unit high cabinet and has minimal controls. Additional controls are available via a virtual control panel provided by an embedded web server. I/O is connected by TRS insert cables or ADAT optical cables. The Model E-1A can be linked for up to 128 channels, and it can link with the Dugan Models E, E-1, E-1A, E-2, E-3, D-2 and D-3. Power is 9–24Vdc or 9–18Vac.

Four models are available. The Model E-1A has unbalanced analog I/O and ADAT optical digital I/O for use in the insert points of analog or digital mixing consoles. The Model E-2 has balanced analog I/O and ADAT optical digital I/O connections for insertion into analog or digital mixers. The model E-3 has AES3 digital I/O and ADAT optical digital I/O connections for insertion into digital mixers. The Dugan-MY16 is a 16 channel automatic mixer controller that plugs into a slot on Yamaha digital consoles.

25.3.4.9 Automatic Microphone Mixers Implemented in Software

If software can control automatic microphone mixer hardware, then automatic microphone mixers can also be completely created in software. This completely digital approach to automatic microphone mixers can be found in software based products offered by Allen & Heath, ASPI, BSS, Crown, Dan Dugan, Gentner, Lectrosonics, Peavey, QSC, Rane, and Shure. To date, the operational concepts used in digital automatic microphone mixers have not varied far from the previously described concepts underlying the analog automatic microphone mixers but the quality of implementation varies. The details of the implementation of automatic mixing concepts are often hidden deep within computer code, and manufacturers may be unwilling to reveal the details of how their automixers actually work. Future operational breakthroughs will likely be kept as closely guarded company secrets. New concepts in automatic mixing might only become public knowledge if patents are granted or technical papers are presented.

The Polycom Vortex EF2280, Fig. 25-28, automatically mixes microphones and other audio sources while canceling acoustic echoes and annoying background noise. It is used in boardrooms, courtrooms, distance learning, sound reinforcement, and room combining. It connects easily to other equipment including codecs, VCRs, or other A/V products. The unit can be programmed from the front panel, or through Conference Composer™ software (included). Conference Composer's Designer™ wizard ensures fast, accurate setup for a variety of applications.

A single Vortex EF2280 unit provides automatic mixing of up to eight microphones plus four auxiliary audio sources. Up to seven additional Vortex EF2280 or Vortex EF2241 units can be linked to the first unit. (NOM) information can be specified across all channels in the linked units. The microphone channels feature acoustic echo cancellation to prevent retransmission of signals to their original locations. A neural network AGC reacts only to valid speech patterns, bringing voices within desired levels. AGC controls are user adjustable, as are settings for the five-band parametric EQ offered on all input and output channels and output delay controls. Fig. 25-29 is the block diagram of the Vortex EF2280.

25.3.4.10 Which Type of Automatic Mixer Works Best?

There is no definitive answer to this question. It is impossible to tell which automatic microphone mixer design will operate best in a given situation by studying technical specifications, believing the marketing literature, poring over circuit schematics, deciphering lines of computer code, or rereading this chapter. Human speech is very complex and human hearing is very discerning. Like so many areas in professional audio, the critical ear is the final judge.

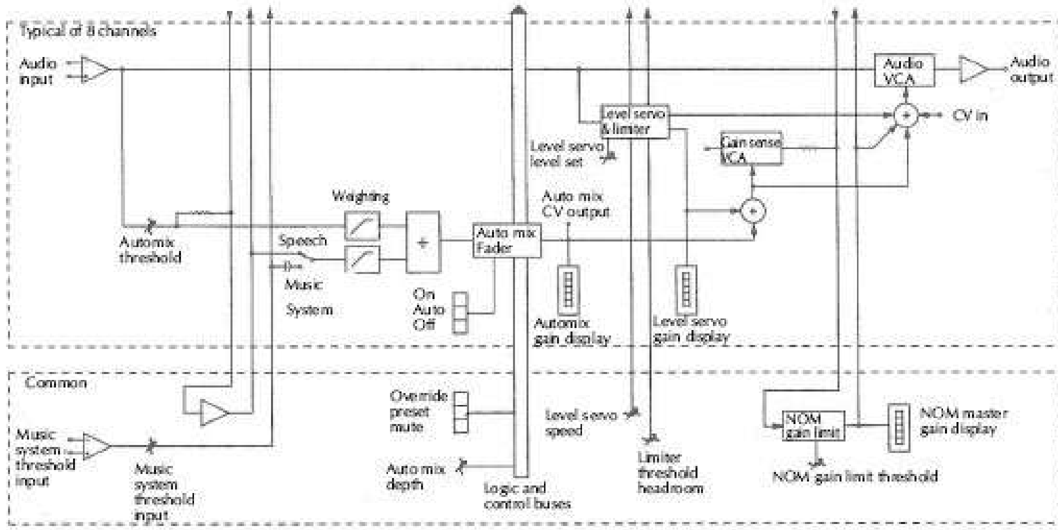


Figure 25-27. Block diagram of the E-1 automatic mixing controller. Courtesy Dan Dugan Sound Design.

Figure 25-28. Vortex EF2280 digital multichannel acoustic echo and noise canceller with a built-in automatic microphone/matrix mixer. Courtesy Polycom, Inc.

25.3.5 Teleconferencing and Automatic Microphone Mixers

Automatic microphone mixers are used in many teleconferencing systems. The design of such systems involves a number of complex issues that do not enter into the design of sound reinforcement systems. This section will discuss important design aspects of such installations.

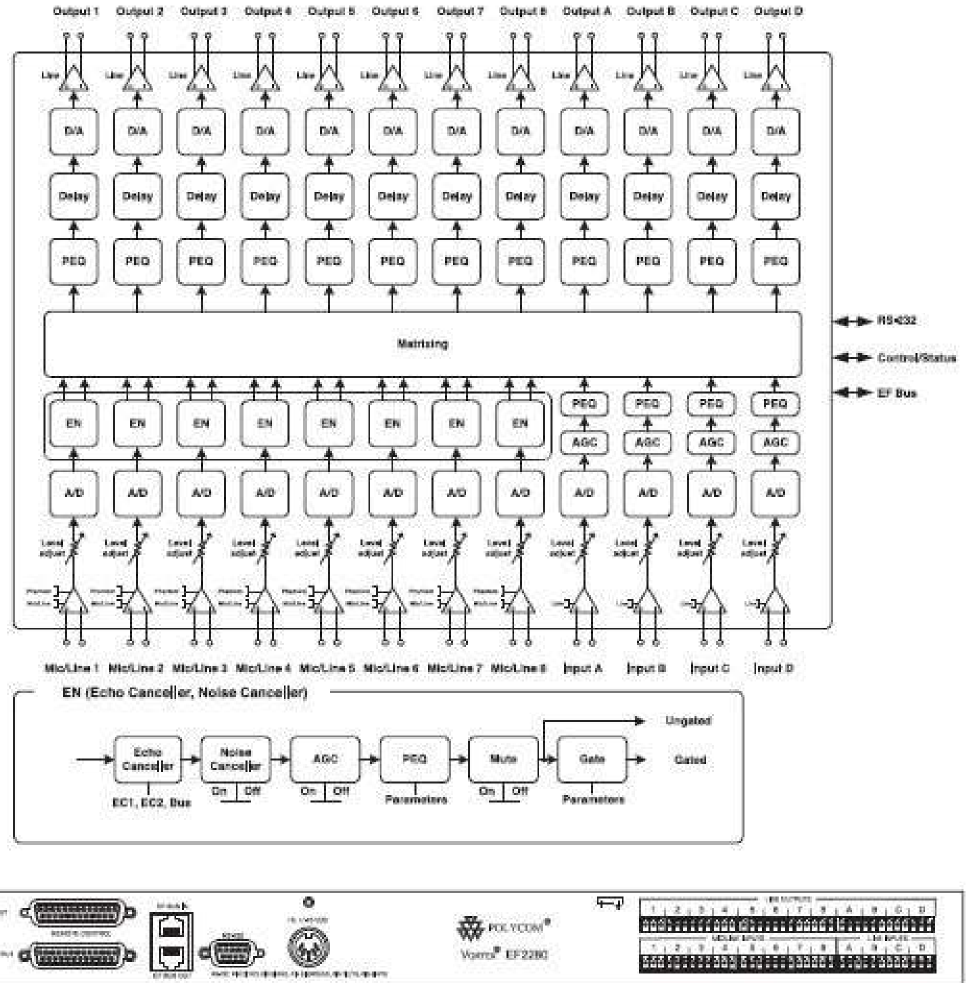


Figure 25-29. Block diagram of the Vortex EF2280 digital multichannel acoustic echo and noise canceller with a built-in automatic microphone/matrix mixer. Courtesy Polycom, Inc.

As practiced in modern communication between separated groups of talkers, teleconferencing has two components—visual and aural. The visual is handled by television cameras, video monitors, and video projectors. The visual may be full motion in real time, slow scan, or single-frame presentation.

It is appropriate to identify the aural part of the teleconferencing system as the audio conferencing system. Considerable attention must be paid to a number of details for acceptable sound quality, intelligibility, and user comfort. Users of teleconferencing systems tend to employ very subjective descriptions that have to be interpreted into quantitative engineering terms that can then be applied, measured, and included in system designs. Teleconference participants expect good speech intelligibility, easy identification of the talker, relatively high *SNR*, and other qualities. They also expect the overall aural experience to be better than a conversation conducted via telephone handsets.

An audio conferencing installation for voice and program has four primary facets:

1. Conference room and building acoustics.
2. Interface with telephone/transmission system.
3. Possible secondary use as a sound reinforcement system.
4. Proper equipment selection and setup.

25.3.6 Room and Building Acoustics

25.3.6.1 Conference Room Noise

The first consideration for a teleconference installation is noise in the room. Obvious noise sources, like heating and air conditioning systems, should be evaluated and specified for acceptable levels. External noise must also be considered:

- Conversations in hallways or adjacent offices.
- Business machines in adjacent spaces.
- Elevators on opposite sides of the wall.
- Water flow in building services.
- Vibration of air conditioners on the roof.
- Loading docks

There will also be unwanted noise generated in the conferencing room itself:

- Fans in projectors and computers.
- Hum from light fixtures.
- Paper shuffling.
- Moving chairs.
- Coughing.
- Side conversations.

All of these undesired sound sources are **much** more obvious, annoying, and detrimental to intelligibility at the remote site of the teleconference than they are in the local site where they originate. Also, as the number of participants increases, the geographic area covered by the participants expands and unamplified speech becomes harder to hear due to greater distances between the talkers and the listeners. Consequently, for comfortable talking and listening, the ambient noise level in a room must be lower for larger groups.

Table 25-1 provides recommended noise level limits for conference rooms. These are levels at the conference table with the room in normal unoccupied operation and at least 2 feet from any surface. Methods for achieving low interfering noise levels are discussed in Chapter 6 *Small Room Acoustics*.

Table 25-1. Ambient Noise Level Limits for Conference Rooms

Conference Size	Maximum Sound Level in dBA	Preferred NC	Acoustic Environment
50 people	35	20–30	Very quiet, suitable for large conferences at 20–

			30ft table.
20 people	40	25–35	Quiet, satisfactory for conferences at a 15 ft table.
10 people	45	30–40	Satisfactory for conferences at 6–8ft table.
6 people	50	35–45	Satisfactory for conferences at 4–5ft table.

More accurate assessment will result if noise criteria (NC) are used because of the strong influence of frequency spectrum on speech interference and listener annoyance, see Chapter 8 *Acoustical Treatments for Indoor Areas*.

Fig. 25-30 shows the maximum microphone/talker distance for a marginally acceptable *SNR* of 20 dB in transmitted speech based on the A-weighted noise at the microphone location. Often microphones are placed at locations very close to noise sources such as air registers in the ceiling or laptop fans on the table. This can result in noise level at the microphone being considerably higher than the average level in the room. The graph applies to omnidirectional microphones. The distance may be increased by 50% for directional microphones if they are aimed at the talker. If more than one microphone is active in the system, the number of open microphones must be taken into account by reducing the predicted *SNR* by 3 dB for each time the number of open microphones doubles. An automatic microphone mixer will alleviate this concern.

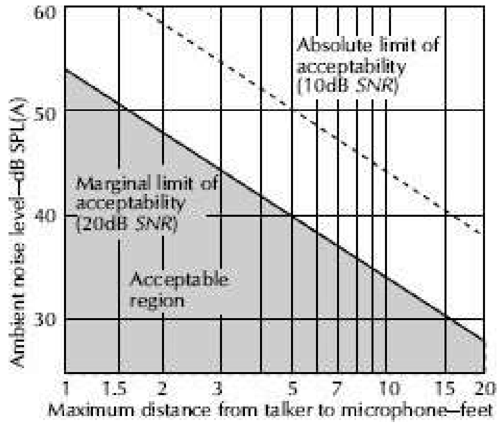


Figure 25-30. Acceptable ambient noise levels.

If an acoustical survey indicates the presence of interfering noise sources, construction techniques must be implemented to provide adequate sound transmission losses, or another room should be considered.

25.3.6.2 Conference Room Acoustic Reflections

Acoustic reflections from the far end of an audio teleconference are made much more obvious and annoying by the propagation delay across the link between rooms, and are often identified by conference participants as the "speaking into a barrel" effect. Some improperly call these acoustic reflections reverberation, but a small room such as typically used for audio teleconferencing almost never supports a true statistical reverberant field. The sources of acoustic reflections are many. For example, there are reflections from hard surfaces in the room where speech is originating at the moment. Requirements for comfortable talking in the room dictate some preferably diffused acoustic reflections, as a bit of acoustic liveliness in a meeting room is desirable. If the microphones are not located close to the mouths of the talkers and an automatic microphone mixer is not employed to reduce the reverberation picked up by unused microphones, the acoustic reflections as heard at the remote site of the teleconference can be deemed excessive and intolerable. In this situation, a very low (and uncomfortable) amount of acoustic reflections at the local site will be required. The potential for reduced intelligibility at the remote site is increased because the remote participants do not have the advantage of separating the speech signal from the acoustic reflections via binaural hearing, plus not having the talker in the same room also tends to dull one’s attention. Listeners at the remote site need about 6dB greater separation between the speech level and the level of the noise and acoustic reflections in order to experience similar intelligibility as listeners who are hearing the speech live.

There are also the acoustic reflections added at the remote site. The incoming signal is reproduced by loudspeakers, the sound propagates around the room, and even more acoustic reflections are overlaid on those coming from the far end. So, unless there is only a telephone handset at the remote site, both sites need to have proper acoustical characteristics. This is often not the case as in many conferencing rooms the visual comfort of the room takes precedence over the aural comfort. Just ask the interior designer!

Room dimensions, angles, and surface treatments should be chosen to minimize standing waves and flutter echoes. If the room already exists, judicious use of acoustically absorbent and diffusive materials are advisable for control of the room’s acoustics.

Direct to reflected ratio (D/R) is often used to predict appropriate talker to microphone distances. D/R =1 is where the direct signal of the talker is equal to reflected signal of the talker, i.e., 50% direct signal and 50% reflected signal. The distance for D/R = 1 in conference rooms is typically in the range of 1–4ft.

For good intelligibility, an omnidirectional microphone should be placed at ½ of the distance where D/R =1 or less from the talker. When a directional microphone is used, the distance between talker and microphone may be increased by up to 50%.

Because the sound decay in the first 60–100ms usually is the most damaging to teleconference conversations, the usual measure of acoustic decay which is the time required for a 60dB reduction in the level of the acoustic reflections, may not be the most appropriate. One manufacturer of conference equipment insists that the room produce a decay of greater than 16dB in the first 60ms. One reason for this has to do with the nature of acoustic echo cancellers (AEC) used with teleconferencing systems. AECs reduce both the direct sound from the loudspeakers into the microphones, and acoustic reflections of that sound from the room, from the outgoing signal sent to the remote site. They only do this for a limited number of ms, after which they provide no reduction of later arriving acoustic reflections. Therefore the level of the later arriving acoustic reflections must be controlled by acoustic means.

The one admonition to anyone faced with the design of a teleconferencing system is *do not ignore the acoustical characteristics of the room*. Insist upon a room that has the right acoustical environment or commit the resources to make it right before proceeding with the rest of the project. Acoustical deficiencies can rarely be corrected by electronic means. If it is new construction, work closely with the architect before the room design is complete.

25.3.6.3 Telephone/Transmission System Interface

Fig. 25-31 shows two teleconference rooms connected by a single two-wire analog plain old telephone service (POTS) telephone line. Each room has a microphone and a loudspeaker with associated amplification. A *hybrid* interface between the send and receive lines and the telephone line serves to reduce loop gain within the room by reducing *sidetone leakage*.

Possible feedback loops are shown. Not only is there potential oscillation in the sending room, but also the coupling through the line to the receiving room and back is an equally probable feedback loop. Basic speakerphones use voice-activated gates to capture the line and permit transmission in only one direction at a time and thus interrupt the feedback path from the remote site. This can cause frequent dropouts in a conversation and forces the communication into a *half-duplex* mode of operation. Half duplex transmission refers to transmission in only one direction at a time.

The ideal is *full duplex*, which allows transmission in both directions all the time. A phone call from telephone handset to telephone handset provides full duplex communication. Full duplex is preferred for audio conferencing because there are no missing words or sentences, and conversations can be conducted in a normal manner. Control of reverberation and room noise is essential in any full duplex system.

An alternative connection system uses four wires as shown in Fig. 25-32. One pair of wires is used for each direction of transmission, thus eliminating the often troublesome hybrid sidetone leakage. As can be seen, there is still the possibility of feedback through the room at either end. However, there is usually cleaner signal transmission with the added expense of a second telephone line. Four-wire systems make full duplex communication easier by removing the limitations of the hybrids required for two-wire phone circuits. Digital connections such as ISDN and VOIP are the equivalent of a four wire circuit and also allow full duplex communication.

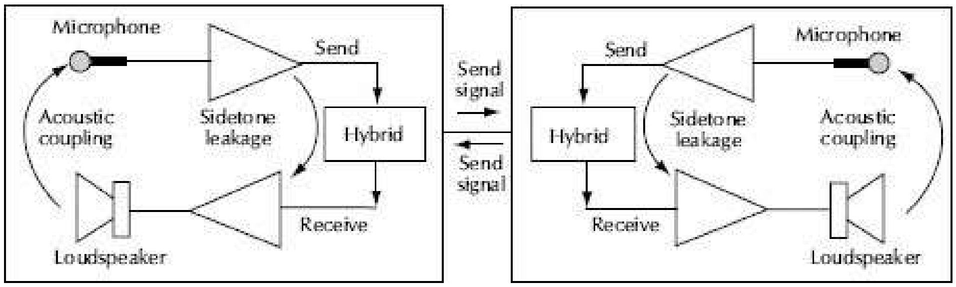


Figure 25-31. Teleconference system with two-wire telephone connection, showing feedback paths.

Frequently audio conferences involve several sites giving rise to point-to-multipoint or multipoint-to-multipoint telephone interconnections. A *conference bridge* is used to connect a number of telephone lines so that all participants will be tied together. Bridging over 20 phone lines is now quite common. The actual bridging may be provided by an external bridging service company or bridging devices may be part of the on-site teleconferencing equipment.

A conference bridge at its simplest is a matrix mixer or mix minus circuit setup so the voice coming in a port from a phone line is sent out all ports except the one it came in on. Such a basic conference bridge can be improved on by passing all the incoming voice signals through an automatic mixer before the mix-minus circuitry.

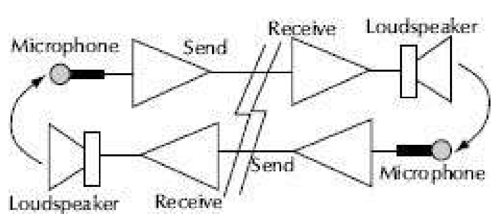


Figure 25-32. Teleconference system with a four-wire telephone connection.

A simple gated conference bridge may limit the number of open ports to two because signal leakage at the far ends can cause retransmission of received audio on telephone lines. As a result, only one two-way conversation can occur and others can only listen. Also, the uncertain and variable quality of analog telephone connections can result in having a noisy line tying up the system and preventing access since the bridging control depends on signal-activated switching. The combination of modern digital phone lines and more advanced conference bridges has reduced these problems and can provide a much more natural user experience.

25.3.7 Teleconferencing Equipment

25.3.7.1 Telephone Interface

The telephone interface for a typical two-wire analog phone line is the hybrid. It converts the two-wire transmission of the connecting lines to internal four-wire paths to isolate the send and receive signals. A hybrid passes the microphone send signal (two of the four wires within the room) to the two-wire telephone line but attenuates it to the receive line. Conversely, a signal being received from the telephone line passes to the receive line (the other two of the four wires in the room) and is attenuated to the microphone send line. For many years, the hybrid in a standard telephone set was a transformer; now electronic equivalents are common.

Simple hybrids such as those in a standard telephone are designed for an assumed phone line impedance of around 600 to

900 Ω Because they are not matched exactly to the phone line impedance the send to receive isolation provided is rarely greater than 10dB. Hybrids that allowed manual tuning to the phone line could improve that performance but required retuning each time a call was placed. Studer developed a self-tuning hybrid for broadcast and teleconference applications that eliminated the need for manual tuning. Many phone lines present an impedance that varied considerably with frequency, which limits the performance of even the best analog hybrid. Steve Church of Telos applied DSP echo cancellation technology to the hybrid with the Telos 10 which was the first modern digital hybrid. The Telos 100 improved on the performance reducing send to receive isolation to 50dB or more even on difficult two wire connections.

Good isolation through the hybrid is necessary. Unless these devices can adapt to variable telephone line conditions, signal leakage may be retransmitted through them. If the boardroom has not been correctly treated acoustically, the combination of room echo and signal leakage creates an undesired feedback path. Many hybrids suppress leakage by less than 15 dB whereas 35 to 40dB is regarded as the minimum acceptable for loudspeaker receive conference installations. The paths of signal leakage are shown in Fig. 25-33. DSP based hybrids are supplied by manufacturers such as Gentner Electronics, ASPI, and Telos. These products provide means for optimizing the impedance match to the telephone line, thereby giving additional suppression of signal leakage. DSP hybrids can make the difference between a marginal and an acceptable

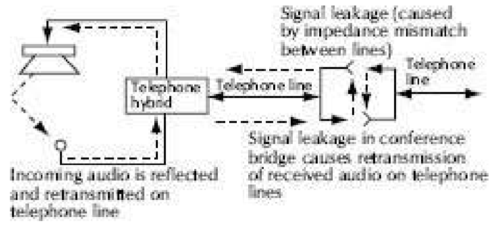


Figure 25-33. The paths for signal leakage and undesired feedback in a typical teleconference system.

Typical telephone line impedances range from 600 Ω to 900 Ω Telephone equipment expects send levels of 0dBm. The receive level standard is –6dBm, but these levels are reported to vary widely, – 10dBm is frequently experienced. The standard telephone line has 48 Vdc (some private exchanges use 24Vdc) for system control that must be blocked with a transformer or capacitors. The dc current through the off-hook relay keeps the line open while the connection is active.

25.3.7.2 Microphone Considerations

For a small group in a conference room, it may be possible to use only one omnidirectional microphone on a table top, typically of the surface-mount type. However, even for a group of four to six people, the equivalent of several directional microphones with an automatic microphone mixer is preferred to reduce the number of open microphones to the minimum necessary for the discussion. Three cardioid microphones in a circle, spaced at 120° intervals, is a typical approach. There are a number of surface-mount microphones that can be used, provided that the distance to the talkers is acceptably short. The typical participant in a teleconference expects, at minimum, the sound quality heard from a handset where the microphone is within inches of the talker’s mouth. Thus, keeping the microphones close to the talkers is very important. One excellent solution is the use of gooseneck microphones to get the microphone element within a few inches of the talker's mouth. If this is not possible then a steerable array microphone from TOA will allow a pickup beam to be aimed at the talker from a greater distance. The microphone tracks the position of the talker to keep the narrow pickup beam correctly aimed.

Keeping the microphone away from noise sources is critical. Ceiling mounted microphones are often near air registers, lighting fixtures, and projectors which make noise. Microphones on the surface of the conference table may have to contend with noise and wind from laptop fans, and are often covered by papers on the table.

25.3.7.3 Microphone Mixing

Larger groups inevitably require a large number of microphones to keep the participant-to-microphone distance within the limits set by the room's D/R = 1 distance. Some form of automatic microphone selection and mixing is essential in this case. Systems can be designed using an automatic microphone mixer (as described earlier in this chapter) connected to a telephone line interface device. Or a system can be implemented using an integrated device where the automatic mixing and the telephone interface are contained in the same chassis. Depending on the complexity required, there are many suitable approaches to system design. Acoustic Echo Cancellation (AEC) which is often used to improve the performance of teleconferencing systems depends on developing a model of the echoes in the room. If a single AEC is used with an automatic mixer the echo pattern changes and microphones are turned on and off. This requires the AEC to constantly adapt to the changing echo pattern. While it is more expensive, providing a separate AEC for each microphone eliminates the need for constant adaptation and can result in better system performance. Consult the equipment manufacturers for specific design suggestions, Fig. 25-34.

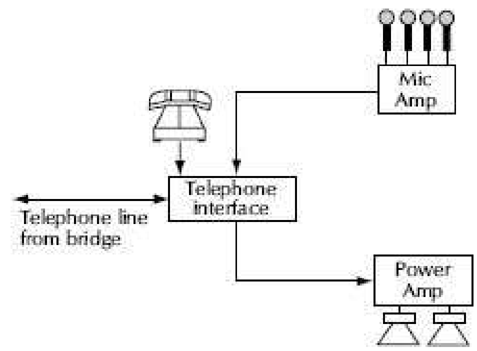


Figure 25-34. The configuration of a multimicrophone audio conference installation, without sound reinforcement.

25.3.7.4 Loudspeaker Considerations

Direct feedback from loudspeakers to microphones in any of the conference sites must be avoided; therefore, loudspeaker placement is critical. Loudspeakers should be placed in the null of the microphone pickup patterns. For cardioid microphones pointing in a horizontal direction, loudspeakers can be placed behind the microphones and aimed upward. Never place loudspeakers in front of microphones as microphones cannot distinguish between talkers in the room (desired sound sources) and talkers heard via loudspeakers (undesired sound sources).

When there is talking in the room, automatic microphone mixers can reduce the level of the loudspeaker signal from the remote site. This is accomplished via attenuating relays, ducking circuits, etc. A better solution is to reduce the acoustic coupling from the loudspeakers into the microphones by placing each microphone in the null between two loudspeakers that carry the same signal but with opposite polarities. Brigham Young University developed this idea and on June 27, 1969 shared the idea with Electro-Voice. In September 1969 Electro-Voice published the idea in their Sound Technique newsletter. Sound Control Technologies offers a system using this concept. The loudspeaker contribution to the send line is claimed to be reduced by 40dB with this arrangement.

If sound reinforcement of conversations within the room (sometimes known as voice lift) must also be provided in addition to audio conferencing, even more attention must be given to reducing the audio coupling between the loudspeakers and the microphones. Such systems can be very difficult to design correctly and must be approached with great caution. The use of an experienced acoustical/audio consultant is highly recommended in these cases.

25.3.7.5 Send Level Control

Send level, i.e., the audio signal voltage supplied to the telephone line, should be within acceptable ranges. Compressors, AGCs and levelers are all devices to consider for this technical requirement.

25.3.7.6 Echo Canceler

Echo cancelers reduce residual echo return in audio conferencing installations. If the local site returns significant signal from its incoming port to its outgoing port, and there is significant propagation delay due to the transmission line, the remote site will hear an annoying echo when someone in the remote site speaks.

The imperfect balancing of hybrids is one path for echo. Signal reflection within the telephone line is another source of echo. Echo also occurs acoustically when loudspeaker sound reaches open (active) microphones that are transmitting speech. The use of satellite transmission links also makes echo problems worse because of the long propagation delays.

A line echo canceller attempts to reduce echoes that are electronic in nature, such as those caused by hybrid leakage. An acoustic echo canceller (AEC) looks at the signal coming into a room in comparison with the signal leaving the room and develops a model of the direct sound and acoustic reflections. It then inserts time-delayed mirror image of the incoming signal as processed through the model into the outgoing signal leaving the room. The idea is to cancel any of the incoming signal and it's reflections that leaks into the outgoing signal path as a result of the acoustical coupling between loudspeaker and microphone.

Echo canceller technology has rapidly advanced due to faster CPU speeds and new research into canceler algorithms. Early echo cancelers were very expensive and thus having a single canceler at each conferencing site was considered adequate. As the price of echo cancelers has declined, manufacturers such a Gentner and ASPI now offer devices that have an echo canceler for each microphone input channel.

25.3.7.7 Historical Examples of Teleconferencing Equipment

Two historical systems will be described in more detail in order to show the number of parameters that must be considered in addition to the usual sound reinforcement needs. The first is an automatic microphone mixer approach as exemplified by the Shure ST3000, first manufactured in the 1980s. The second is the Sound Control Technologies system that does not use automatic microphone mixing.

Shure ST3000—An Analog Speakerphone. A simplified block diagram of the ST3000 is shown in Fig. 25-35. A conference call connection is made by taking the telephone handset from its cradle and dialing the desired number. When it is determined there is a good connection with the dialed party, the controller conference switch can be depressed to turn on the conference system. Green talk LEDs turn on and the handset may be returned to its cradle. The controller loudspeaker volume may next be adjusted if necessary. Levels for any auxiliary equipment may also be adjusted. Use the mute switches to prevent the called party from hearing local conversation. Red LEDs indicate muted status. The conference is terminated by depressing the controller telephone switch for at least 1 second.

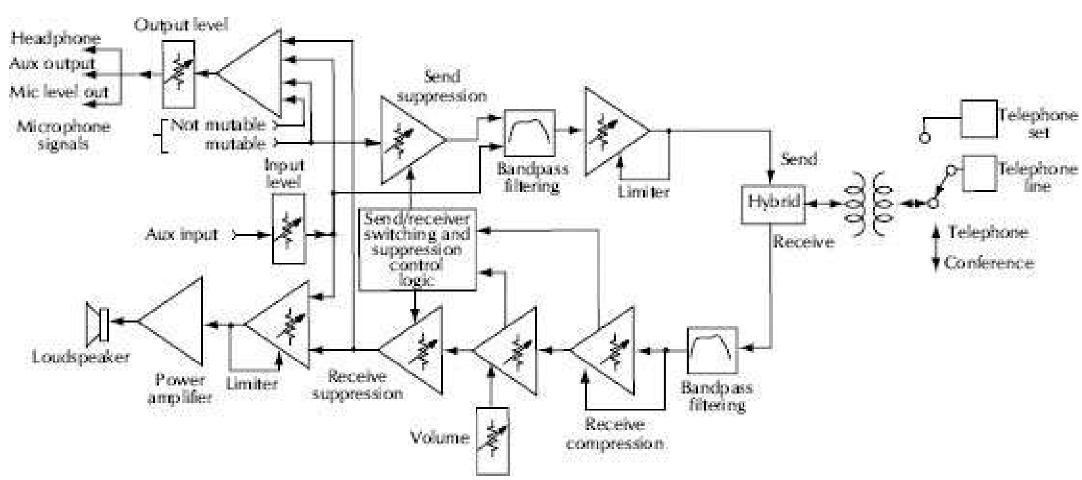


Figure 25-35. Block diagram of an analog teleconferencing system using automatic microphone control techniques. Courtesy Shure Incorporated.

In Fig. 25-35, the upper left mixer amplifier feeds the various auxiliary outputs. Below this amplifier, the conference microphone inputs are shown. Only the mutable (i.e., for automatic mixing) microphone inputs feed the send signal path to the hybrid. The receive path leads to the power amplifier and loudspeaker. Relative send and receive signals in the room are controlled by the send/receive switching and suppression logic. The suppression logic causes either the send amplifier or the receive amplifier to attenuate its signal depending on the presence of a receive signal. Because standard voice quality telephone lines have restricted bandwidth requirements, bandpass filtering is included in the send channel. Bandpass filtering in the receive channel reduces the possibility of extraneous noise from the telephone line.

In the early 1990s, digital technology replaced analog devices such as the Shure ST3000. Polycom is now one of the prime suppliers of digital, full duplex speakerphones.

Sound Control Technologies Ceiling Systems. Two configurations have been supplied by Sound Control Technologies. Loudspeakers and microphones are mounted in the ceiling over the conference participants. In one configuration, two loudspeakers are driven in antiphase (180% out of polarity) and a small microphone is mounted midway between them. Direct sound from the loudspeaker to the microphone is balanced for a null of 20dB for receive signals. The basic element is shown in Fig. 25-36.

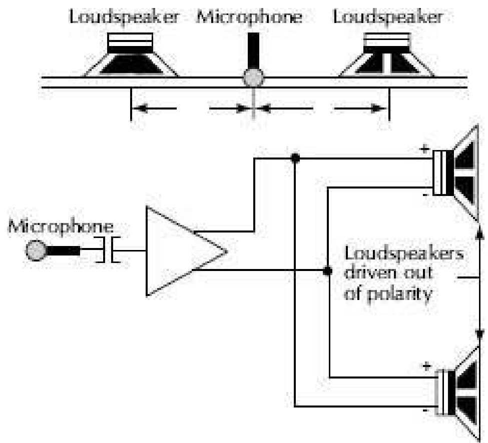


Figure 25-36. A method of using acoustic cancellation to reduce acoustic leakage in a conference room by driving loudspeakers out of polarity.

The second configuration uses a microphone and a loudspeaker mounted precisely 12 inches apart in reflecting baffle ceiling-mounted panels. Pairs of these loudspeaker/microphone units are placed above the conference table. All loudspeakers are driven in the same phase, while the microphones of symmetrically located units are mixed and balanced antiphase. A block diagram of such a system is shown in Fig. 25-37.

The microphone signals being mixed and balanced in antiphase feed the bus from which both the sound reinforcement (voice lift) and telephone send signals are derived. Notch filters are used for adjustment of spectrum balance. Delay may be included in the reinforced sound feeds if the room is large. The telephone return signal also feeds the sound reinforcement loudspeakers. An echo canceler is included to reduce the effects of telephone line echo or room acoustic echo.

As with the Shure system described previously, a telephone connection is made with a handset. Upon completion of the connection, the status of the line is determined by transmission of a group of tone bursts that allows the hybrid to electronically balance for the complex impedance of the telephone line. A pushbutton switch converts the connection to conference and the handset may be placed in its cradle.

25.3.7.8 The Present and Future of Teleconferencing

Most basic teleconference systems are sophisticated speakerphones with full duplex capability. Mid-level teleconferencing systems employ automatic microphones mixers and digital hybrids. The most sophisticated systems feature integrated teleconferencing devices that include multiple inputs with automatic mixing and echo cancellation, mix-minus signal routing capability, real-time feedback and level control, and operation via touchscreen.

Personal computers and digital signal processing (DSP) are becoming the dominant technologies that drive new developments in teleconferencing.

DSP advances are leading to teleconferencing systems that provide each participant with a customized electro-acoustical environment, unique to his or her own talking and hearing requirements. Advances in background noise reductions via electronic means are already impressive, as long as the noise has a repetitive nature. Microphone arrays that can be steered to best pick up a talker and steerable loudspeaker arrays are more prevalent.

But no matter how dominant digital technology becomes in teleconferencing, the speech input to the system from the human mouth will be analog, and the acoustical output to the human ear will also be analog. And that is the only technology forecast that will be 100% accurate.

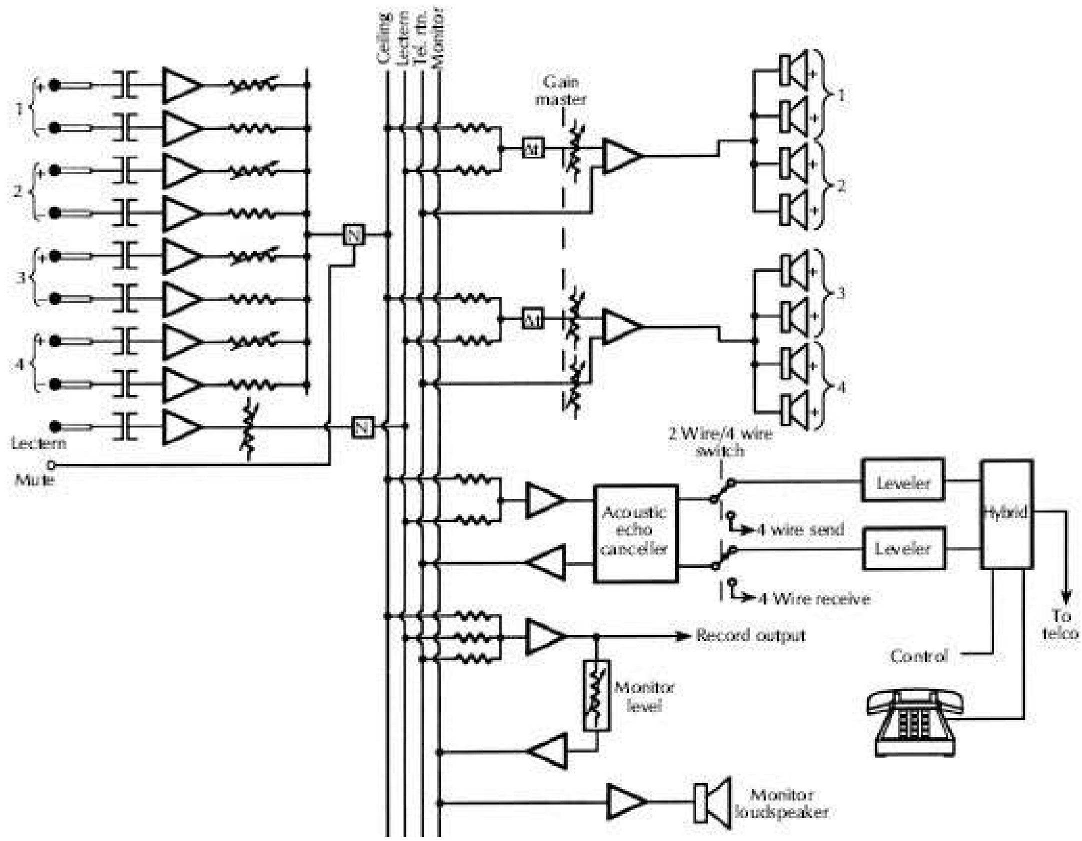


Figure 25-37. Schematic of an teleconference system in a board room that uses all loudspeakers in phase and pairs of microphones in antiphase. Courtesy of Sound Control Technologies, Inc.

References

1. C. Motchenbacher and F. Fitchen, *Low-Noise Electronic Design*, New York: John Wiley & Sons, 1973, p. 10.

2. Ibid, pp. 34–35.

3. D. Jensen, *High Frequency Phase Specifications—Useful or Misleading?* AES 81st Convention Preprint 2398, 1986, pp. 5–9.

4. W. Leach, Jr., "The Differential Time-Delay Distortion and Differential Phase-Shift Distortion as Measures of Phase Linearity," *AES Journal*, Vol. 37, no. 9, 1989, pp. 712–714.

5. D. Jensen and G. Sokolich, *Spectral Contamination Measurement*, AES 85th Convention Preprint 2725, 1988.

Patent References

D. Dugan, U.S. Patent 3,814,856; June 1974.
D. Dugan, U.S. Patent 3,992,584; November 1976.
R. Peters, U.S. Patent 4,149,032; April 1979.
R. Ponto, U.S. Patent 4,239,938; December 1980.
R. Ponto, H. Martin, U.S. Patent 4,374,300; February 1983.
C. Anderson, W. Bevan, R. Schulein, A. Smith, U.S. Patent 4,489,442; December 1984.
S. Julstrom, U.S. Patent 4,658,425; April 1987.
S. Sakamoto, U.S. Patent 5,148,491; September 1992.
T. Sims, Jr., U.S. Patent 5,402,500.
T. Sims, Jr., U.S. Patent 5,414,776.
Anderson, M., U.S. Patent 6,137,887; October 2000.

Bibliography

F. Ampel, “A Survey of Automatic Mixers,” *Sound & Video Contractor;* October 1983. Bell Labs, “Acoustics of Teleconference Rooms,” Publication 42901, August 1977.

C. Stockbridge, “Acoustics of Teleconferencing,” *Journal of the Acoustical Society of America*, vol. 74, no. 81, p. 85.

M. Pettersen, “Automatic Microphone System Updates,” Shure Inc.; 34 technical newsletters; 1983–1995.

R. Baum, “Automatic Mic Mixers,” *Sound & Communications;* series in four parts: July 1993, September 1993, November 1993, March 1994.

M. Pettersen “Broadcast Applications for Voice Activated Microphones,” National Association of Broadcasters Conference Proceedings; May 1989.

M. Pocino, “Conference Audio Systems Design,” *Sound & Video Contractor*, March 2001.

S. Wooley, “Echo Cancellation Explained,” *Sound & Communications*, January 2000.

M. Burkhard, “Preamplifiers and Mixers,” *Handbook for Sound Engineers;* 2nd edition; Howard W. Sams; 1991.

S. Julstrom and M. Pettersen, “Room Acoustics and Their Effect on Microphone Placement,” Invited paper, Acoustical Society of America, 118th meeting, November 1989.

D. Dugan, “The Theory Behind Automatic Mixers,” *Sound & Video Contractor*, March

1994.

Chapter 26

Attenuators

by Glen Ballou

26.1 General

Most of the circuits today do not require passive attenuators and/or impedance matching devices as their input impedance is high and their output impedance is low. However, if a low-impedance output feeds a long line to a high-impedance input, high-frequency losses will occur if the line is not terminated with a matched impedance. This may be thousands of feet or a few feet when using older equipment that was designed for matched operation. When connecting to external circuits, the signal must often be attenuated to meet standards, a good place for low-maintenance passive attenuators.

An attenuator or pad is an arrangement of noninductive resistors in an electrical circuit used to reduce the level of an audio- or radio-frequency signal without introducing appreciable distortion. Attenuators may be fixed or variable and can be designed to reduce the signal logarithmically or any other curve.

The 1st through the 4th editions of the *Handbook for Sound Engineers* discussed many different passive attenuators and pads. This edition only discusses those which might still be useful in older circuits or to fix a gain problem in the field.

Attenuators and pads may be *unbalanced* or *balanced*. In an unbalanced attenuator, the resistive elements are on one side of the line only, Fig. 26-1. In the balanced configuration, the resistive elements are located on both sides of the line, Fig. 26-2.

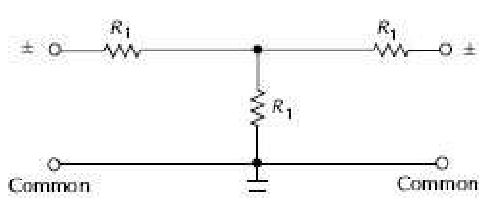


Figure 26-1. An unbalanced T-type attenuator.

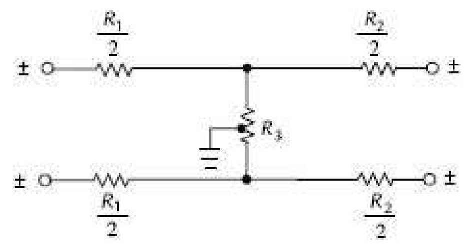


Figure 26-2. A balanced T-type attenuator.

An *unbalanced pad* should be grounded to prevent leakage at the higher frequencies. The line without the resistor elements, called the *common*, is the only line that should be grounded. If the side with the resistors is grounded, the attenuator will not work properly, in fact, the signal will probably be shorted out. A *balanced attenuator* should be grounded at a center point created by a balancing shunt resistance.

Balanced and unbalanced configurations cannot be directly connected together; however, they may be connected by the use of an isolation transformer, Fig. 26-3. If the networks are not separated electrically, half of the balanced circuit will be shorted to the ground, as indicated by the broken line in Fig. 26-4. Here severe instability and leakage at the high frequencies can result. The transformer will permit the transfer of the audio signal inductively while separating the grounds of the two networks. Even if the balanced network is not grounded, it should be isolated by a transformer. Transformers are usually designed for a 1:1 impedance ratio; however, they have taps for other impedance ratios. Chapter 36, *Grounding and Interfacing*, discusses the proper way to connect equipment to eliminate ground problems.

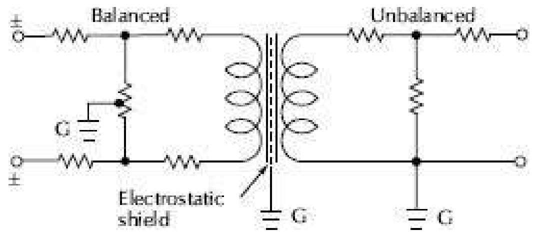


Figure 26-3. Correct method of connecting balanced and unbalanced networks through a transformer.

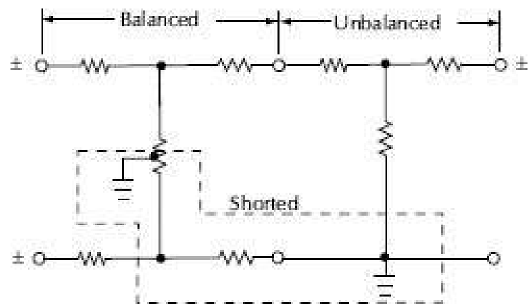


Figure 26-4. Two networks, one balanced and one unbalanced, connected incorrectly.

26.1.1 Loss

The term *loss* is constantly used in attenuator and pad design. Loss is a decrease in the power, voltage, or current at the output of a device compared to the power, voltage, or current at the input of the device. The loss in decibels may be calculated by means of one of the following equations

$$dB_{loss} = 10log\frac{P_1}{P_2} \tag{26-1}$$

$$dB_{loss} = 20log\frac{V_1}{V_2} \tag{26-2}$$

$$dB_{loss} = 20log\frac{I_1}{I_2} \tag{26-3}$$

where,
P_1 is the power at the input,
P_2 is the power at the output,
V_1 is the voltage at the input,
V_2 is the voltage at the output,
I_1 is the current at the input,
I_2 is the current at the output.

The *insertion loss* is created by the insertion of a device in an electrical circuit. The resulting loss is generally expressed in decibels.

A *minimum-loss pad* is a pad designed to match circuits of unequal impedance with a minimum loss in the matching network. This minimum loss is dependent on the ratio of the terminating impedances.

The minimum loss for attenuators of unequal impedance may be read from the graph in Fig. 26-5.

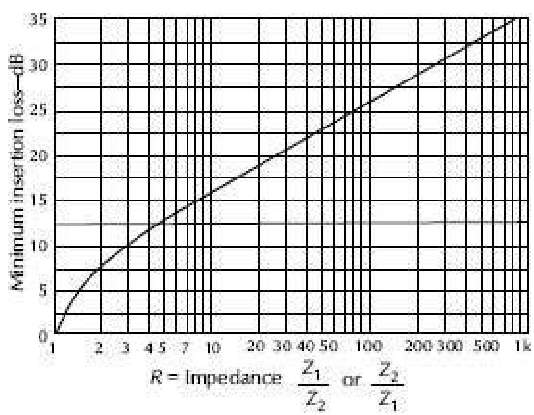


Figure 26-5. Minimum loss graph for networks of unequal impedances.

26.1.2 Impedance Matching

An *impedance-matching network* is a noninductive, resistive network designed for insertion between two or more circuits of equal or unequal impedance. When properly designed, the network reflects the correct impedance to each branch of the circuit.

If two resistive networks are mismatched, generally the frequency characteristics are not affected; only a loss in level occurs. If the impedance mismatch ratio is known, the loss in level may be directly read from the graph in Fig. 26-5 or with the equation

$$dB_{loss} = 20 \log\left(\sqrt{\frac{Z_1}{Z_2}} + \sqrt{\frac{Z_1}{Z_2} - 1}\right) \quad (26\text{-}4)$$

where,
Z_1 is the higher impedance in Ω,
Z_2 is the lower impedance in Ω.

In a simple circuit, Fig. 26-6, where the input to the attenuator is greater than the load, R_1 can be found with the equation

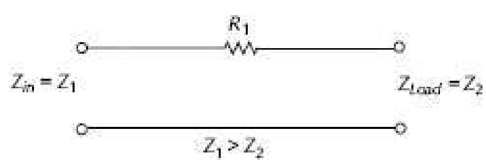


Figure 26-6. Impedance matching a low-impedance load to a high-impedance source.

$$R_1 = Z_1 - Z_2 \quad (26\text{-}5)$$

If the lower impedance is to be matched, R is found with

$$R = \frac{Z_1 Z_2}{Z_1 - Z_2} \quad (26\text{-}6)$$

The resistor is shunted across the line, Fig. 26-7.

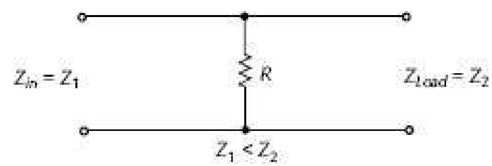


Figure 26-7. Impedance matching a high-impedance load to a low-impedance source.

26.1.3 Measurements

The resistance of an attenuator can be measured with an ohmmeter by terminating the output with a resistance equal to the terminating impedance and measuring the input resistance. The resistance as measured by the ohmmeter should equal the impedance of the pad. If the attenuator is variable, the dc resistance should be the same for all steps.

If the impedance of an attenuator is not known, its value can be determined by first measuring the resistance looking into one end with the far end open and then shorted. The impedance (Z) is the geometric mean of the two readings

$$Z = \sqrt{Z_1 Z_2} \tag{26-7}$$

where,

Z_1 is the resistance in ohms measured with the far end open

Z_2 is the resistance in ohms measured with the far end shorted.

This measurement will hold true only for pads designed to be operated between equal terminations. If the dc resistance of the two ends differs, the pad was designed to be operated between unequal impedances.

If an attenuator is to be converted to a different impedance, the new resistors can be calculated by

$$R_x = \frac{Z_x R}{Z} \qquad (26\text{-}8)$$

where,
Z_x is the new impedance in Ω,
Z is the known impedance in Ω,
R is the known value of resistance in Ω,
R_x is the new value of resistance in Ω.

Any balanced or unbalanced attenuator may be directly connected to another, provided the impedance match is satisfied and the configurations are of such nature they will not cause an unbalanced condition.

26.2 Types of Attenuators

26.2.1 L Pads

L pads are the simplest form of attenuator and consist of two resistive elements connected in the form of an L, Fig. 26-8. This pad does not reflect the same impedance in both directions. An impedance match is afforded only in the direction of the arrow shown in the figures. If an L-type network is employed in a circuit that is sensitive to impedance match, the circuit characteristics may be affected.

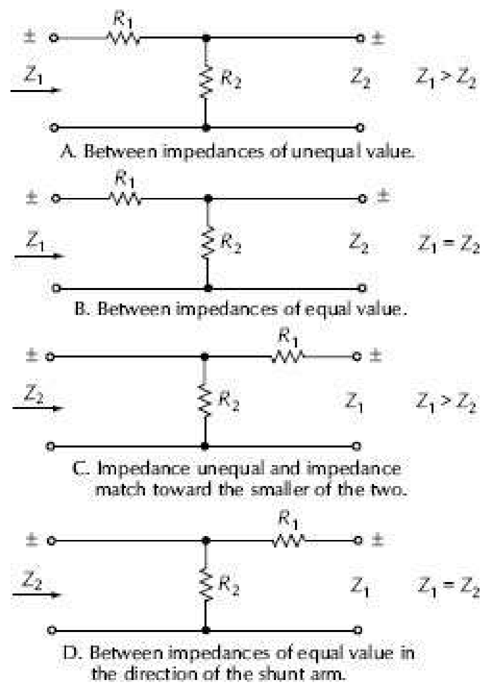


Figure 26-8. Configurations of L-type networks.

If the network is designed to match the impedance in the direction of the series arm, the mismatch is toward the shunt arm. The mismatch increases with the increase of loss, and, at high values of attenuation, the value of the shunt resistor may become a fraction of an ohm, which can have a serious effect on the circuit to which it is connected.

The configuration for an L-type network operating between impedances of unequal value, Z_1 and Z_2, is shown in Fig. 26-8A. The impedance match is toward the larger of the two impedances, Z_1, and the values of the resistors are

$$R_1 = \frac{Z_1}{S}\left(\frac{KS-1}{K}\right) \quad R_2 = \frac{Z_1}{S}\left(\frac{1}{K-S}\right) \qquad (26\text{-}9)$$

where,

S is $\sqrt{\frac{Z_1}{Z_2}}$,

$K = 10^{\frac{db}{10}}$.

If a minimum-loss, L attenuator is used to match two impedances of unequal value, as in Fig. 26-8A, the resistor values will be

$$R_1 = \sqrt{Z_1(Z_1 - Z_2)} \qquad (26\text{-}10)$$

$$R_2 = \frac{Z_1 Z_2}{R_1} \qquad (26\text{-}11)$$

where,

R_1 is the series resistor in ohms connected on the side of the larger impedance;

R_2 is the shunt resistor in Ω.

The loss through the attenuator will be

$$dB_{loss} = 20log\left(\sqrt{\frac{Z_1}{Z_2}} + \sqrt{\frac{Z_1}{Z_2} - 1}\right) \qquad (26\text{-}12)$$

For a condition where the impedances are equal and the impedance match is in the direction of the arrows, Fig. 26-8B, the values of the resistors may be calculated by the equation

$$R_1 = Z\left(\frac{K-1}{K}\right) \qquad (26\text{-}13)$$

$$R_2 = Z\left(\frac{1}{K-1}\right) \qquad (26\text{-}14)$$

When the impedances are unequal and the impedance match is toward the smaller of the two impedances, Fig. 26-8C, the values of the resistors are determined by the equations

$$R_1 = \frac{Z_1}{S}(K - S) \qquad (26\text{-}15)$$

$$R_2 = \frac{Z_1}{S}\left(\frac{K}{KS - 1}\right) \qquad (26\text{-}16)$$

For the conditions shown in Fig. 26-8D, resistors R_1 and R_2 are calculated by

$$R_1 = Z(K - 1) \qquad (26\text{-}17)$$

$$R_2 = Z\left(\frac{1}{K - S}\right) \qquad (26\text{-}18)$$

26.2.2 Dividing Networks

Dividing or *combining networks* are resistive networks designed to combine several devices or circuits, each having the same impedance, Fig. 26-9A. The resistors may be calculated with the equation

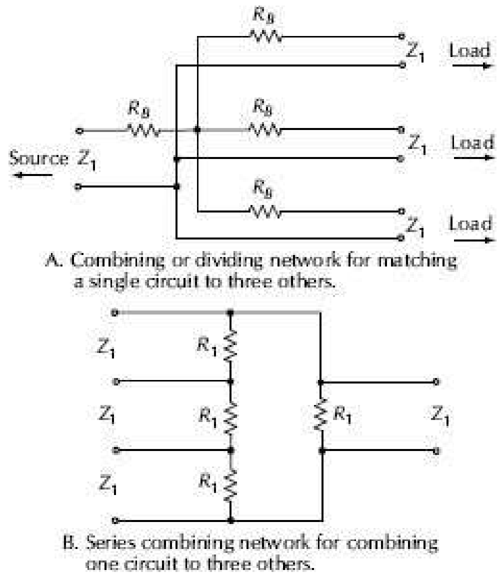


Figure 26-9. A combining or dividing network for matching a single circuit to three circuits.

$$R_B = \frac{N-1}{N+1}Z \qquad (26\text{-}19)$$

where,
R_B is the build-out resistor in Ω,
N is the number of circuits fed by the source impedance,
Z is the circuit impedance in Ω.

The loss of the network is

$$dB_{loss} = 20log(N-1) \qquad (26\text{-}20)$$

where,
N is the number of input or output circuits.

Unused circuits of a dividing or combining network must be terminated in a resistive load equal to the normal load impedance.

This same circuit can be reversed and used as a combining network. This circuit was often used in the design of sound mixers.

Combining or branching networks may also be designed as a series configuration, Fig. 26-9B. For equal impedances the equation is

$$R_1 = \frac{N-1}{N+1}Z \qquad (26\text{-}21)$$

where,
R_1 is the terminating resistor in Ω,
N is the number of branch circuits,
Z is the circuit impedance in Ω.

The insertion loss is calculated with

$$dB_{loss} = 20logN \qquad (26\text{-}22)$$

where,
N is the number of branch circuits.

A series configuration can only be used in an ungrounded circuit. The insertion loss of a combining network may be avoided by the use of an active combining network (see Sections 26.2.15 and 26.2.16).

26.2.3 T Attenuators

A *T-type attenuator* is an attenuator network consisting of three resistors connected in the form of a T, Fig. 26-1. The network may be designed to supply an impedance match between circuits of equal or unequal impedance. When designed for use between circuits of unequal impedance, it is often referred to as a *taper pad*.

A *T type attenuator* may be designed for any value of loss if designed to operate between equal impedances.

A *balanced T pad* is called an *H pad*. The pad is first calculated as an unbalanced T configuration. The series resistance elements are then divided and one-half connected in each side of the line, Fig. 26-2. The shunt resistor remains the same value as for the unbalanced configuration. A tap is placed at the exact electrical center of the shunt resistor for connection to ground.

The average noise level for a T pad is −100dB and constant. Therefore, the signal-to-noise level varies with the amount of attenuation.

26.2.4 Bridged T Attenuators

A *bridged T pad* is an attenuator network containing four resistive elements, Fig. 26-10. The resistors are equal in value to the line impedance; therefore, they require no calculation. This network is designed to work between impedances of equal value only. The contact arms for resistors R_5 and R_6 are connected mechanically by a common shaft and vary inversely in value with respect to each other.

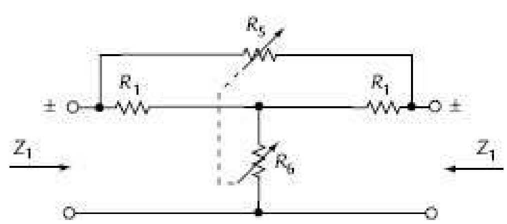


Figure 26-10. A bridged T attenuator. For variable pads, the arms R_5 and R_6 are made variable.

A *balanced bridged T attenuator* is a configuration similar to the unbalanced bridged T attenuator, except each resistor element is divided by 2 and placed in each side of the line. The greatest impedance variation occurs as the attenuator arm approaches zero attenuation and amounts to about 80Ω.

26.2.5 π or Δ Attenuators

A π or Δ *attenuator* is a resistive network resembling the Greek letter pi (π), or delta (Δ), Fig. 26-11. Such networks may be used between impedances of equal or unequal values.

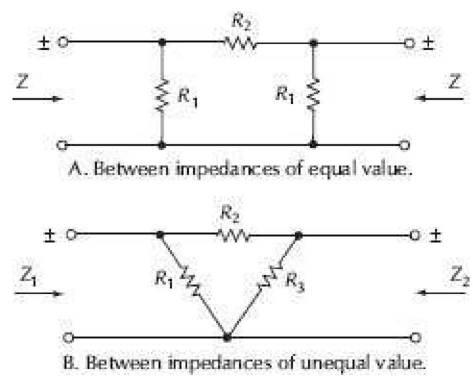


Figure 26-11. π or Δ attenuators.

26.2.6 Ladder Attenuators

Ladder-type pads, Fig. 26-12, are so named because they look like a ladder laying on its side. The ladder pad is actually a group of *pi* attenuators in tandem, R_2 being common to each section. Because of resistor R_4, this type of attenuator has a fixed 6 dB loss, exclusive of the attenuator setting, which must be taken into account when designing a ladder attenuator. The ladder attenuator does not have a constant input and output impedance throughout its range of attenuation. However it does reflect a stable impedance into its source.

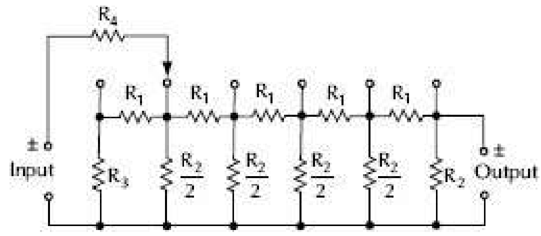


Figure 26-12. Unbalanced ladder attenuator with five fixed steps of loss.

Ladder potentiometers for mixer control use may be obtained in two types of construction—slide-wire and contact types.

For mixers, the slide-wire type control is generally employed because it permits a smooth, even attenuation over a wide range. The contact type, although not quite as smooth in operation as the slide-wire, has only one row of contacts, which reduces the noise and maintenance.

Ladder networks may also be designed for balanced operation. This is accomplished by connecting two unbalanced networks side by side, Fig. 26-13.

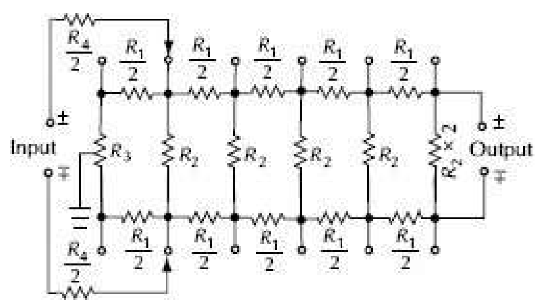


Figure 26-13. Balanced ladder attenuator.

The noise level for a ladder attenuator is on the order of −120dB, and as the attenuation increases, the *SNR* increases.

26.2.7 Simple Volume and Loudness Controls

A simple volume control consists of a potentiometer with the two ends connected to the source and the wiper and one end connected to the load, Fig. 26-14. The volume control should be a high impedance with respect to the source so it will not load it, and the load impedance should be a high enough so as not to affect the control. The output voltage is calculated with the following equation

$$V_{out} = V_{in}\frac{\left(\frac{R_2 Z_2}{R_2 + Z_2}\right)}{R_1 + \left(\frac{R_2 Z_2}{R_2 + Z_2}\right)} \qquad (26\text{-}23)$$

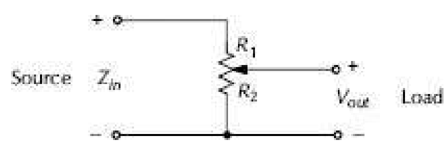


Figure 26-14. Simple volume control.

where,
R_1 is the upper section of control,
R_2 is the lower section of control,
Z_2 is the load impedance.

If the load impedance is high compared to R2, the equation is simplified to

$$V_{out} = V_{in}\left(\frac{R_2}{R_1 + R_2}\right) \qquad (26\text{-}24)$$

The attenuation is

$$dB = 10log(4)\left[\frac{R_1 + \left(\frac{R_2 Z_2}{R_2 + Z_2}\right)^2}{Z_1 Z_2}\right] \qquad (26\text{-}25)$$

Normally, volume controls have a logarithmic taper, so the first 50% of the pot only represents a change of 7-8%, following the ear's sensitivity. If a special taper is required, a linear pot can be altered to change its characteristics by shunting a fixed resistor from one end of the potentiometer to the wiper. Three methods of shunting a straight-line potentiometer are shown in Fig. 26-15. In the first method, the shunt resistor is connected from the wiper to ground. With the correct value shunt resistance, the potentiometer will have a taper relative to the angular rotation, as shown below the schematic. The second method makes use of a second potentiometer ganged with the straight-line potentiometer. In the third method, two shunt resistors connected at each side of the wiper result in a taper resembling a sine wave. A fourth method, not shown, uses a shunt resistor connected from the wiper to the top of the potentiometer.

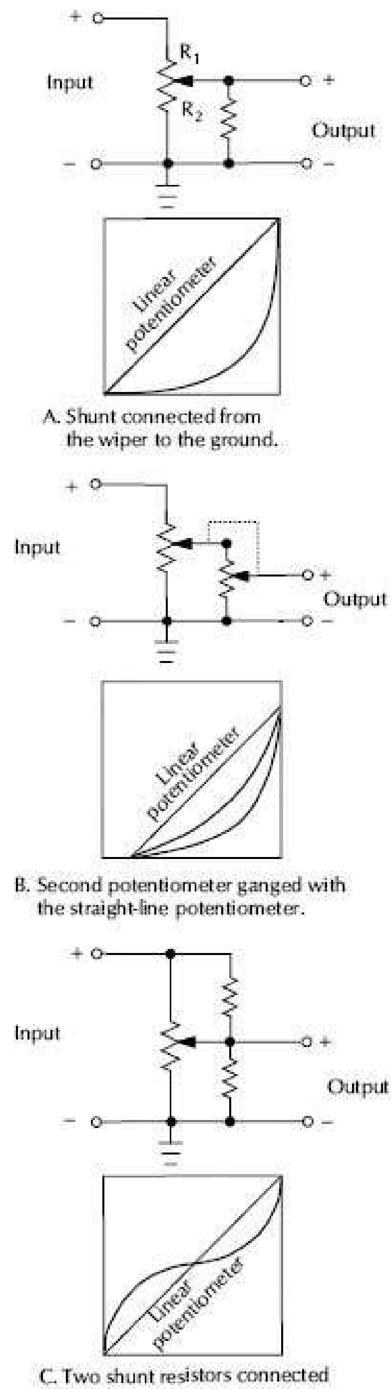


Figure 26-15. Method of varying the response of a simple potentiometer.

A loudness control incorporates a circuit to alter the frequency response to follow the Fletcher-Munson curves of equal loudness, i.e., the softer the level, the more the low frequencies are be boosted with respect to 1 kHz and above. To approximate this a capacitor is tapped off the volume control at about 50% rotation. As the wiper is rotated below the tap, the signal has the high frequencies rolled off, giving the effect of low-frequency boost.

26.2.8 Light-Dependent Attenuators

A *light-dependent attenuator* (LDA) is one where the attenuation is controlled by varying the intensity of a light source on a light-dependent resistor (LDR) (cadmium sulfide cell). LDAs were popular before op-amps and are still useful for remote control as they are not affected by noise or hum on the control line. LDAs eliminate problems of noisy potentiometers as the potentiometers operate the lamp circuit that has an inherent lag time. This type of circuit is also very useful for remote control as the remote control line carries lamp control voltage so it is not susceptible to hum and extraneous pickup.

A simple volume control is shown in Fig. 26-16. *R* and LDR form an attenuator. When the light source is bright, the resistance of the LDR is low; therefore, most of the signal is dropped across R. When the light intensity is decreased, the resistance of LDR increases and more signal appears across the LDR. This circuit has constantly varying impedances. The advantages of a LDA are:

1. No wiper noise.
2. One control can operate many attenuators.
3. Controls can be remoted from the attenuator.

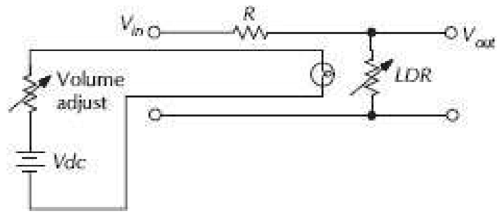


Figure 26-16. Volume control using a light-dependent resistor.

The disadvantages are:

1. Lamp burnout or aging.
2. Slow response time.

26.2.9 Feedback-Type Volume Control

In a *feedback-type volume control* attenuation is controlled by the amount of feedback in the circuit. Feedback-type volume controls have the advantage of reduced hum and noise as they reduce the gain of the active network rather than reducing just the signal level.

A noninverting op-amp feedback gain controlled amplifier is shown in Fig. 26-17. Feedback resistor R_2 is used to adjust the gain of the op-amp and therefore the output. When R_2 is zero, gain will be one as the system has 100% feedback. Increasing the value of R_2 decreases feedback, consequently increasing gain by the ratio of R_2/R_1. Gain can be determined with the equation

$$E_0 = E_{in}\left(\frac{R_1 + R_2}{R_1}\right) \qquad (26\text{-}26)$$

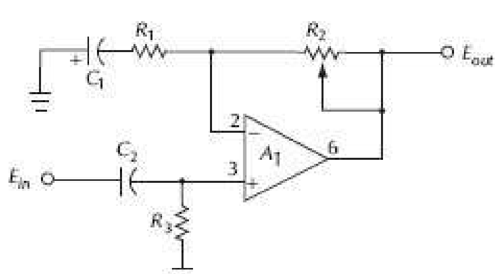


Figure 26-17. Noninverting linear-feedback, gain-controlled amplifier.

26.2.10 Voltage-Controlled Amplifiers

A *voltage-controlled amplifier (VCA)* is used as an attenuator by varying a dc control voltage. VCAs are often used for automatic mixing since the control voltage can be stored in analog or digital form and on command can be programmed back into the console and VCA.

VCAs are also useful for remote control operation and in compressors or expanders. VCAs have attenuation ranges from 0-130dB and response time better than 100μs. A typical circuit is shown in Fig. 26-18. Since the input is a virtual-ground summing point, R1 is used so as not to load the preceding circuit. The output circuit must feed a virtual ground so an operational amplifier current-to-voltage converter (any operational amplifier with a resistor from output to inverting input and with the noninverting input grounded) must be used. The circuit can be used with a linear taper potentiometer to give a linear control characteristic.

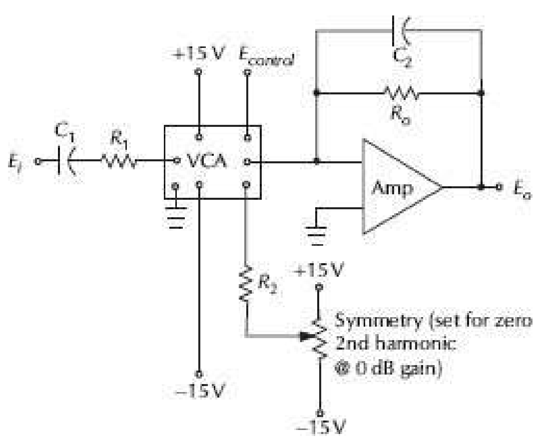


Figure 26-18. VCA volume control.

26.2.11 Field Effect Transistor Attenuators

A field effect transistor attenuator is one where an FET is used to control gain. Field effect transistors have characteristics much like a tube, i.e., high-input impedance and moderate-output impedance. In its simplest form, the FET is used as the lower leg of a voltage divider, Fig. 26-19A.

The voltage out is

$$\begin{aligned} V_{out} &= V_{in} r_{DS(on)} + V_{out(max)} \\ &= \frac{V_{in}}{R + r_{DS(on)}} \end{aligned} \qquad (26\text{-}27)$$

where,
r_{Ds} is the resistance of the drain to source.

To improve distortion and linearity, feedback is required around the FET as in Fig. 26-19B. If a low-output impedance is required, an op-amp can be used in conjunction with the FET, Fig. 26-19C. In this circuit, the op-amp is used to match impedances. The FET can also be used to control feedback, Fig. 26-19D. The gain in this circuit is

$$AV = 1 + \frac{R_f}{r_{DS}} \qquad (26\text{-}28)$$

where,
Rf is the feedback resistor.

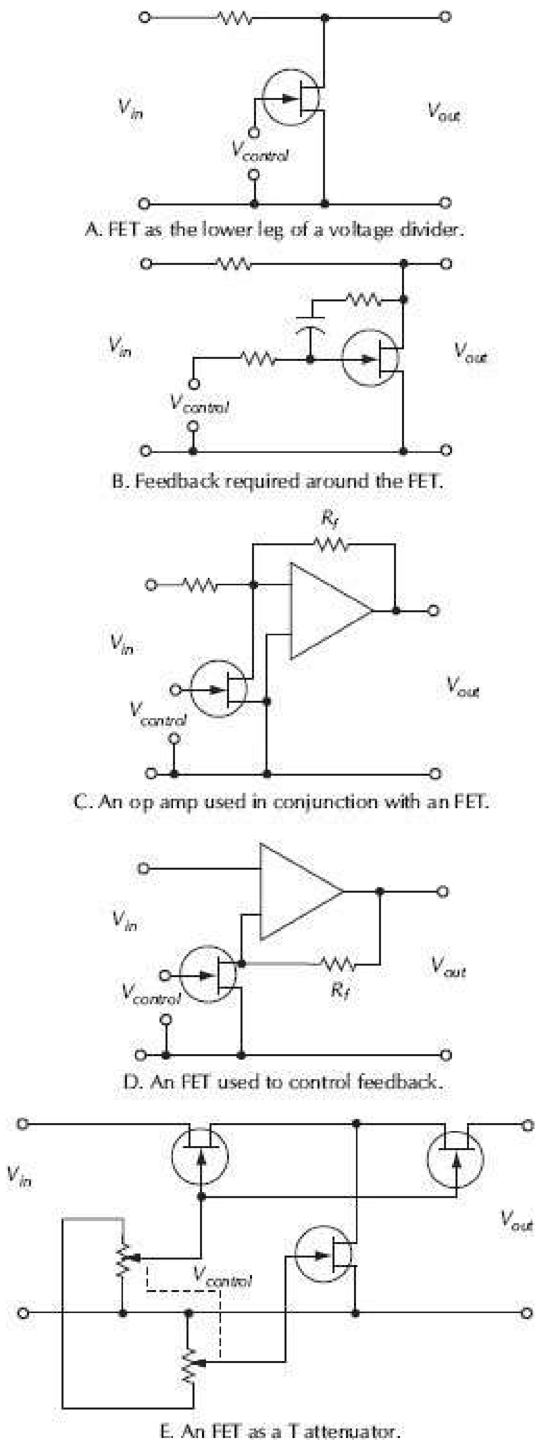


Figure 26-19. FET attenuators.

When r_{DS} is minimum, gain is maximum as most of the feedback is shorted to ground. The FET can also be used as a T attenuator, Fig. 26-19E. This provides optimum dynamic linear range attenuation and tends to hold the impedances more even.

26.2.12 Automated Faders

In an *automated fader*, the fade control can be programmed into a data storage device and used to adjust the fader settings during mixdown, Fig. 26-20. The fader is adjusted manually, and when the desired setting is made, a write voltage is injected into the programmer (encoder) that supplies data to the data track of the tape recorder. During playback, the data track is decoded and, through the read control, adjusts the attenuator to the recorded level. If the mixdown is not proper, any control can be adjusted or updated and the tape played over again.

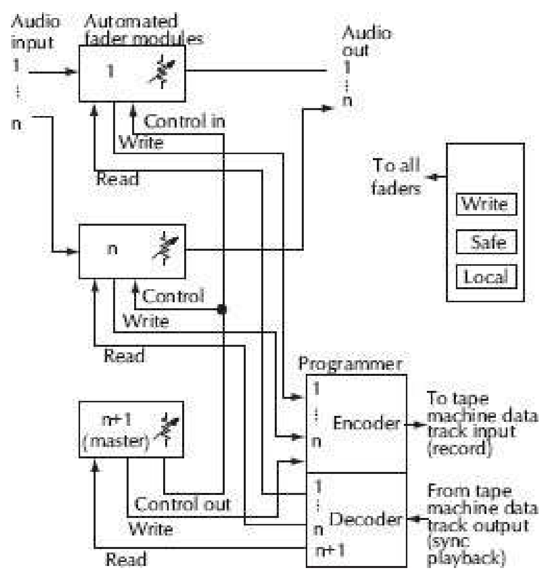


Figure 26-20. Functional block diagram of an automated fader.

Automatic Attenuators

In an *automatic attenuator*, the attenuation varies automatically between two points, usually off and a prescribed setting. Automatic attenuators are often voice operated but can be manually operated.

They are used to automatically turn off unused inputs, as a Ducker and gating.

26.3 Mixers

A mixer is a device used to mix two or more signals into one composite signal. Mixers may be adjustable or nonadjustable and either active or passive.

A passive mixer uses only passive devices (i.e., resistors and potentiometers), Fig. 26-21.

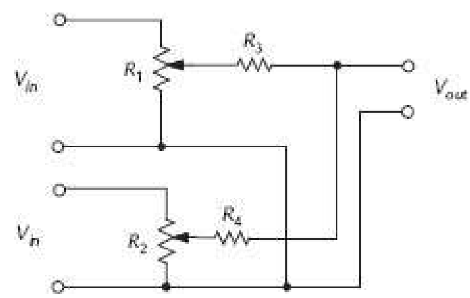


Figure 26-21. A passive mixer circuit.

The main disadvantage of passive mixing is that an amplifier is required after mixing to boost the gain back to the level at the input of the mixer. As the attenuator controls are lowered, the signal on the mixing buzz is reduced; however, the mixing buzz noise remains the same, so the *SNR* is reduced, causing more apparent noise at low levels where high *SNR* is most important. This can be seen in the analysis of Fig. 26-22.

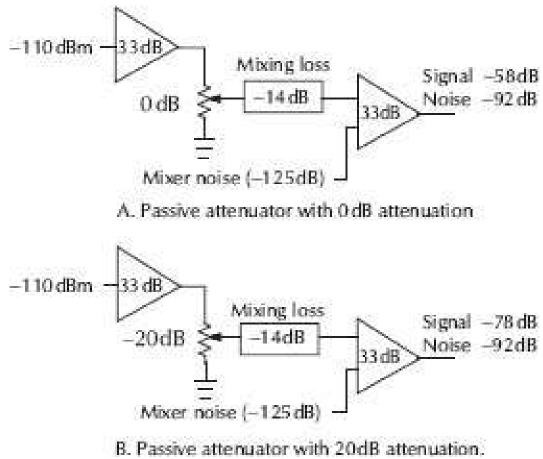


Figure 26-22. Signal-to-noise analyses of passive attenuators.

In Fig. 26-22A, the input signal of −110dBm is not attenuated; therefore, the signal going into the booster amplifier is −91dB and out of the booster amplifier is −58dB [−110 + (+33) + (−14) + (+33)]. The mixer noise going into the booster amplifier is −125dB; therefore, the output noise is −92dB [−125 + (+33)] or 34dB below the signal.

In Fig. 26-22B, the input signal of −110dBm is attenuated 20dB in the mixer so the signal to the booster is −111 dB and the signal output is −78 dB. The mixer input noise is still −125 dB into the booster and −92dB out of the booster, a difference between the signal and the noise of only 14dB, hardly enough to be useful.

An active mixer is one that uses operational amplifiers (op-amps) or some other active device along with resistors and/or potentiometers to control gain or attenuation.

A unity-gain current-summing amplifier is used for a standard active mixer. The mixer is usually designed for an input impedance of about 5-10kΩ, an output impedance of less than 200 Ω, and a gain of 0 to 50. A typical active mixer is shown in Fig. 26-23.

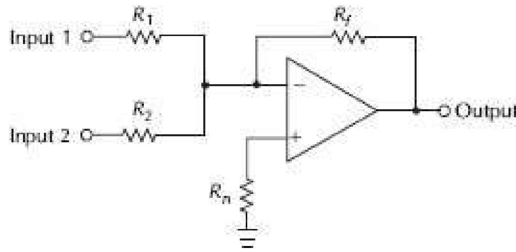


Figure 26-23. An active mixer block diagram.

In unity-gain current-summing amplifiers feedback to the minus or inverting input presents an extremely low apparent input impedance or virtual ground on the inverting input.

The positive input is also essentially ground since the current through R_n will only produce about 0.5mV. While the positive input can be grounded, it is better to make the R_n a value about the same as the parallel combination of $R_1 + R_2 + R_f$ to reduce offset voltage.

Any small, positive-going input applied to the input of R_1 is amplified by the high-gain op-amp driving the output negative since the input signal is on the inverting input. The output signal is fed back through R_f, the feedback resistor, and it continuously attempts to drive the voltage on the input to ground.

Since the input is a virtual ground, the input impedances are determined by R_1 and R_2. The gain of the circuit is

$$\text{input } 1_{gain} = \frac{R_f}{R_1} \tag{26-29}$$

$$\text{input } 2_{gain} = \frac{R_f}{R_2} \tag{26-30}$$

If the gain of both inputs were to be the same, R_1 and R_2 would remain constant and R_f would be varied. Mixers, however, usually require separate gain control for each input so R_1 and R_2 are varied to change the gain of the system. Increasing R_1 or R_2 decreases the gain. The main disadvantage of this system is that the input impedance varies with gain.

The advantage of an active mixer is that gain is included in the mixing circuit; therefore, it does not need a gain makeup amplifier that amplifies both the signal and the mixing noise after the mixer. With active mixing, the mixing noise is also reduced along with the signal, improving the *SNR*, particularly at low level.

26.4 Summing Amplifiers

A standard audio circuit function is the linear combination of a number of individual signals into a common output without crosstalk or loss. This function is well suited for the summing amplifier, which is often referred to as *active combining network*. Summing amplifiers operate much like the mixer in Section 26.3. Fig. 26-24 shows a 10-input summing amplifier using one op-amp. Channel isolation is important in summing amplifiers to eliminate crosstalk.

The primary determinant of interchannel isolation is the nonzero summing-bus impedance presented by the virtual ground of the inverter and, to a lesser extent, by the source impedances at the inputs. To illustrate the method of calculating interchannel isolation, refer to Fig. 26-25. There are two attenuations that a signal must undergo in order to leak from one channel to an adjacent channel. The first attenuation consists of R_i and R_{in}; the second consists of R_i and R_s.

The equation for calculating isolation is

$$\text{Isolation from } E_{in_a} \text{ to } E_{in_b} = \left(\frac{R_{i1} + R_{in}}{R_{in}}\right)\left(\frac{R_{i2} + R_{sb}}{R_{sb}}\right) \tag{26-31}$$

where,

R_{sb} is the E_{sb} source resistance in Ω,

R_{in} is the A_1 closed-loop input impedance in Ω or

$$R_{in} \cong \frac{R_f}{A_{vo}\beta}.$$

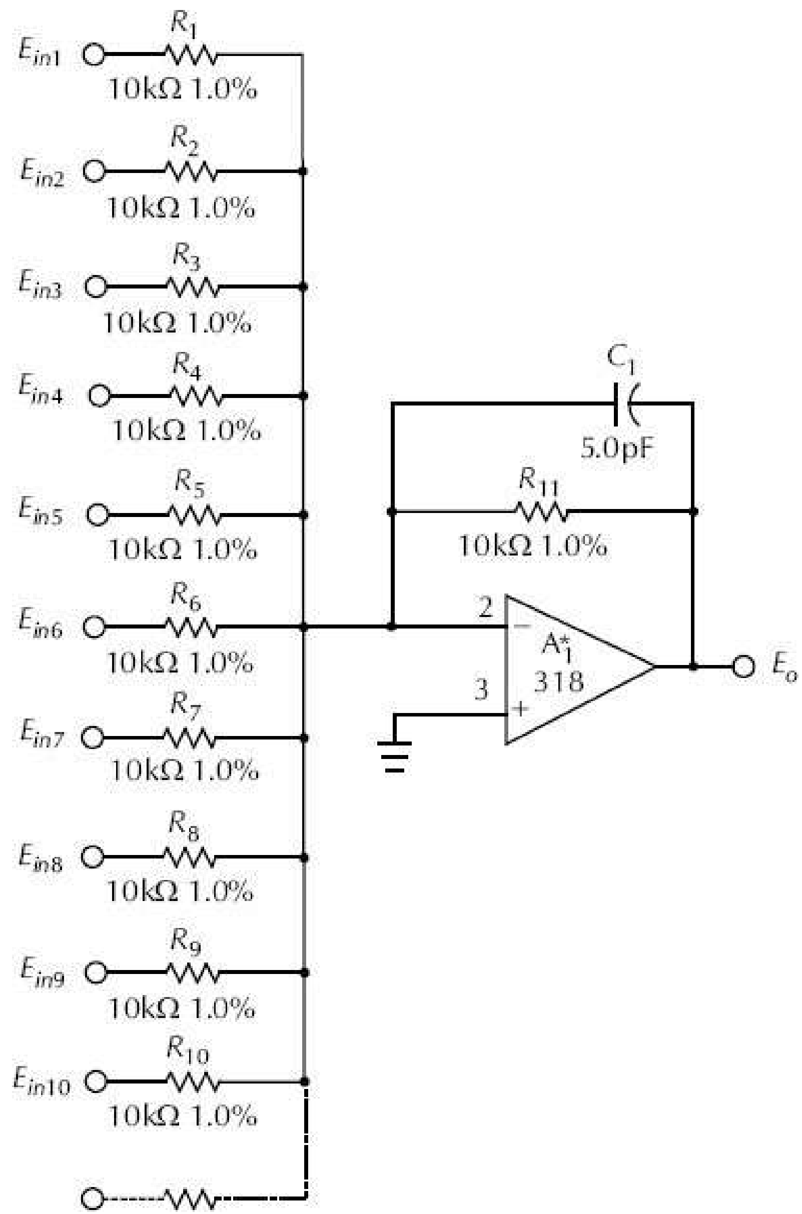


Figure 26-24. Summing amplifier (active combining network).

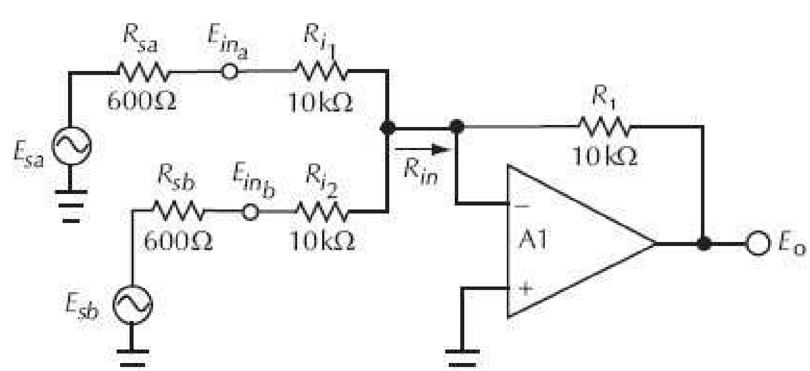


Figure 26-25. Method of calculating interchannel isolation for a summing amplifier.

Chapter 27

Filters and Equalizers

by Steven McManus

27.1 Filter and Equalizer Definitions

A *filter* is a device or network for separating signals on the basis of their frequency. Filters can either be defined in terms of their *pass band* only where, the frequencies of interest are allowed through, or in term of their *stop band*, where certain frequencies are removed, Fig. 27-1. The default design mode for most filters is as a *low pass* where all frequencies below a *cutoff* frequency, and extending down to dc, are allowed to pass. A simple re-arrangement usually allows for a *high pass* to be made, where all frequencies above a *cutoff* frequency, and extending upward, are transmitted. Other mode complex responses such as *bandpass* are constructed from these basic elements.

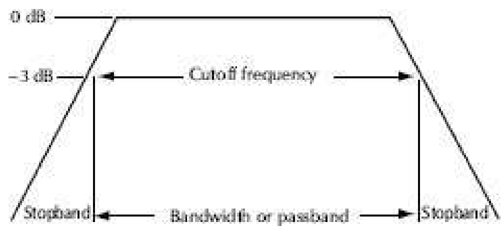


Figure 27-1. Pass bands and Stop bands of a filter.

Passive filters have no amplification components in the circuit. They cannot add energy to the signal so can only act to attenuate signals.

Active filters use transistor or operational amplifier-based gain stages allowing the option of boosting some of the, or the whole, spectrum.

An *equalizer* is a device that uses filters to compensate for undesirable magnitude or phase characteristics of a systems response.

27.1.1 Pass Band

The *pass band* is a band of frequencies that pass through a filter with a loss of less than 3 dB relative to the nominal gain of the filter.

27.1.2 Stop Band

The *stop band* is a band of frequencies that pass through a filter with a loss of greater than 3 dB relative to the nominal gain of the filter.

27.1.3 Cutoff Frequency

A *cutoff frequency* is the frequency at which the gain first falls to 3 dB below the nominal gain of the filter, as you move out of the pass band.

27.1.4 Corner Frequency

A *corner frequency* is the frequency at which the rate of change of a response makes a noticeable change. In the case of a low-pass or high-pass filter, this is the same as the cutoff frequency, but other filters such as shelving filters may have additional corner frequencies.

27.1.5 Bandwidth

The *bandwidth* is the difference between the upper and lower cutoff frequencies on either side of the pass band.

27.1.6 Transition Band

The *transition band* is the range of frequencies over which the gain the filter falls from its level at the cutoff frequency to the nominal attenuation level in the stop band.

27.1.7 Center Frequency

The *center frequency* of a band of frequencies is defined as the geometric mean of the lowest and highest frequencies of the band.

$$f_m = \sqrt{f_1 \times f_2} \qquad (27\text{-}1)$$

where,
f_1 is the cutoff frequency of the high-pass filter,
f_2 is the cutoff frequency of the low-pass filter.

27.1.7.1 Geometric Symmetry

A response showing mirror image symmetry about the center frequency when plotted on a log scale is said to have *geometric symmetry*. This is the natural response of many electrical circuits as the response function tends to contain multiplicative terms.

27.1.7.2 Arithmetic Symmetry

A response showing mirror image symmetry about the center frequency when plotted on a linear scale is said to have *arithmetic symmetry*. A bandpass filter with a constant envelope delay will have arithmetic symmetry in both phase and amplitude. The center frequency in this case will be given by the arithmetic mean

$$f_c = \frac{f_1 + f_2}{2} \qquad (27\text{-}2)$$

27.1.8 Order

The *order* of a filter is determined by the number of *reactive* elements in the circuit. These can either be *inductive* or *capacitive* and generally only include those added for purposes of the frequency response within the audio band and not for stability or RF suppression. If all of the elements act as either low pass or high pass, the roll off in the stop band will approach 6dB/oct per order. A fourth-order low pass will have a roll-off of 24dB per octave above the cutoff frequency, but a fourth-order band pass will have 12dB/oct on either side of the center frequency.

27.1.9 Phase Angle

The *phase angle* at a particular frequency is a measure of the relative time for a particular frequency to pass through a system from input to output. Phase angle is a relative measure and is usually expressed in degrees where 360° represents one wavelength. In most formulas, phase is used in terms of radians where 2π represents one wavelength. The instantaneous phase of a sinusoidal signal is given by

$$\begin{aligned} \alpha &= \omega t \\ &= 2\pi \times ft \end{aligned} \tag{27-3}$$

27.1.10 Phase Delay

The *phase delay* of a system at a given frequency is the equivalent time offset that would induce the same phase offset as measured on a sinusoid of the same frequency.

$$\begin{aligned} \tau_p &= \frac{\alpha}{\omega} \\ &= \frac{\alpha}{2\pi f} \end{aligned} \tag{27-4}$$

27.1.11 Group Delay

A filter can exhibit a *group delay* over a group of frequencies covering a section of the audio spectrum if those frequencies are all subject to the same time delay. The group delay is given as the first derivative of the phase with respect to frequency

$$\tau_g = \frac{d}{d\omega}\phi(\omega) \qquad (27\text{-}5)$$

The threshold of perceptibility for group delay has been shown to be between 1 to 3 ms over the 500Hz to 4 kHz range of the audio spectrum.[1]

27.1.12 Transient Response

The *transient response* of a filter is the time response to an input stimulus. *Impulse* and *step* inputs are common stimuli for this measurement. Narrow bandwidth filters, when subjected to rapidly changing input, ring because it takes a certain amount of time for the energy in the network to change upon application or removal of the signal. Ringing can most clearly be seen as a damped tail on a signal after it has been removed, Fig. 27-2.

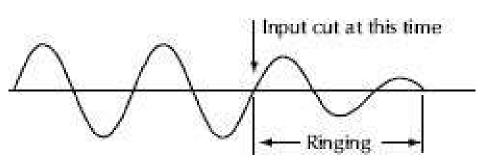


Figure 27-2. Ringing of a filter after the removal of a signal.

27.1.13 Minimum Phase

A *minimum phase* system is one for which the phase shift at each frequency can be uniquely determined from the magnitude response using the *Hilbert transform*. A filter with more than one path from input to output, in which the different branches have a different *group delay*, will be a *linear time invariant* (LTI) system but may be nonminimum phase.

27.2 Passive Filters

Passive filters do not have any amplification components in the circuit and as such cannot put out more energy than is put in. A passive filter can never have a boost in the energy response, although with some resonant circuits, instantaneous voltages may be higher than the input voltage. In this case the output impedance will rise, preventing any significant current from being driven. To build a passive filter with boost, we must construct a filter that cuts all other frequencies and then use a separate amplifier to increase the overall gain.

27.2.1 First-Order L and C Networks

Inductor- and capacitor-based filter networks may be analyzed in terms of their impedances by reducing the circuit to its resistance and reactance components.

The impedance may be represented as a single complex number where the real part is the resistance and the imaginary part is the reactance. The imaginary part of a complex number is given by the magnitude multiplied by the square root of negative one. The mathematical notation for this number is *i* but in engineering, *j* is commonly used to avoid confusion in expressions involving current.

$$Z = R + jX \tag{27-6}$$

Analyzing a network in term of complex impedance allows the calculation of both magnitude and phase at any frequency according to

$$\theta = \tan^{-1}\left(\frac{imaginary}{real}\right) \qquad (27\text{-}7)$$

$$A = \sqrt{imaginary^2 + real^2} \qquad (27\text{-}8)$$

where,
θ is the phase angle of the complex number,
A is the magnitude of the complex number.

27.2.1.1 Capacitive Networks

The capacitor has impedance that approaches a short circuit at high frequency and an open circuit at low frequency. The reactance of a capacitor is given by

$$X_C = \frac{1}{2\pi f C} \qquad (27\text{-}9)$$

where;
X_c is the capacitive reactance,
f is the frequency in Hz,
C is the capacitance in F.

If a capacitor is connected in series with the signal path as in Fig. 27-3A, the capacitor and the resistor form a *potential divider*. Low frequencies will be attenuated as the impedance of the capacitor increases at lower frequencies.

$$V_{out} = V_{in}\frac{R}{(R + Z_c)} \qquad (27\text{-}10)$$

The cutoff frequency of this filter is at the frequency where R = $|Z_C|$, so substituting into Eq. 27-8, we find

$$f = \frac{1}{2\pi RC} \qquad (27\text{-}11)$$

Using the complex analysis in Eq. 27-8, we can determine the phase at this frequency.

$$\begin{aligned} V_{out} &= \frac{V_{in} \times jR}{R + jR} \\ &= \frac{j}{1 + j} \\ &= \frac{1 + j}{2} \end{aligned}$$

So according to Eqs. 27-7 and 27-8, the magnitude is 0.707 or −3dB and the phase angle is −45°.

If a capacitor is connected in parallel with the signal path as in Fig. 27-3B, the capacitor and the resistor form a *potential divider*. High frequencies will be attenuated as the impedance of the capacitor reduces at higher frequencies.

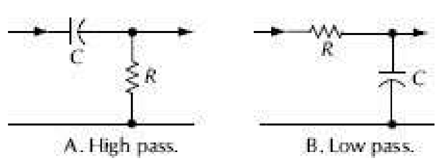


Figure 27-3. Simple filter networks using only a capacitor and a resistor.

$$V_{out} = V_{in}\frac{Z_c}{(R+Z_C)} \quad (27\text{-}12)$$

The cutoff frequency of this filter is at the frequency where $R = |Z_C|$, so substituting into Eq. 27-10, we find that again

$$f = \frac{1}{2\pi RC} \quad (27\text{-}13)$$

Using the complex analysis in Eq. 27-10, we can determine the phase at this frequency.

$$\begin{aligned} V_{out} &= \frac{V_{in} \times jR}{R+jR} \\ &= \frac{j}{1+j} \\ &= \frac{1+j}{2} \end{aligned}$$

So according to Eqs. 27-7 and 27-8, the magnitude is 0. 707 or −3dB and the phase angle is +45°.

27.2.1.2 Inductive Networks

The inductor has impedance that approaches an open circuit at high frequency and a short circuit at low frequency. The reactance of an inductor is given by

$$X_L = 2\pi fL \quad (27\text{-}14)$$

where,
X_L is the inductive reactance in Ω,
f is the frequency in Hz,

L is the inductance in H.

Inductors are prone to parasitic resistances especially for large inductances where a long coil is wound. In a large value inductor, the parasitic resistance is reduced by using heavier gauge wire, which causes the size to grow rapidly as the inductance becomes larger. The full expression for the impedance of an inductor is

$$Z_L = R_L + j2\pi fL \quad (27\text{-}15)$$

where,
Z_L is the impedance of the inductor,
R_L is the dc resistance of the inductor.

If an inductor is connected in series with the signal path as in Fig. 27-4B, the inductor and the resistor form a *potential divider*. High frequencies will be attenuated as the impedance of the inductor increases at higher frequencies.

$$V_{out} = V_{in}\frac{Z_c}{(R + Z_L)} \quad (27\text{-}16)$$

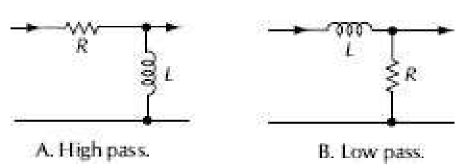


Figure 27-4. Filters using only an inductor and a resistor.

The cutoff frequency of this filter is at the frequency where $R = |Z_L|$, so substituting into Eq. 27-14, we find that

$$f = \frac{L}{2\pi R} \quad (27\text{-}17)$$

Using the complex analysis in Eq. 27-10 and ignoring the parasitic resistance, we can determine the phase at this frequency.

$$
\begin{aligned}
V_{out} &= \frac{V_{in} \times R}{R + jR} \\
&= \frac{j}{1 + j} \\
&= \frac{1 - j}{2}
\end{aligned}
$$

So according to Eqs. 27-7 and 27-8 magnitude is 0.707 or −3dB and the phase angle is −45°. Note that this is the opposite phase angle to the capacitor-based low-pass filter.

27.2.2 Second-Order L-Type Networks

An L-type filter consists of an inductor in series with a capacitor, with the outputs across one or more of the components. Since there are two reactive elements in the circuit, it forms a *second-order* filter with a roll-off of 12dB per octave. There are two configurations of this network, Fig. 27-5.

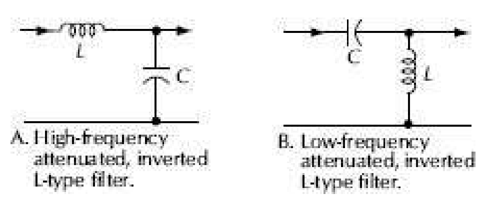


Figure 27-5. Two configurations of L-type filters.

The insertion loss for the low-pass configuration as shown in Fig. 27-5A is given by

$$IL_{dB} = 10\log\left[1 + \left(\frac{f}{f_c}\right)^2\right] \qquad (27\text{-}18)$$

The insertion loss for the high-pass configuration as shown in Fig. 27-5B is given by

$$IL_{dB} = 10\log\left[1 + \left(\frac{f_a}{f}\right)^2\right] \qquad (27\text{-}19)$$

where,
f_c is the frequency of a 3 dB insertion loss,
f is any frequency,
IL_{dB} is the insertion loss in dB.

These configurations are commonly used in basic loudspeaker crossover networks as in Fig. 27-6. Both a high-pass and a low-pass response may be derived from the same circuit. The L-type filter in this application presents constant impedance to the input port. The impedance of the inductor Eq. 27-12 and the capacitor Eq. 27-7 vary with frequency and are chosen such that, at the crossover frequency, their impedances equal the characteristic impedance, Z_0. Each port is in parallel with a load, which for simplicity of analysis we will consider to be constant and of value Z_0

$$Z_L = \frac{1}{\left(\frac{1}{2\pi f_x L} + \frac{1}{Z_0}\right)} \qquad (27\text{-}20)$$

$$Z_C = \frac{1}{\left(2\pi f_x C + \frac{1}{Z_0}\right)} \qquad (27\text{-}21)$$

where,
Z_0 is the circuit impedance,
F_x is the crossover frequency.

The total impedance at the input is

$$Z_{in} = Z_C + Z_L \qquad (27\text{-}22)$$

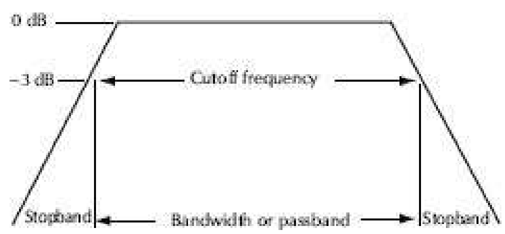


Figure 27-6. Passive crossover using an L-type filter.

When the frequency is very much lower than the crossover frequency, the value of Z_L becomes $2\pi f_x L$, which is very small. At the same time, the value of Z_C becomes Z_0 as $2\pi f_x C$ becomes smaller. The total impedance becomes Z_0.

At the crossover frequency, the inductor and capacitor impedances equal Z_0, so the total circuit impedance also equals Z_0.

When the frequency is very much higher than the crossover frequency, the value of Z_L becomes Z_0 as $2\pi f_x L$ becomes larger. At the same time, the value of Z_C becomes $1/(2\pi f_x C)$, which is small. The total impedance becomes Z_0.

27.2.3 T and π Networks

T and π networks are classes of *constant-k* filters. They are formed by combing L-type filters with one leg being in common. The line impedance Z_0 is a critical parameter in the design of these filters. The impedance presented by a T network to the input and output transmission lines is symmetrical and is designated Z_T. This impedance is equal to the line impedance in the passband and progressively decreases in the stop band. The impedance presented by a π network to the transmission lines is also symmetrical and is designated Z_P. This impedance is equal to the line impedance in the passband and progressively increases in the stop band.

The full T and π networks have twice the attenuation of the L-type half sections.

27.2.3.1 Low Pass

A T-type *low-pass* filter has two inductances: L_1 in series with the line and a capacitance C_2 in parallel, Fig. 27-7. As frequency increases, the inductive reactance increases, presenting an increasing opposition to transmission. As frequency increases, capacitive reactance reduces, so the parallel capacitor becomes more effective at shunting the signal to ground. The design equations for the component values are

$$C_2 = \frac{1}{2\pi f_c Z_0} \tag{27-23}$$

$$L_1 = \frac{Z_0}{2\pi f_c} \tag{27-24}$$

where,
f_c is the cutoff frequency,
Z_0 is the line impedance.

These equations are the same as for the L-type network. In the T network, the actual value of the capacitor is $2C_2$, where the capacitors from two low-pass L-type networks are combined in parallel. In the π network, the actual value of the indictor is $2L_1$, where the inductors from the L-type network are combined in series.

27.2.3.2 High Pass

The basic designs of constant-k high-pass filters are shown in Fig. 27-8. The positions of the inductors and capacitors are opposite to those in the low-pass case. The design equations are

$$C_1 = \frac{1}{2\pi f_c Z_0} \tag{27-25}$$

$$L_2 = \frac{Z_0}{2\pi f_c} \tag{27-26}$$

where,
f_c is the cutoff frequency,
Z_0 is the line impedance.

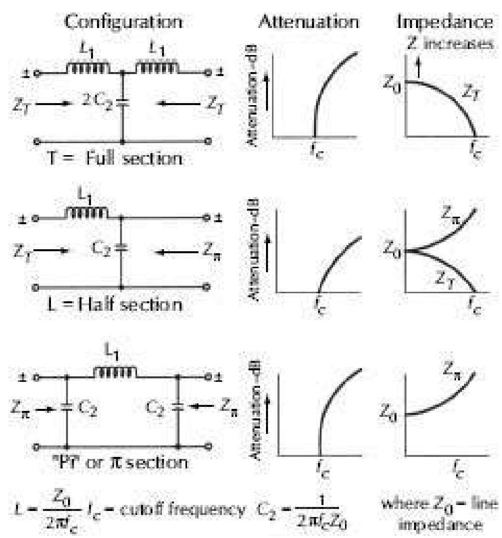


Figure 27-7. Configuration and characteristics of low-pass filters.

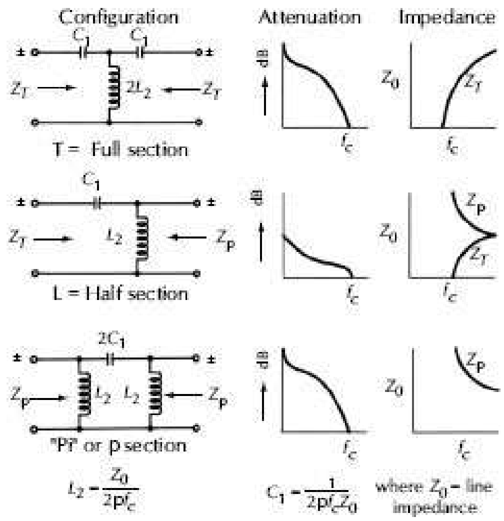


Figure 27-8. Configuration and characteristics of high-pass filters.

27.2.3.3 Parallel Resonant Elements

A *parallel resonant* circuit element has impedance that is at a maximum at the resonant frequency, Fig. 27-9. The impedance of the element is given by

$$Z = \frac{X_L \times X_C}{X_L + X_C} \qquad (27\text{-}27)$$

where,
Z is the impedance,
X_L is the reactance of the inductor,
X_C is the reactance of the capacitor.

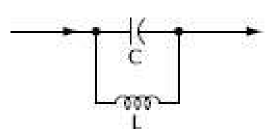


Figure 27-9. Parallel resonant circuit.

At very low frequencies, the reactance of the inductor approaches a short circuit, reducing the overall impedance. At high frequencies, the reactance of the capacitor approaches a short circuit, reducing the overall impedance.

27.2.3.4 Series Resonant Elements

A *series resonant* circuit element, Fig. 27-10, has impedance that is at a minimum at the resonant frequency. The impedance of the element is given by

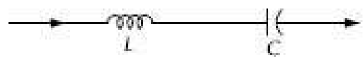


Figure 27-10. Series-resonant circuit.

$$Z = \frac{X_L \times X_C}{X_L + X_C} \qquad (27\text{-}27)$$

where,
Z is the impedance,
X_L is the reactance of the inductor,
X_C is the reactance of the capacitor.

At very low frequencies, the reactance of the capacitor approaches an open circuit, increasing the overall impedance. At high frequencies, the reactance of the inductor approaches an open circuit, increasing the overall impedance.

27.2.3.5 Bandpass

The impedance characteristics of the *series* and *parallel* resonant elements can be used to form a bandpass filter as in Fig. 27-11. The frequencies f_1 and f_2 are the cutoff frequencies of the pass band. The design equations for the component values are

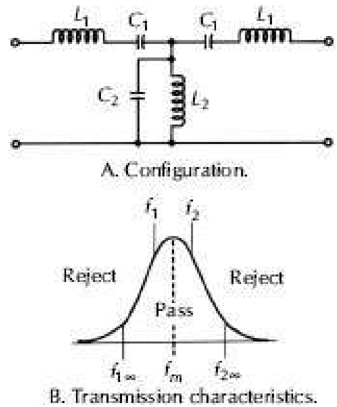


Figure 27-11. T network bandpass filter.

$$L_1 = \frac{Z_0}{2\pi(f_2 - f_1)} \tag{27-29}$$

$$L_2 = \frac{(f_2 - f_1)Z_0}{2\pi f_1 f_2} \tag{27-30}$$

$$C_1 = \frac{f_2 - f_1}{2\pi f_1 f_2 Z_0} \tag{27-31}$$

$$C_2 = \frac{1}{2\pi(f_2 - f_1)Z_0} \tag{27-32}$$

where,
f_1 is the lower cutoff frequency,
f_2 is the upper cutoff frequency,
Z_0 is the line impedance.

27.2.3.6 Band Reject

The configuration for a *band reject* filter using series and parallel resonant elements is shown in Fig. 27-12. The configuration is the reverse of the bandpass T-network filter. In this case the frequencies f_1 and f_2 are at the edge of the reject band. The design equation for the component values are

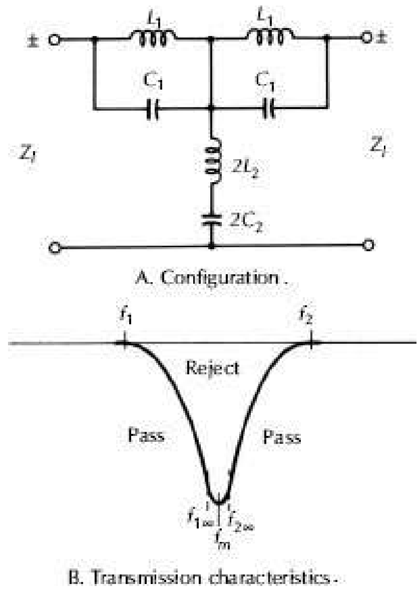


Figure 27-12. T network band-reject filter.

$$L_1 = \frac{(f_2 - f_1)Z_0}{2\pi f_2 f_1} \tag{27-33}$$

$$C_1 = \frac{1}{2\pi(f_2 - f_1)Z_0} \tag{27-34}$$

$$L_2 = \frac{Z_0}{2\pi(f_2 - f_1)} \tag{27-35}$$

$$C_2 = \frac{f_2 - f_1}{2\pi f_1 f_2 Z_0} \tag{27-36}$$

27.2.3.7 Ladder Networks

Passive filters of arbitrary length may be constructed by adding RC, RL, or LC L-type sections into a network of arbitrary length called a *Cauer network*. The interaction between the various stages in this topography starts to become important as the impedance of one section loads the next.

27.2.4 Filter Design

As the number of components in a filter increases, the number of possible transfer functions also increases. Increasing the order of a filter by adding more of the same sections will not necessarily produce the optimum results. Consider chaining two low-pass filters with a cutoff frequency of f_c. The attenuation at the cutoff is 3 dB, so with two sections in series, the attenuation at f_c is 6dB. This means that the 3 dB cutoff point has moved somewhat lower.

We can analyze a filter in the *Laplace domain* in terms of input signals of the form e^{st} with s defined as

$$s = \sigma + j\omega \qquad (27\text{-}37)$$

where,
σ is a value for exponential decay,
ω is $2\pi f$, f being the frequency.

This gives us a transfer function that can be expressed as polynomial functions in s. A first order low-pass filter is of the form

$$h_1(s) = \frac{1}{(s + p_0)} \qquad (27\text{-}38)$$

The value of p_0 defines the cutoff frequency. Adding more sections in series progressively multiplies more terms,

$$h_1(s) = \frac{1}{(s + p_0)(s + p_1)} \qquad (27\text{-}39)$$

For a *normalized* version of Eq. 27-37, p_0 is set to be one, and all other values in the sequence of p_n can be defined according to a formula. The exact formula used depends on the most important characteristic of the filter you are designing.

27.2.4.1 Butterworth

The *Butterworth* filter is maximally flat and has the most linear phase response in the pass band but has the slowest transition from pass band to stop band for a given order. The polynomial transfer function in the form of Eq. 27-37 can be constructed using a formula.

$$B_n(s) = \prod_{k=1}^{\frac{n}{2}} \left[s^2 + 2\cos\left(\frac{2k+n-1}{2n}\pi\right)s + 1 \right] \qquad (27\text{-}40)$$

Eq. 27-40 gives the polynomials for an even order of filter. To calculate the polynomial for an odd order, add a term (s + 1), and then apply the formula with $n = n-1$. Table 27-1 gives the calculated values for the Butterworth polynomials up to fifth order.

27.2.4.2 Linkwitz-Riley

The *Linkwitz-Riley* filter[2] is used in audio crossovers. It is formed by cascading two Butterworth filters so that the cutoff at the crossover frequency is –6dB. This means that summing the low-pass and high-pass responses will have a gain of 0dB at crossover and all

Table 27-1. Butterworth Polynomials

Order	Polynomial
1	(s + 1)

2	$(s^2 + 1.414s + 1)$
3	$(s^2 + 1)(s^2 + s + 1)$
4	$(s^2 + 0.765 + 1)\ (s^2 + 1.848s + 1)$
5	$(s + 1)(s^2 + 0.618s + 1)(s^2 + 1.618s + 1)$

27.2.4.3 Chebyshev I and II

Chebyshev filters have a steeper roll-off than the Butterworth filters but at the expense of a ripple in the response. There are two forms of the Chebyshev filter. Type I has a ripple in the pass band and maximum attenuation in the stop band. Type II is the reverse, with a flat pass band and a ripple in the stop band that limits the average attenuation.

The filter's transfer function is defined in terms of a ripple factor ε

$$H(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 C_n^{\,2} \frac{\omega}{\omega_0}}} \qquad (27\text{-}41)$$

where,

C_n is the polynomial for the order n, as given in Table 27-2.

The magnitude of the ripple in decibels is

$$ripple_{dB} = 20\log\left(\frac{1}{\sqrt{1 + \varepsilon^2}}\right)\ \mathrm{dB} \qquad (27\text{-}42)$$

Table 27-2. Chebyshev Polynomials

Order	Type I	Type II
1	s	$2s$
2	$2s^2 - 1$	$4s^2 - 1\ s$
3	$4s^3 - 3s$	$8s^3 - 4s$

4	$8s^4 - 8s^2 + 1$	$16s^4 - 12s^2 + 1$
5	$16s^5 - 20s^3 + 5s$	$32s^5 - 32s^3 + 6s$

27.2.4.4 Elliptical

The *elliptical* filter has a ripple in both the pass band and the stop band, with the shortest possible transition band for the order of the filter with a given ripple. The ripple in the pass band and the stop band are independently controllable. This is a generalized form of the *Butterworth* and *Chebyshev* filters. If the pass band and stop band ripple is set to zero, we have a Butterworth filter. If the pass band has a ripple and the stop band does not, we have a Chebyshev Type I. If the stop band has a ripple and the pass band does not, we have a Chebyshev Type II.

The transfer function is the same form as Eq. 27-39 with a different polynomial

$$H_n(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 E_{n,\zeta}{}^2\left(\frac{\omega}{\omega_0}\right)}} \qquad (27\text{-}43)$$

where,

$E_{(n,\xi)}$ is the elliptical polynomial for the order n and selectivity factor ξ.

27.2.4.5 Normalizing

Normalizing is the process of adjusting the values of filter components to a convenient frequency and impedance. For analysis, the frequency is usually normalized to 1 rad s^{-1} and the impedance to 1Ω. For designing practical audio circuits the filter is normalized to 1 kHz and 10kΩ.

27.2.4.6 Scaling

Scaling is the design process of changing the normalized frequency or impedance values for a filter by varying resistor and capacitor values. Frequency can be changed relative to the normalized frequency by either changing all of the resistor values or all of the capacitor values by the ratio ρ of the desired frequency to the normalized frequency. From Eq. 27-11, frequency varies inversely with the product of the capacitor and resistor value.

$$\rho = \frac{f_{norm}}{f_1} \qquad (27\text{-}44)$$

where,
ρ is the scaling factor,
f_1 is the new frequency.

By multiplying all of the resistor values by a factor, and dividing all of the capacitor values by that same factor, we can change the normalized impedance of the network without changing the RC product, thus keeping the frequency unchanged.

$$\rho = \frac{Z_1}{Z_{norm}} \qquad (27\text{-}45)$$

where,
ρ is the scaling factor,
Z_1 is the impedance.

27.2.5 Q and Damping Factor

A *damping factor, d*, or its reciprocal, Q, appears in the design equation of some filters. The circuit behaves differently depending on the value of d.

When d is 2, the damping is equivalent to the isolated resistance-capacitance filters.

When d is 1.41 (square root of 2), the filter is *critically damped* and gives maximum flatness without overshoot.

As d decreases between 1.414 and 0, the overshoot peak increases in level with its being 1dB at d = 1.059, 3dB at d = 0.776.

When d is 0, the peak becomes so large that the filter becomes unstable, and if gain is applied it can become an oscillator.

27.2.6 Impedance Matching

Source and load impedance have an effect on a passive filter's response. They can change the cutoff frequency, attenuation rate, or Q of the filter. Fig. 27-13 shows the effects of improper source and load impedance on three different passive filters. The peaks in the response before the cutoff frequency lead to a ringing in the filter, making it potentially unstable at these frequencies. The bridged T filter is not affected by the impedance mismatch because of the resistors in the filter; however, these resistors create an insertion loss.

27.3 Active Filters

Any passive filter may be turned into an active filter by using amplification at the input and output to provide the option of gain, Fig. 27-14. This also provides important buffering, giving the circuit a high-input impedance and low-output impedance, guarding the circuit against external impedance mismatches. This allows active filter sections to be connected together without concerns for mutual interference.

More advanced active filters use filter components in the feedback loop of a gain stage to add functionality with fewer components. Active filters have advantages over passive filters in that they can be made much smaller, especially for low-frequency filters that would otherwise use bulky inductors. The removal of inductors also makes active filters less prone to low-frequency hum interference. The disadvantages of active filters are that they are more complex, having more components to fail; require a power supply; and have a dynamic range limited at the top by the power supply and at the bottom by high-frequency self-noise in the amplifiers.

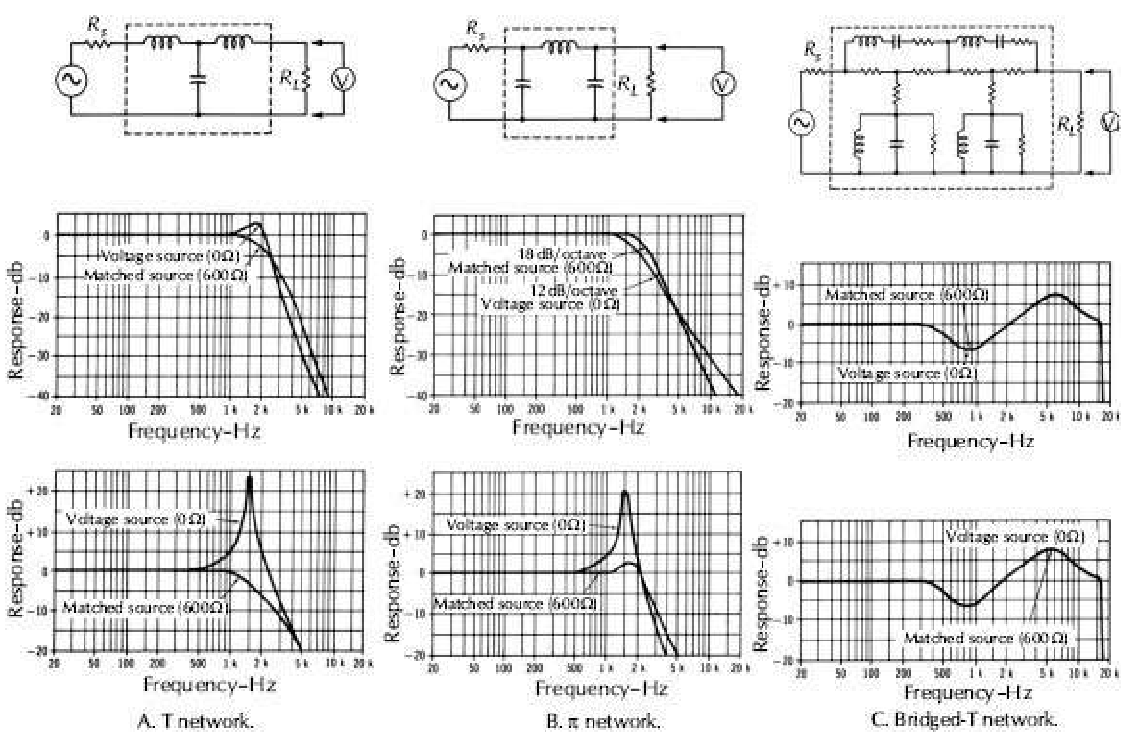


Figure 27-13. Effects of termination impedance on three types of filter sections.

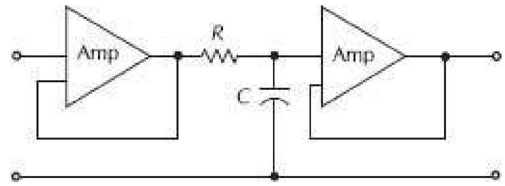


Figure 27-14. Simple buffering of an active filter.

27.3.1 Filter Topologies

27.3.1.1 Sallen-Key

Sallen-Key filters are second-order high-pass or low-pass sections exhibiting a 12dB per octave cutoff slope in the stop band. *Equal component value* filters are the easiest to design, with the frequency-determining resistors being of equal value and the frequency-determining capacitors being of equal value. They have the advantage of being able to high pass or low pass simply by interchanging their positions.

In the second-order low pass of Fig. 27-15, frequency is changed by scaling the values of *R* and *C* in the input network in accordance with Eq. 27-42. To keep the offset at a minimum, it is best to have R_0 equal to the input impedance of 2R. Damping factor, *d*, is controlled by the ratio of R_f and R_0 such that

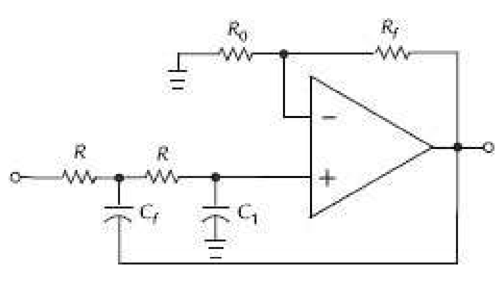


Figure 27-15. Sallen-Key low-pass filter.

$$R_f = (2-d)R_0 \tag{27-46}$$

The gain of the circuit is fixed at

$$\begin{aligned} gain &= 1 + \frac{R_f}{R_0} \\ &= \frac{1 + (2 - d)R_0}{R_0} \qquad (27\text{-}47) \\ &= (3 - d) \end{aligned}$$

where
d is the damping factor,
Rf is the op-amp feedback resistance,
R_0 is the resistance between ground and the inverting input.

The second-order high-pass filter of Fig. 27-16 is constructed by reversing the locations of *R* and *C* in Fig. 27-15. The gain and damping factor follow the same equations as for the low pass.

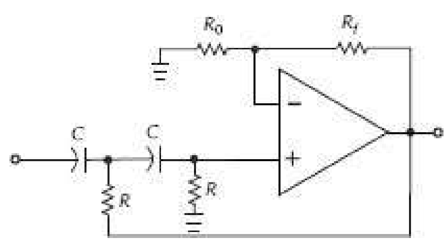


Figure 27-16. Sallen-Key high-pass filter.

A *unity gain* Sallen-Key filter can also be made. To independently control frequency and damping, the ratio of the capacitors must be changed such that in the low pass

$$C_f = \left(\frac{4}{d5}\right) C_1 \qquad (27\text{-}48)$$

The cutoff frequency is still determined by the product of *R* and C, so it can be adjusted with the value of *R* or by scaling C_f and C_1

together.

Fig. 27-17 is a Sallen-Key filter implemented as a bipolar junction transistor circuit.

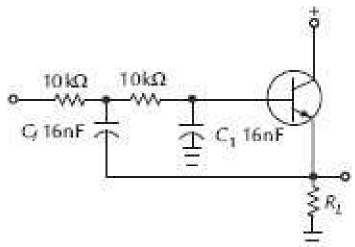


Figure 27-17. Sallen-Key filter implemented as a Bipolar Junction Transistor (BJT) circuit.

27.3.1.2 State Variable

The *state variable* filter consists of two low-pass filters and a summing stage. High-pass, bandpass, and low-pass outputs are all available from the circuit. The operation relies on both the magnitude and phase characteristics of the low-pass sections to generate the outputs.

At high frequency, the low-pass sections attenuate the signal so that the feedback signal is small, leaving the unaffected signal at the high-pass output. As the input frequency approaches the center frequency, the levels at both the bandpass and low-pass outputs begin to increase. This leads first to an increase in positive feedback from the bandpass section giving a damping dependent overshoot. When the input frequency is below the center frequency, the net phase shift of both low-pass sections is 180°, leading to negative feedback and an attenuation of the high-pass output.

The cutoff frequency of the filter in Fig. 27-18 can be changed as in the preceding circuits by varying R_1 and R_2 or C_1 and C_2 while keeping other values identical. The damping factor is varied by changing the band-pass feedback gain, controlled by the ratio of R_3 and R_4.

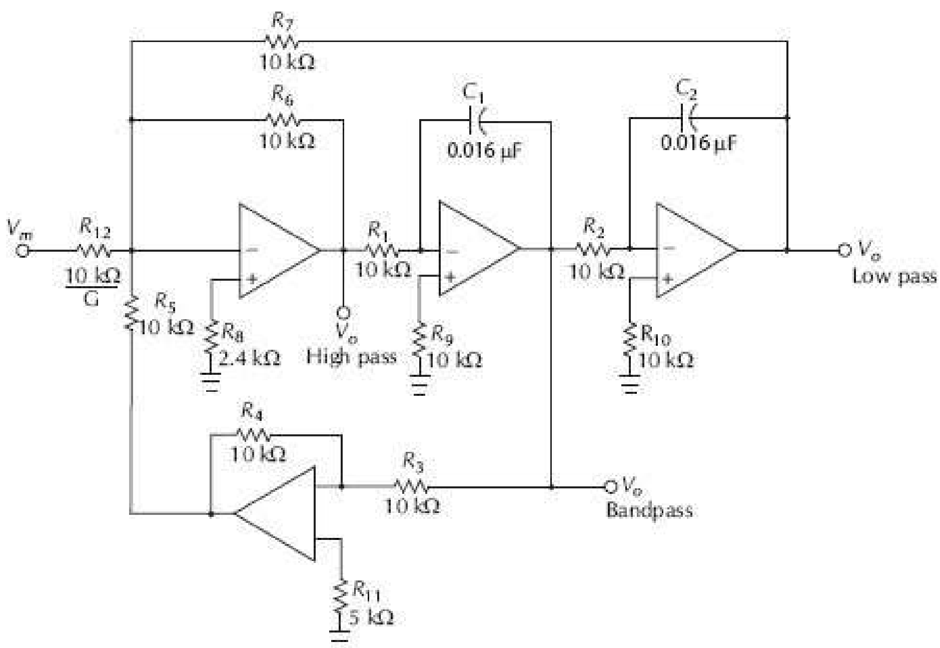


Figure 27-18. Second-order state variable filter.

$$d = \frac{R_4}{R_3} \tag{27-49}$$

The overall gain is controlled by R_{12}. If R_1, R_2 and R_{12} are equal, the gain is one.

$$gain = \frac{R_1}{R_{12}} \tag{27-50}$$

The values of R_8, R_9, R_{10}, and R_{11} are not critical and should be chosen for minimum dc offset at each op-amp stage.

27.3.1.3 All-Pass Filter

The circuit shown in Fig. 27-19 is an all-pass amplifier with unity gain at all frequencies and having a phase shift proportional to frequency according to

$$\theta = 2\tan^{-1}\left(\frac{f_0}{f}\right) \qquad (27\text{-}51)$$

where,
θ is the phase shift from input to output,
f_0 is $1/2\pi RC$).

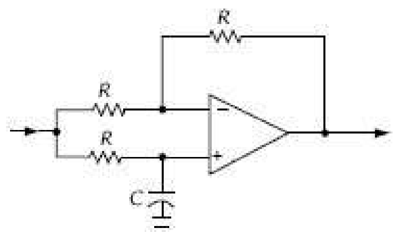


Figure 27-19. All pass unity gain amplifier.

The phase shift is approximately proportional to the frequency over a range of frequencies below and above f_o. These circuits can be cascaded to induce more phase-shift over the same frequency range or each designed with a different f_o to extend the range over which phase is proportional to frequency.

$$\begin{aligned}\theta &= \omega t \\ &= 2\pi f t\end{aligned} \qquad (27\text{-}52)$$

Since phase is proportional to frequency, and from Eq. 27-50 phase is an expression of time, these circuits may be used to introduce a small amount of delay.

27.3.2 Pole-Zero Analysis

A pole-zero plot, Fig. 27-20 is graphical way of representing the complex transfer function of a filter. The pole-zero plot describes a surface that has peaks of infinite magnitude that stretch the surface upward and zeros that do the same downward. The height of the surface along the ω axis, where σ = 0, is the normal magnitude response.

If the expression for the function is reduced to a factored form in the *s-plane* where *s* is the Laplace domain variable Eq. 27-35, then the transfer function of a system can be represented as

$$H(s) = \frac{P(s)}{Q(s)} \tag{27-53}$$

where,
$P(s)$ and $Q(s)$ are polynomials expressible in the form

$$P(s) = (s - p_1)(s - p_2)\ldots(s - p_n),$$
$$Q(s) = (s - q_1)(s - q_2)\ldots(s - q_n).$$

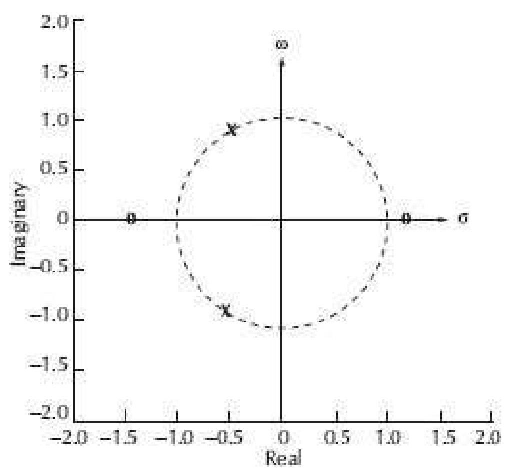


Figure 27-20. Pole-zero plot with poles represented as X and zeros as 0.

27.3.2.1 Zeros

The *zeros* of the function are the values of s at which $P(s)$ is zero and consequently $H(s)$ is zero. These occur at values p_1, p_2, and so on, and represent frequencies at which the transfer function exhibits maximum attenuation.

27.3.2.2 Poles

The *poles* of the function are the values of s at which $Q(s)$ is zero and consequently H(s) is infinite. These occur at values q_1, q_2, ..., and represent frequencies at which the transfer function exhibits maximum gain.

27.3.2.3 Stability

A pole or a zero in the right-hand side of the s-plane means that for that value of s, σ is greater than zero. In the time domain representation, the signal is given as

$$f(t) = \int_0^{\infty} e^{st} F(s) ds \qquad (27\text{-}54)$$

The term e^{st} may be expanded to $e^{\sigma t} \times e^{j\omega t}$. If the value of σ is greater than zero, the expression represents an exponentially increasing factor, meaning that the filter is unstable. This situation cannot arise in a passive filter so they are inherently stable.

27.4 Switched Capacitor Filters

Any active filter based on resistive and capacitive components may be reconfigured as a *switched capacitor filter*. The resistive elements are replaced by an equivalent switched capacitive element. The advantages of using switched capacitors in place of resistors is that they are easier to implement in silicon, since capacitors take up less space than resistors, and tolerances of capacitor-to-capacitor ratios can be more easily controlled the resistor-capacitor products.

The circuit shown in Fig. 27-21 transfers charge, and therefore current, between the two voltage sources under control of the switch. The charge ΔQ transferred every switch period of length t_s may be expressed in terms of current Eq. 27-53 or voltage Eq. 27-54.

$$\Delta Q = It_s = \frac{I}{f_s} \tag{27-55}$$

$$\Delta Q = C(v_1 - v_2) \tag{27-56}$$

Combining these two equations we can find the equivalent resistance.

$$\frac{I}{f_s} = C(v_1 - v_2)$$

$$R = \frac{(v_1 - v_2)}{I} = \frac{I}{Cf_s} \tag{27-57}$$

The equivalent resistor value in Eq. 27-55 has a fixed capacitive term and a frequency term. Its value may be controlled by varying the switching frequency. This makes switched capacitor filters ideal for filters that need to be tuned.

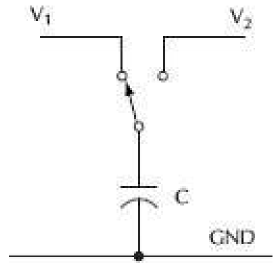


Figure 27-21. Switched capacitor equivalent of a resistor.

27.5 Digital Filters

Filters may be implemented using entirely mathematical means from their transfer function representations in the time domain. The time and frequency domains are related by the *Fourier transform*. Digital filters make use of extensively recursive algorithms involving additions and multiplications, for which digital signal processors (DSPs) are optimized. The precision of the sampled data in magnitude and time is an important factor, not only at the input and output but all through the calculations.

27.5.1 FIR Filters

A *finite impulse response* (FIR) filter performs the convolution in the time domain of the input signal and the impulse response of the filter. While FIR filters are simple in concept and easy to design, they can end up using large amounts of processing power relative to other designs. Hundreds of multiplications per sample are often needed. They are, however, inherently stable as there are no feedback loops that can get out of control when finite precision arithmetic is used. They can also be designed to have *linear phase*, preserving wave shape and having a constant time delay for all frequency components.

The FIR filter structure is shown in Fig. 27-22. Each Z^{-1} is a delay that represents one unit of time equivalent to the sample period of the system. The notation derives from the *Z* domain transform, which is a way of expressing transfer functions in a discrete time form. The recursive nature of the algorithm is apparent, with the multiply and add sections being repeated for every sample in the stored impulse response.

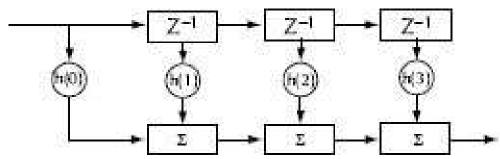


Figure 27-22. Block diagram of an FIR filter.

The FIR filter treats each incoming sample as an input impulse stimulus and generates an output that is a truncated copy of the impulse response scaled by the magnitude of that sample. The summing of the results from each successive sample by *superposition* generates the full output signal. The result for each output sample in a filter with *M* coefficients is

$$y(n) = \sum_{m=0}^{M} x(n-m) \times h(m) \qquad (27\text{-}58)$$

where,
n is the sample number,
x(*n*) is the *n*th input sample value,
h(*m*) is the *m*th filter coefficient value,
y(*n*) is the *n*th output sample value.

This requires the storage of *M* − 1 previous input samples and is executed in *M* multiply and add operations per sample.

27.5.1.1 FIR Coefficients

The coefficient values for an FIR filter are generally computed in advance and stored in a look-up-table for reference while the filter is operating.

Consider the ideal, or brick wall, digital low-pass filter with a cutoff frequency of ω_0 rad s^{-1}. This filter has magnitude 1 at all frequencies less than ω_0 and magnitude 0 at frequencies between ω_0 and the Nyquist frequency. The impulse response sequence $h(n)$ for a filter normalized for frequencies between 0 and π is

$$\begin{aligned} h(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H(\omega)e^{jw\omega n}d\omega \\ &= \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{j\omega n}d\omega \qquad (27\text{-}59) \\ &= \frac{\omega_0}{\pi}\operatorname{sin}c\left(\frac{\omega_0}{\pi}n\right) \end{aligned}$$

This filter cannot be implemented as an FIR since its impulse response is infinite. To create a finite-duration impulse response, we truncate it by applying a window. By retaining the central section of impulse response in this truncation, you obtain a linear phase FIR filter. The length of the filter primarily controls the steepness of the cutoff, while the choice of window function allows you to trade off between pass band and stop band ripple, Fig. 27-23.

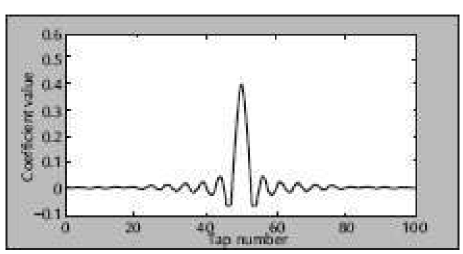


Figure 27-23. Coefficients of a 100-tap low pass filter at 0.2 times the sample rate.

27.5.1.2 FIR Length

The required number of taps (*N*) in an FIR filter at a sample rate (f_S) for a given transition band specified by a width (f_t) and attenuation in dB(A) can be estimated as

$$N = \frac{f_S A}{f_t 22} \qquad (27\text{-}60)$$

As an example, we can calculate how may taps a 100Hz, fourth order high-pass filter in a 48kHz system would use. Fourth order gives a roll-off of 24dB per octave, so the response will be 24dB down by 50Hz. The transition band is for a 20 dB minimum attenuation in the stop band and therefore is 50 × 20/24 or 42Hz wide and the desired attenuation is 24dB so the equation gives us 48,000 × 20 ×/(42 × 22) = 1049 taps. This is a very long filter and introduces 520 samples or 10.8ms of delay at the 48kHz sample rate.

If we consider the same example, but for a 1000 Hz cutoff frequency, everything scales by a factor of 10, giving a much more acceptable filter length of 105 samples with a delay of 1.1 ms. This illustrates the limitation of using FIR filters for low frequencies, Fig. 27-24.

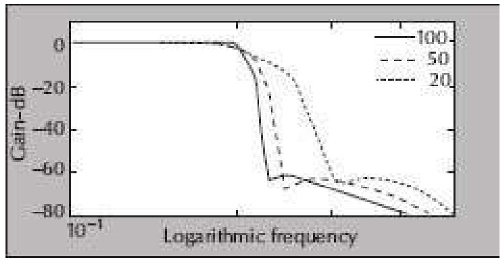


Figure 27-24. Increasing filter steepness with number of FIR taps.

27.5.2 IIR Filters

There are many possible configurations of *infinite impulse response* (IIR) filters, two of which are shown in Figs. 27-25 and 27-26. They show the direct form of a biquad filter in which the input and output samples are passed into the delay line. The transpose form has a sum between every delay and scaled copies of the input and output samples are inserted into the delay line. Direct form I is better suited to fixed point implementation where it is important that the delayed terms maintain as much precision as possible.

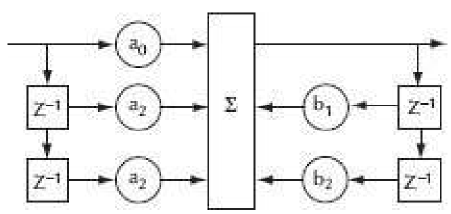


Figure 27-25. IIR Filter implementing a bi-quad section in Direct Form I.

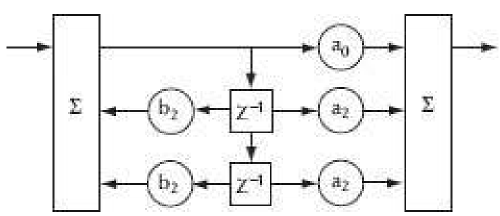


Figure 27-26. IIR filter implementing a bi-quad section in Direct

Form II.

The biquad IIR filter is a second-order filter, and forms the most common basis for higher-order IIR filters. This form fits well with the transfer function equations such as the Butterworth polynomials in Table 27-1. The feedback coefficients correspond to the poles of the filter and the direct coefficients correspond to the zeros. Each section can be represented as a short FIR filter, but unlike the direct implementation of an FIR, these sections can have complex coefficients, even if the input and output are to be real only.

Calculation of Coefficients from Poles and Zeros

IIR filters are designed in terms of the *Z* transform. In this transform, the time domain representation of the filter is used and the notation z^{-1} is used in place of the common exponential terms in the discrete Fourier transform

$$z^{-n} = e^{j\omega n\Delta t} \qquad (27\text{-}61)$$

This gives us the expression for the *Z* transform

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n} \qquad (27\text{-}62)$$

The biquad has two poles in the denominator and two zeros in the numerator. It may be expressed in the factored form as:

$$H(z) = \frac{G(z - r_{01}e^{-jq_{01}})(z - r_{02}e^{-jq_{02}})}{(z - r_{p0}e^{-jq_{p0}})(z - r_{p1}e^{-jq_{p1}})} \qquad (27\text{-}63)$$

where,
G is the gain,
r_o denotes the real part of the zero location,
q_o denotes the imaginary part of the zero location,
r_p denotes the real part of the pole location,
q_p denotes the imaginary part of the pole location.

Table 27-3 lists the equations for the individual coefficients for a purely real implementation of a biquad filter, given the locations of the poles and zeros.

Table 27-3. Relation of Biquad Coefficients to Pole and Zero Location

Zeros	Poles
$a_0 = 1$	
$a_1 = -2r_o cos(q_o)$	$b_1 = -2r_p cos(q_p)$
$a_2 = r_o^2$	$b_2 = r_p^2$

27.6 Equalizers

Equalizers are devices or components that are designed to compensate for undesirable characteristics in the magnitude or phase response of another part of the system and thus make the response equal again. Equalizers consist of filters implemented in such a way as to provide control over the frequency response in terms of how the operator thinks of the response curve that they are trying to recreate. Equalizers give control over one or more of the parameters that affect the response over the audio range, usually 20Hz–20kHz, and ideally do so such that the parameters do not interact. Controls are arranged in terms of center frequencies, bandwidths, and gains rather than actual circuit values that control these things. This means that often the controls are dual-ganged so that the ratio of two resistor values may be kept constant while their absolute values are changed.

27.6.1 Tone Control

The simplest form of equalizer is the *tone control* as used on portable radios. The control only acts to attenuate the high frequency. Another version of this type of equalizer that is becoming more prevalent than the tone control is the *bass boost*, which as its name suggests acts as the exact opposite to add a controlled gain to the low frequencies.

The tone control circuit shown in Fig. 27-27 includes transistor-based buffer amplifiers around the passive filter section in the middle. This allows the operation of the equalizer to be independent of source and load impedances.

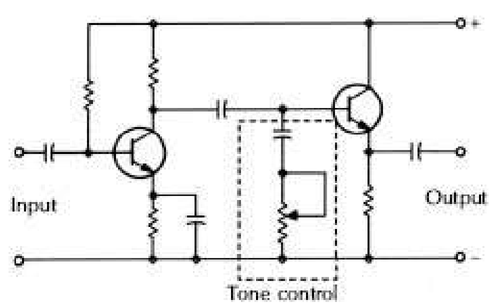


Figure 27-27. Simple low-pass tone control

27.6.2 Graphic Equalizers

A *graphic equalizer* is used to shape the overall spectrum of program material. The term *graphic* refers to the way that the controls are set out on the front panel such that the positions of the slider controls draw the desired frequency response. Graphic equalizers typically use 1/3-octave band filters but may be constructed with any spacing. The 1/3-octave refers to the spacing between adjacent filters and not necessarily the width of the filter.

A graphic equalizer is constructed using a series of filters with fixed frequency and width. The centers of the filters are typically on the ISO preferred frequencies rather than the mathematically correct 1/3-octave spacing. This means that in order to cover the spectrum completely, some of the filters must have different widths. The output of each filter is added to the original signal to a degree controlled by a slider control. The levels add together and are prone to producing a ripple in the response between the centers. In Fig. 27-28, the four sliders for 800Hz, 1000 Hz, 1250Hz, and 1600 Hz were set to +5dB. The overall peak is greater than desired and a ripple of 2dB is induced across the band.

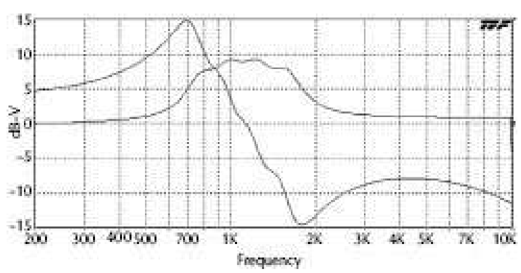


Figure 27-28. Magnitude and phase response of a graphic equalizer.

Transversal Equalizers

Fig. 27-28 shows an example of how graphic equalizers based on tuned filters exhibit ripple in the response when groups of adjacent controls are used. The actual response that we were trying to create would have been better achieved using a single filter as in Fig. 27-29. The Transversal *equalizer* configures as a graphic equalizer produces ripple-free response for any equal or flat setting of the controls. It produces minimum phase response curves and avoids phase mismatch anomalies at the band edges that can be a problem in other equalizers. The response curve is mathematically a best match for the desired response.

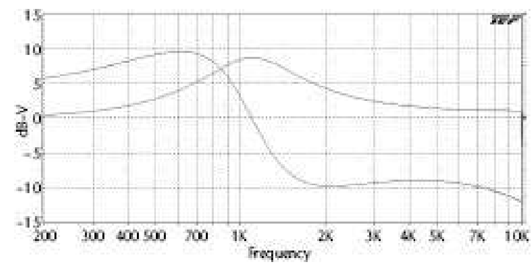


Figure 27-29. Magnitude and phase response of a single bi-quad filter.

The FIR filter discussed previously is a digital implementation of a transversal filter. Whereas a conventional tuned filter operates in the frequency domain, a transversal filter operates in the phase or time domain. If a unity gain all-pass circuit stage, Fig. 27-19, is substituted for each Z^{-1} delay element in Fig. 27-22, an analog transversal filter is created. The coefficients are implemented by summing the outputs of the successive delays via different weighting resistors to a summing amplifier.

27.6.3 Parametric Equalizers

Parametric equalizers allow adjustment of the filters in term of the three main parameters that define a filter.

- The boost or cut in dB.
- The center frequency.
- The bandwidth or *Q*.

It is difficult to make a parametric filter that provides completely independent control over all three parameters over a wide frequency range. Several filter components have to be varied with one control. For this reason, parametric equalizers sometimes have one of the controls as a multiposition switch instead of continuously variable. This allows a band of calibrated components to be switched into place rather than having to worry about how variable component values track.

Parametric equalizers are always active and typically there are several second-order sections in a unit. Each band's center frequency is adjustable over a limited frequency range so that the parameters' independence can be maintained. This means that each section in a unit typically covers a slightly different frequency range, each section having a ratio of between 10:1 and 25:1 between the highest and lowest center frequency. The lowest band will adjust down to 20Hz and the highest band up to 20kHz. Each section will typically provide more scope for cutting levels than for boosting. Typical boost level is up to 15 dB while the available cut may be down to −40dB. The bandwidth or *Q* is not consistent in its labeling between manufacturers. Some specify bandwidth in Hz, some specify *Q*, and others specify octave fraction. In terms of *Q*, the range of this control will typically be between 0.3 and 3, with the critically damped value of 0.707 being in the center position of the control.

An overall gain is usually provided to help maintain the average level and to maximize headroom by avoiding clipping.

27.6.3.1 Semi-Parametric Equalizers

A reduced version of the parametric equalizer is commonly found on mixing consoles. This is the *semi-parametric* or *swept frequency equalizer*. This type has only the center frequency and cut or boost controls. The *Q* is usually set to be a midrange critically damped value but can also be configured so that the *Q* varies with gain.

27.6.3.2 Symmetric or Asymmetric Q

Straightforward designs produce *constant Q* filters that have the same *Q* for any amount of boost or cut. If the frequency response curves for the same amount of boost as cut are mirror images of each other across the unity gain axis, the response characteristic is called *reciprocal* or *symmetrical*. This means that the bandwidth of frequencies affected when boost is applied is greater than that affected when cut is applied. Fig. 27-30 shows that in the symmetrical response, the cutoff frequency in attenuation mode F_c is less than that in boost mode F_b.

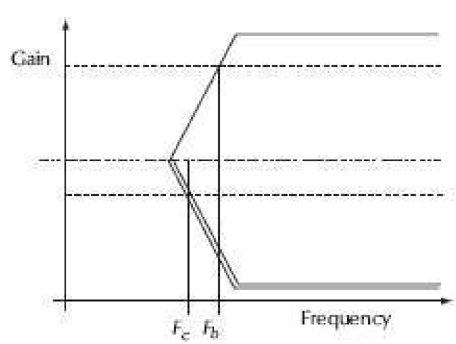


Figure 27-30. Symmetrical response with different bandwidth in cut and boost.

This is not always the most musically useful response. It is more common in spectrum shaping to want to gently apply boost to a broader region. Boosting a narrow region tends to lead to instability. At the same time, it is more useful to be able to notch out a fairly precise frequency, without removing a large portion of the surrounding spectrum. For this reason, equalizers tend be designed so that the bandwidth increases with gain.

27.6.4 Programmable Equalizers

All types of equalizers can be programmable. In digital equalizers, the filter coefficients are stored in memory and may be recalled or modified at will. Unless a digital equalizer implements only a fixed set of coefficients it is inherently programmable.

In programmable analog equalizers, a digital control system is used to physically manipulate the analog filters. This can be either by controlling switches that swap components in or out of the circuit, or by using voltage-controlled gain to alter the filter's response. In the case of switched capacitor filters, the digital control system can adjust the filters by manipulating the switching frequencies to adjust the equivalent resistor values and thus the filter characteristics.

27.6.5 Adaptive Equalizers

The *adaptive equalizers* have long been used in communications systems for multipath echo cancellation. They are the ultimate equalizers for sound systems that must adapt to acoustic conditions that may change at any time. A common example of an adaptive equalizer in sound reinforcement is a feedback suppressor. In this application, the equalizer monitors the signal passing through it for the characteristic exponential increase in level of a frequency that is associated with feedback buildup. When this increase is detected, a very narrow and deep notch filter is placed at that frequency to suppress the feedback. This can typically operate in a fraction of a second such that you were unaware that the event occurred.

References

1. J. Blauert, and P. Laws, "Group Delay Distortions in Electroacoustical Systems," *JASA*, Vol. 63, no. 5, pp. 1478–1483 (1978).
2. S. Linkwitz, and R. Riley, "Passive Crossover Networks for Non-coincident Devices," *JAES*, Vol. 26, no. 3, pp. 149–150 (1978).

Chapter 28

Delay

by Steven McManus

28.1 Introduction

Delay is relative. For a delay to have an effect on a sound it must be heard in conjunction with the original, nondelayed sound. There are two ways that this can occur. A single sound can arrive at the listener via two different length paths, such as a direct sound and a reflected sound, or two signals with different delays can be added electrically and then heard from a single location, Fig. 28-1.

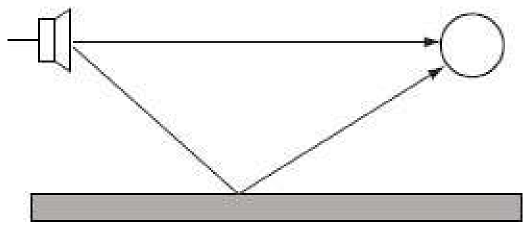


Figure 28-1. Different sound paths through air.

28.1.1 Comb Filter

Two copies of the same signal at different delay times combine to add or subtract depending on the relative phase of each frequency as shown in Fig. 28-2. If the waves are a whole period apart, they combine to give a peak in level, if they are half a period apart, they cancel out to an extent controlled by their relative levels. This effect sets up a comb filter, so named for its appearance on a frequency plot as shown in Figs. 28-3 and 28-4. The series of peaks in the response fall first at dc, then at every frequency whose period is equal to an integer multiple of the delay time. The cancellation notches occur at the exact midpoints between these frequencies.

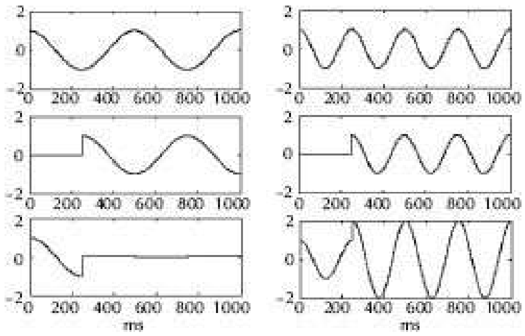


Figure 28-2. Effects of adding signals of different frequencies with the same delay.

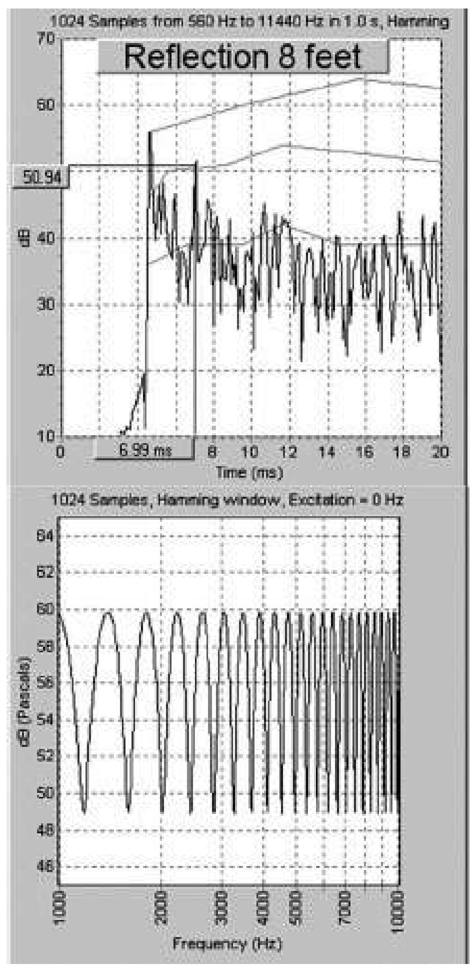


Figure 28-3. A reflection that will alter the perceived direction of the sound.

28.1.2 Directional Perception

In the case of sounds traveling through the air, the path lengths with their corresponding travel times are different for every point in space, resulting in a different comb filter for every location.

The brain uses the results of the different comb filters that are in effect at the location of each ear and combines this information in conjunction with the arrival times, relative levels, and directional filtering due to the shape of the pinnae to determine the originating direction of a sound. Other cues such as the ratio of direct to reverberant energy are used to help determine distance.

A completely dry sound heard in a set of headphones will appear to originate inside your own head. Gradually adding reverberation to it will make the sound appear to move away out in front of you. The sound can be made to appear to move from side to side by alerting the relative levels in each ear, as is commonly done in a pan control, but the same effect can be achieved by altering the relative delay of the dry sound to each ear. The reasons that this delay technique is not commonly used are that the level control is much simpler to implement and the result is compatible with monaural reproduction when the left and right channels are summed.

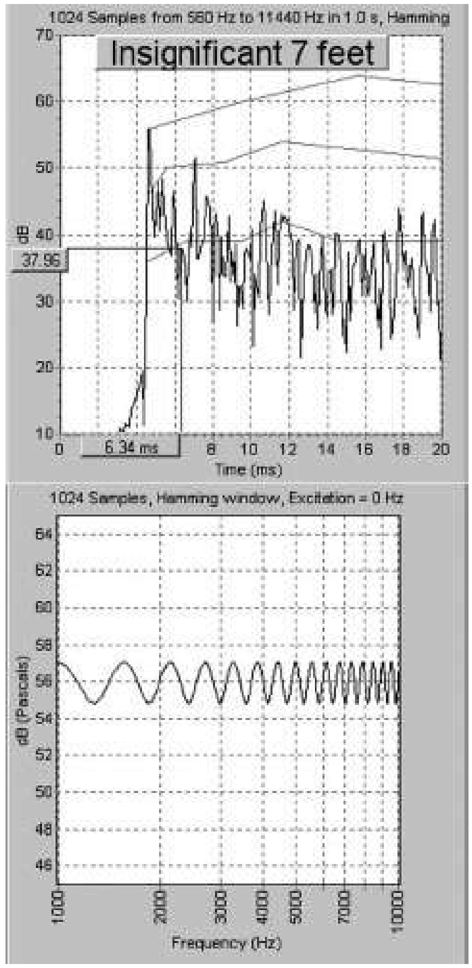


Figure 28-4. A reflection that is too low in level to affect perception of the sound.

A sound is perceived as originating in the location at which it was first heard. This is generally the correct location as the direct sound will always arrive before any reflected sounds. The same sound coming from a second location will be perceived in different ways depending on its timing and level relative to the first:

- If the second sound is more than 30ms after the first it will be heard as a distinct echo.
- If the second sound is more than 10 dB louder than the first, it will be heard as a distinct echo.

- If the second sound is within 10 dB and less than 30ms after the first, it will cause an image shift in where the source location is perceived.
- If the second sound is more than 10 dB below the first, it will contribute to the spatial feel of the sound but will not be heard as a distinct sound or alter the apparent location of the first.

These rules of thumb are approximations of the psychoacoustic effects in operation. The perception curves are more complex than the rules of thumb suggest. The actual values are plotted in Fig. 28-5 and tabulated in Table 28-1.

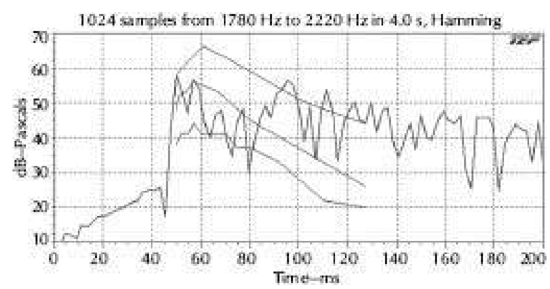


Figure 28-5. Perception curves.

28.2 Uses of Delay

Delay is sometimes useful. It should also be noted that there can be undesirable delays in a system. This is particularly true with digital recessing equipment where there is always a conversion delay in and out of the processor plus any processing delay. It is not uncommon for processors to have a minimum delay of a few milliseconds, and these delays should be considered when calculating the amount of delay that you actually want to use.

28.2.1 Delay in Loudspeaker Systems

A sound amplified through a loudspeaker system will be subject to image shifts and audible echoes only if there is a reference point against which to judge it. This is usually the case in sound reinforcement with the original sound being the reference or in multiple loudspeaker setups where another loudspeaker can act as the reference.

Table 28-1. Perception Curves of Fig. 28-5 Tabulated

	Echo	Image Shift	Spacious
Ms after direct	**dB**	**dB**	**dB**
0	0	−10	−20
1		−6	
2			−17
4	4	−5	
5			−17
7		−2	−14
10			−17
11	8		
17	6	−5	
20	5		−17
25			−21
30		−13	−21
40			−25
50	−7		
60	−10		−36
77	−14	−32	−38

It is not generally desirable for loudspeakers in a system to appear to be generating echoes as this will have a detrimental effect on the intelligibility of the system. Whether the image shift effects are important depends on the application. In a stage system, it is desirable to have the apparent sound source at the stage, regardless of the placement of the loudspeakers. In a distributed announcement system, the creation of a coherent source image is not as important as the intelligibility.

Sound travels at 334m/s or 1130 ft/s. A sound traveling 33 ft will be delayed by 30ms, so with sound sources greater than 33 ft apart, delay should be used to avoid the creation of echoes.

28.2.2 Setting Delay Times

In Fig. 28-6 the sound from the source, a person talking, is to be augmented by a loudspeaker and the apparent source of the sound is to be kept on the stage. To achieve this, the sound from the source must arrive at the listener before the sound from the loudspeaker. The time taken for the signal to arrive at the listener from the loudspeaker is a combination of the distance traveled in air from the loudspeaker to the listener and the negligible time taken for the signal to arrive electrically at the speaker. We must delay the signal to the loudspeaker by an amount that allows the direct sound traveling more slowly through the air to catch up and overtake the sound from the loudspeaker. The delay should slightly exceed the time taken for the sound to travel the difference in distance between the source and the loudspeaker so that the direct sound will be heard first and localized to the source. The loudspeaker can then add up to 10dB of level 5 to 10 ms later to increase the level of the sound without changing its apparent position.

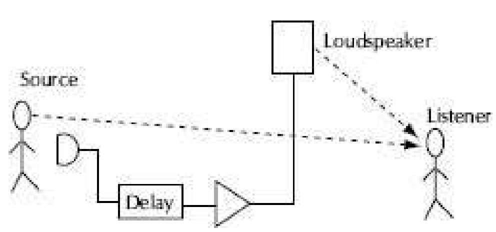


Figure 28-6. A delay in the sound system corrects for the differences in path length between the source and listener and the loudspeaker and listener.

A graphical method for setting delays is shown in Fig. 28-7. The positions of the source and loudspeakers are plotted and a series of concentric circles drawn around them at 30ms (33ft) intervals. The *SPL* level from the polar response pattern of the loudspeaker can also be plotted, but for simplicity in this example, omnidirectional sources are used where the level decreases by 6 dB per doubling of distance.

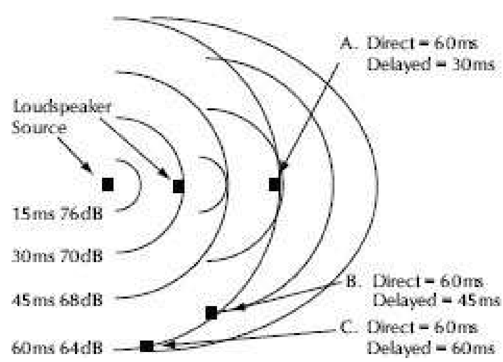


Figure 28-7. Graphical method of setting delays.

If we look at point A, where source and loudspeaker are in a direct line, the time difference is 30ms. If we add a small amount to this to allow the direct sound to be heard first, we come up with a delay setting of 35ms. We can now analyze the sound at each of the three points:

1. The loudspeaker is 6dB louder than the source and 5 ms later. The overall sound is heard as originating at the source but 6dB louder.
2. The loudspeaker is 2dB louder than the source and 20 ms later.

The overall sound is heard as originating at the source but 2dB louder.

3. The loudspeaker is the same level as the source but 35 ms later. At this point the delay is too long and a distinct echo is heard.

In reality, the coverage pattern of the loudspeaker should be chosen to ensure that the level of the sound is sufficiently attenuated outside the area where the delay works effectively.

28.2.3 Reverberation Synthesis

Reverberation is the result of many reflections of the original sound. The general pattern of events, as shown in Fig. 28-8 is that there is first a direct sound, followed by a short gap, referred to as the initial time gap (ITG). Next come the first distinct early reflection echoes caused by sound bouncing off surfaces near either the source or the listener. Thereafter the reflected sounds start to generate their own second, third and higher-order reflections and the energy level settles down to a constant decay rate. This decay rate is related to the distances traveled and the amount of absorption in the room.

Delay is used as the basis for reverberation synthesis because it provides a convenient method for storing the signal and releasing it at a later time, much as reflections from the surfaces of a room arrive at the listener at a later time than the direct sound. Typical applications for synthetic reverberation include the enhancement of program material in the production of recordings, the introduction of special effects in live entertainment productions, and compensation for poor or lacking natural reverberation in entertainment spaces.

Requirements for good reverberation synthesis are essentially the same as for an acoustically well-designed hall. There are many parameters that need to be considered to help achieve realism in reverberation simulation.

- **Distance from Source**. The perception of distance is controlled primarily by the relative energy levels of the direct components and the decay components.
- **Room size.** The perceived room size is controlled by the delay time from the first grouping of direct sound and early reflections to the start of the decay tail and by the length of the decay tail. The requirements for the decay tail are the same as for an acoustically well-designed room. A relatively smooth decay rate is desirable, with longer decay times at lower frequencies than at higher frequencies to simulate the high-frequency losses as sound travels through the air.
- **Brightness.** The spectral balance of the decay tail determines the character of the reverberation. A lot of high-frequency roll-off simulates a room with a lot of absorption from carpets, curtains, or designed absorption devices and gives a dark sound. Less height frequency roll-off simulates a hard-surfaced room such as the inside of a stone church, giving a bright sound.
- **Character.** The smoothness of the decay determines the character of the sound. A room with large opposing flat surfaces will exhibit a *flutter echo* where the sound bounces back and forth between the walls with little diffusion. A room with more architectural features or multiple surfaces will tend to scatter the sound more, creating a denser and more evenly distributed decay.

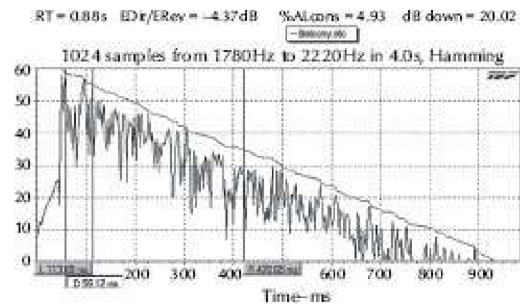


Figure 28-8. Energy Time Curve showing the delay of sound in a room.

- **Envelopment.** The sense of the reverberation coming from all around you rather than a specific location is controlled, making different patterns for the different playback channels. This can be effectively achieved using two-channel stereo as well as in systems with multiple dedicated speakers. The most important differences are in the pattern of the early reflections. The decay tail portions should keep the same with the direct to reverberant energy ratio and decay time but can be randomized to produce a denser sound field. Portions of the randomized signal can be altered in their frequency response to mimic the ear's nonuniform response to sounds from behind, further increasing the sense of envelopment.

A tapped delay a shown in Fig. 28-9 is suitable for creating early reflections. Delays T_1 ... T_N are unequal in length and in the range of 10ms to 30ms with amplitudes set by g_1 ... g_N as appropriate for the character of the room. In Fig. 28-10, a reticulating path is provided via *gf* that produces the exponentially decaying portion. This could simply be fed into the start of the reflection generator, but a more satisfactory result is obtained by using a separate decay section, where the delay taps are set more densely and may be over a wider range, typically between 5ms and 100ms. The delay tap times should be chosen to not be harmonic products of each other to minimize the buildup of standing waves and comb filters. Any gain product in the reticulating path must be less than one, otherwise, the sound will exponentially increase until distortion occurs.

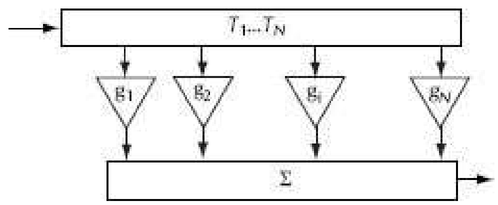


Figure 28-9. A delay with multiple taps for creating early reflections.

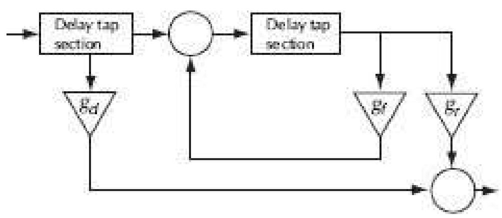


Figure 28-10. Delay section with feedback for producing a decay tail.

28.2.4 Delay-Based Effects

Flange is an audio effect caused by mixing an original (dry) copy of a sound with a delayed (wet) copy. The amount of delay is varied over time, creating a varying pattern of comb filters that sweep up and down through the audio spectrum. Chorus is used to make one voice or instrument sound like many and has the same topography as a flanger, but with longer delays.

28.3 Implementations

The implementation of a delay requires some means of storing the signal and then releasing it after a controlled period of time. This can be done either by storing a continuous record of the sound or by breaking it up into samples that are stored separately. Some preparation of the signal is usually required to make it compatible with the chosen storage medium and method and may involve some postprocessing to restore the stored signal to a usable form.

28.3.1 Small Delays

Small delays may be realized using the phase shift characteristics of an all-pass filter. Such a circuit, illustrated in Fig. 28-11, is limited to delays of the order of less than a wavelength of the highest frequency. These sections may be chained together to produce longer delays but become impractical for delays longer than a few milliseconds. This method is sometimes used in active crossover systems for a loudspeaker. The delays needed to time-align drivers within a cabinet are small and fixed, and the circuit may be easily combined with the frequency filtering requirements.

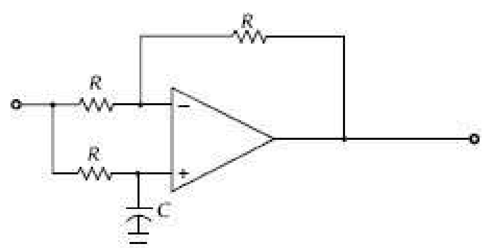


Figure 28-11. An all-pass amplifier having phase shift proportional to frequency and exhibiting a small amount of delay.

28.3.2 Acoustic Delay Methods

One way to implement a long delay is to use the speed of sound and send the signal to be delayed through a fixed air path, such as a tube with a loudspeaker at one end and a microphone at the other. For such a device to work effectively, the tube must be damped to prevent internal reflections and have an absorber at one end to prevent the establishment of standing waves. This type of system has many disadvantages as the system becomes very large for any useful delay time. The frequency response changes with tube length due to the damping material and the signal attenuates as it travels, meaning that large amounts of gain are required. The large gain in turn leads to the requirement that the tube must be mechanically isolated from vibration and outside sounds to prevent these from being added to the delayed sound.

28.3.3 Tape Delay

A more practical early implementation method for continuous delay was to use a magnetic tape loop. An example of such a device is shown in Fig. 28-12. The sound is recorded onto the tape at the record head and then is read back by one or more playback heads. The tape then passes an erase head and loops back to the start. The delay time is given as

$$Time = \frac{distance\ between\ heads}{tape\ speed} \qquad (28\text{-}1)$$

Figure 28-12. Tape loop delay system.

Only the length of the tape limits the maximum delay time. The performance of this system depends on the quality of the recording system. Dynamic range, frequency response, and *SNR* are affected by the tape speed and track width. These parameters may be improved by using the usual tape recording tricks such as compression and various forms of preemphasis/deemphasis noise reduction. Such systems need regular maintenance including head cleaning and replacement of the tape to maintain optimum performance.

28.3.4 Analog Shift Register Delays

Analog shift registers as illustrated in Fig. 28-13 appear in two forms, the bucket brigade and the analog charge coupled device (CCD). They both operate in a very similar manner and differ at the silicon level in the type of switches: a metal oxide-semiconductor capacitor (MOSC) structure for a CCD and a metaloxide–semiconductor junction FET capacitor (MOSJC) structure for the bucket brigade.[2] They are classed as shift registers because they move single samples of a signal in the form of an electrical charge from one stage to the next in response to timing signals. The delay (T) of a shift register is proportional to the number of register elements (N) and inversely proportional to the frequency (f_s) of the timing signal.

$$T = \frac{N}{f_s} \qquad (28\text{-}2)$$

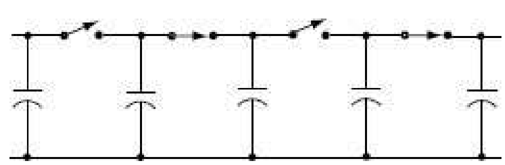


Figure 28-13. Bucket brigade: switched alternately open and close to hand charges along the line.

The term *charge transfer device* (CTD) has been applied to both the bucket brigade and CCD-based structures of an analog delay. The term *CCD* has become colloquially associated with a type of light-sensitive array used in cameras but actually refers to the method used to read the information off these devices. The idea of a CTD is that it stores a sample of analog information as a packet of charge on a capacitor and, under control of a timing signal, transfers it to the next storage site. All the requirements of sampling theory for audio-frequency band limiting should be met. The performance parameters for a CTD include transfer efficiency (ε), the fraction of charge left behind in each transfer; the leakage of charge from a cell during the holding period; and the leakage of charge into a cell due to semiconductor thermal effects. Taken together, these effects degrade the *SNR* of the signal as it passes through the CTD and also lead to distortion due to the nonlinear nature of the leakages. The practical use of CTD is limited to applications requiring less than 100ms of delay or longer where SNR and distortion may be tolerated. CTDs have largely fallen out of use as delay lines because digital systems have become cheaper.

28.3.5 Digital Delays

Digital delay operating principles have undergone a number of important changes since they were first used in sound systems. Probably the most significant is in the type of storage or memory. Shift registers were almost universally used in the first commercially produced units. Digital shift registers are conceptually similar to analog CTD devices with the important advantage that only the presence or absence of a charge carries the significant signal information. Now random access memory (RAM) provides flexibility and economic tradeoffs for design. Until recently, the cost of memory was the dominant factor in delay design considerations. Currently, with the trend for DSPs to include large amounts of on-board memory, the systems have vastly reduced in cost and now the dominant cost factor is in the A/D and D/A converters.

28.4 Sampling in time

Both analog CTD delays and digital delays rely on breaking the delayed signal up into discrete samples. These samples are created by looking at a signal's amplitude at regular intervals and disregarding its amplitude at all other times. The procedure is shown in Fig. 28-14. The sequence of pulses (B) controls a switch that turns on the signal (A) for a brief instant, then disconnects it for the remainder of the sampling period. The result is an amplitude-modulated pulse train (C) where each pulse has amplitude equal to the instantaneous signal value. According to the sampling theorem, a continuous bandwidth-limited signal that contains no frequency components higher than a frequency f_c can be recreated if it is sampled at a rate greater than $2f_c$ samples/s. This rate is called the *Nyquist frequency*. Since the real world never completely satisfies theoretical conditions, sampling frequencies are usually chosen to be higher than $2f_c$. Thus, 20kHz bandwidth delays will typically be found with a sampling frequency of 48 kHz rather than the bare minimum of 40 kHz.

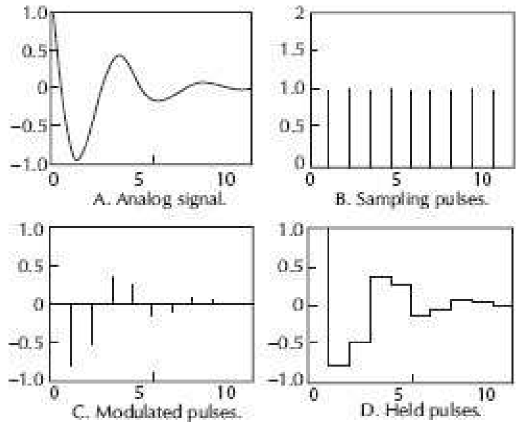


Figure 28-14. The process of sampling a signal.

28.4.1 Aliasing

Sampling of the audio signal is a form of modulation. Modulation of a bandwidth-limited signal with an upper frequency of f_c, by the sampling frequency f_s produces additional copies of the original spectrum centered on frequencies f_s $2f_s$, $3f_s$, etc. If the sampling frequency is not high enough or the bandwidth is not adequately limited, part of the spectrum centered on f_s will fold over into the original signal spectrum as in Fig. 28-15. The fold-over components become part of the signal in the recovery process, producing unwanted frequencies that cannot be filtered out.

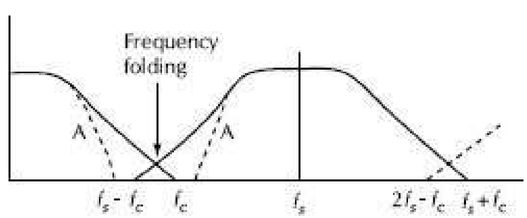


Figure 28-15. Frequency spectrum folding over around the sampling frequency.

An example of the effect of aliasing can be seen on moving wagon wheels in a movie that appear to reverse direction. The sampling rate of the film is lower than the rate at which individual spokes pass the top of the wheel. When the image is reconstructed, the spoke frequency has folded over and the wheel appears to move at a different rate. This phenomenon is known as *aliasing*. Fig. 28-16 shows that aliasing where the sample points lie have the same amplitude on two waveforms of different frequencies.

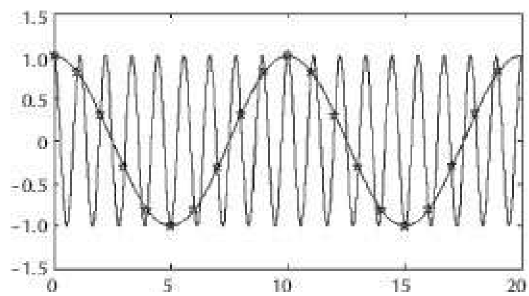


Figure 28-16. Aliasing.

Aliasing must be eliminated or at least largely reduced by selection of a high sampling rate and an adequately sharp *antialiasing* filter. At the output side, a similar low-pass *antiimaging* filter must be used to reduce the number of high-frequency glitches due to the switching at the sample rate.

The economics of delay design dictate that a relatively low sampling rate be used as this reduces the amount of storage that is required for a given length of signal. The number of storage locations required is the product of the sample rate and the length of the delay.

The required cutoff rate of the antialias filter is governed by the separation of the upper frequency f_c and the Nyquist frequency $f_s/2$. As these two frequencies become closer, the number of poles required in the filter increases, adding cost to the filter.

Antialias filters may be implemented as either analog or digital circuits. A digital antialiasing filter still requires some form of an analog antialias filter but relies on a high rate of *oversampling* to ease the design requirements. The digital filter does not have the demanding memory requirements of the delay line, so it can operate at a much higher sample rate than the storage section.

28.4.2 Capturing a Sample

A *sample and hold* circuit takes a very fast snapshot sample of the instantaneous voltage of an analog signal and then changes into a hold mode to preserve that voltage. A hold circuit forces the amplitude of the sample to have constant value throughout a sample period. In Fig. 28-14D, the sample amplitude is shown being set at the beginning of the sample period.

A basic sample and hold circuit is shown in Fig. 28-17. The signal amplitude is frozen for a brief period of time on a capacitor until the next sample period is initiated, at which time the new signal amplitude is transferred to the capacitor. The switch is momentarily closed, under the control of the sample pulse, and then reopened. The amplifier A_1 must have low-output impedance to make it capable of driving enough current to charge the capacitor to the appropriate voltage during the brief ontime of the sampling pulse. The output amplifier A_2 must have high-input impedance so as not to draw excessive charge from the capacitor as any leakage of current will cause a change in the voltage. The capacitor should also be low leakage to help hold the voltage stable. An analog delay may be constructed entirely out of sample and hold circuits that transfer charge from one to another.

28.4.3 Errors in Sampling Magnitude

Any sampling system, digital or analog, will take a finite time to convert the input voltage into a form suitable for storage. This time is called the *aperture time* and relates to the amplitude resolution of the conversion. The sampling error ΔV is equal to the amount that the input voltage, V, changes during the aperture time t_α.

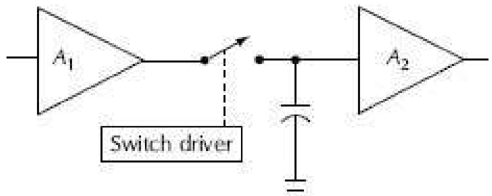


Figure 28-17. A commonly used sample-and-hold circuit.

$$\Delta V = t_\alpha \frac{dV}{dt} \tag{28-3}$$

For a sinusoidal input with peak amplitude A

$$\Delta V = t_\alpha \frac{d}{dt} A \sin \omega t \tag{28-4}$$

$$\Delta V = t_\alpha A \omega \cos \omega t$$

where,
ω is $2\pi f$.

The rate of change of voltage is greatest at the zero crossing when $t = 0$

$$\Delta V = t_\alpha A \omega \tag{28-5}$$

Expressing this error, e, a fraction of full scale,

$$e = \frac{\Delta V}{2A} = \pi f t_{\alpha} \tag{28-6}$$

where,

V is voltage,

A is peak amplitude,

f is frequency,

t_{α} is aperture time.

As an example, a 20kHz signal samples to a resolution of 16 bits (1 part in 65,536 or 0.0000152) requires an aperture time of 0.0000152/20,000π, or 0.24ns. This is a very short time interval for an analog-to-digital converter to operate in. A sample and hold circuit is used to preserve the voltage long enough for the conversion to take place. The aperture time of the system becomes the open switch time of the sample and hold rather than the conversion time of the analog-to-digital converter (ADC).

28.5 Analog-to-Digital Conversion

Details for the large number of analog-to-digital conversion methods are outside the scope of this chapter, but the efficiency with which it is accomplished is so important to the success and acceptability of a digital delay or reverberation system that an overview of the common conversion principles is useful.

28.5.1 Pulsed Code Modulation

Pulsed code modulation (PCM) uses a number to represent the value of each sample. The continuously varying analog signal is divided up in time by *sampling* and divided up in amplitude by *quantization*.

The quantization resolution is defined by the number of bits used in the binary number and defines the amplitude resolution of the signal. The number of possible states for a number with n bits is 2^n. For a 16-bit number, there are 2^{16} or 65,536 different voltages that may be represented. For a 1V peak-to-peak signal, this is equivalent to a 30µV resolution. An error in the representation of the analog value results because there is a range of voltages that yield the same output code. This error is called the *quantization noise* and is given by

$$Q = \frac{A}{2^n} \qquad (28\text{-}7)$$

where,

Q is the smallest analog difference that can be resolved by the converter,

A is the maximum amplitude,

n is the number of bits.

Another way of expressing this error is as the dynamic range of the converter.

$$\begin{aligned} DR &= 20\log 2^n \\ &= 20n\log 2 \qquad (28\text{-}8) \\ &= 6.02n \end{aligned}$$

A 16-bit coding system will therefore have a dynamic range of 6.02^{16} = 96dB.

The multibit binary word represents the amplitude of samples at regular intervals, usually in *twos complement* form. In this scheme the codes vary between 2^{n-1} and $2^{n-1} - 1$. The *most significant bit* (MSB) indicated the sign, with all negative values having MSB = 1. The code is often used in it fractional form, where the numbers represent values between −1 and 0.999.

28.5.2 Delta Modulation

Delta modulation is based on whether the newest sample in a sequence is less than or greater than the last. A delta modulator produces a stream of single bits representing the error between the actual input signal and that reconstructed by the demodulator.

A simple delta modulator, as shown in Fig. 28-18, consists of three parts: a comparator whose output is high or low depending on the relative levels of the input signal (SST) and the reconstructed signal $y(t)$, a D-type flip-flop that stores the comparator output under control of a sampling clock, and a reference decoder that integrates the binary output to reconstruct the signal $y(t)$. The demodulator is a simple integrator circuit with the same characteristics as those used in the reference path of the modulator. It will reconstruct the reference signal $y'(t)$, which is a close approximation of the original input.

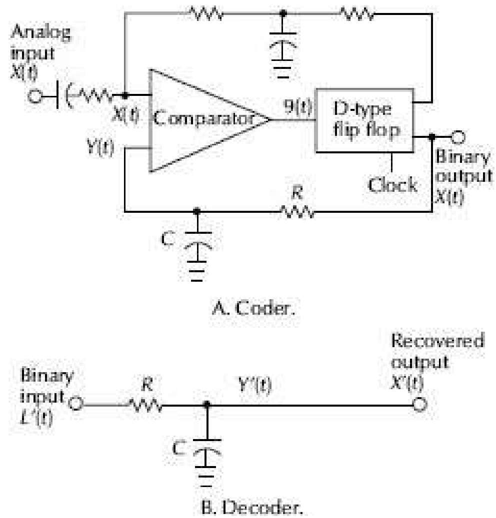


Figure 28-18. A delta-modulator system.

The simplicity of the coding and decoding schemes has resulted in use of the delta modulator for communication and motor control applications. The simplest integrating network consists of a resistor and a capacitor, but the quantization noise from this is quite high so more practical systems use double integration in the reference path.

Distortion in delta modulation occurs if the rate of change in the input signal is greater than the maximum rate of change of the output of the integrator. The maximum rate at which a sinusoidal signal, $A\sin\omega t$, varies is $A\omega$. The maximum required charge in voltage per sample is therefore

$$\Delta V = A\omega\Delta t \qquad (28\text{-}9)$$

The value of ΔV relative to the maximum amplitude A defines the amplitude resolution of the system. The *SNR* is proportional to the sampling frequency and inversely proportional to the signal bandwidth. When presented with a fixed input, the delta modulator will hunt for the value by changing the output between 1 and 0 every sample. The resulting output is a tone of magnitude ΔV at the sampling frequency and is called the *idling noise*.

Delta modulation is more immune to errors in storage or transmission than PCM. A single-bit error in the output has a resulting error in the analog signal of ΔV In a PCM system a single-bit error could cause an error of up to half the full-scale value. When compared to a PCM system in term of bits per second, delta modulation will have comparable dynamic range but a smaller frequency range. At lower bit rates, delta modulation can have a better *SNR* and dynamic range than a PCM system and this has implications for delay lines, where the total number of bits that must be stored can be reduced for the same quality of signal.

28.5.3 Sigma-Delta Modulation

By reorganizing the sequence of operations, the delta Modulator becomes a *sigma-delta modulator*. A firstorder SDM as shown in Fig. 28-19, has one low-pass filer integrator in the signal path and a direct feedback path to a summing point that produces an analog error signal. The comparator of the delta modulator is replaced with a quantizer, which is a comparator against a fixed zero reference. Demodulation is accomplished using a low-pass filter as in delta modulation.

Both delta modulators and sigma-delta modulators use sampling frequencies much larger than the Nyquist frequency, typically of the order of 100 times. This places the quantization noise energy at very high frequencies where it can easily be removed by filtering.

28.5.4 Decimation

The process of sampling well above the required Nyquist frequency is called *oversampling*. The objective is to cause the modulation noise inherent in the sampling process to appear at frequencies further removed from the audio signal so that it can be more easily removed by filtering. The high SNR of an oversampled system can be preserved while reducing the overall bit rate by the process of *decimation*.

The quantization noise from PCM encoding decreases by 6dB for every bit added but decreases by only 3dB for every doubling of the sample rate. Decimation of the oversampled PCM data can result in a large reduction in the overall bit rate. Decimation filters are used to achieve this and can be thought of as performing an interpolation on the existing data to fill in the additional bits in the output.

A sigma-delta modulator can have as much as a 15 dB SNR improvement for a doubling of the sample rate. Decimation of the SDM signal can be used to convert the single bit data stream to a multibit PCM format suitable for storage in RAM and processing by DSP.

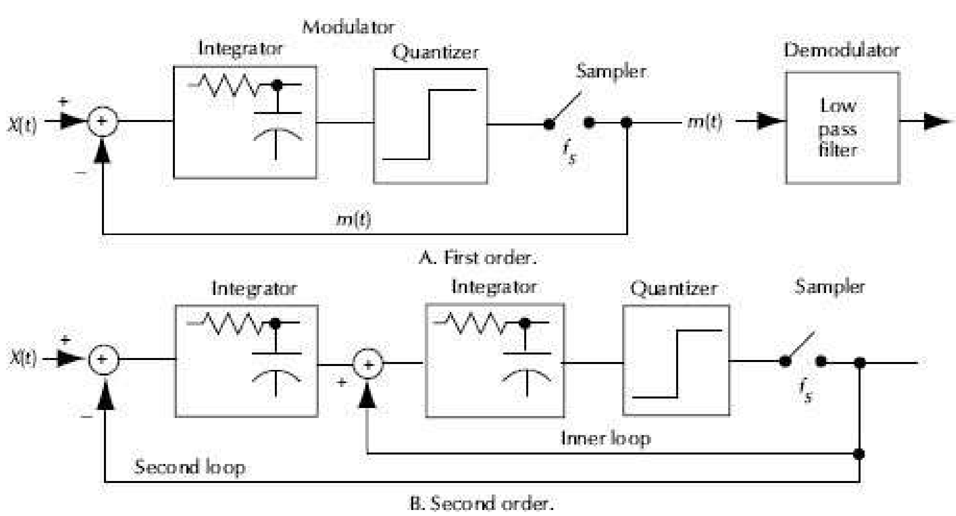


Figure 28-19. Sigma-delta modulators.

References

1. H. Haas, “The Influence of a Single Echo on the Audibility of Speech,” *J. Audio Eng. Soc.*, Vol. 20 (March 1972), pp. 145–159.

2. L. G. Heller, W. H. Chang, A. W. Lo, “A Model of Charge Transfer in Bucket Brigade and Charge-Coupled Devices,” *IBM Journal of Research and Development*, Vol. 16, no. 2, p. 184, 1972.

Chapter 29

Consoles, DAWs, and Computers

by Steve Dove

29.1 Introduction

Digital won.

By far the majority of professional audio manipulation, processing and recording, the province of consoles and work stations, is digital.

Yet analogue still has it surrounded.

Cognizant of both, this chapter describes analogue signal processing elements, and the architecture of systems in which they were and still are used, since for the most part the work-flow in digital environments follows almost exactly, of operational necessity. It's what people need to do with audio—the reasoning behind both is the same. Additionally, the workings of digital consoles are explored in the context of an actual practical design which has evolved over the years with freshly developing technologies.

Within this, the point is made that DAWs (Digital Audio Workstations) are little more than a specialized case of digital console; their internal workings and architectures are highly recognizable as such; that the recording medium is highly integrated is the main technological marker. Consoles abound in forms quite unlike the traditional big sea of knobs of yore, and yes, DAWs are the present majority embodiment of consoles. Productions huge and modest are regularly done ITB ("In The Box") with a screen and a mouse—indeed, nearly everywhere operations don't require immediate access to controls; those that do, mostly live, still use larger control-surfaces. But consoles they are; they're just hiding in unfamiliar shells; their schema are, despite outward appearances, directly traceable to traditional audio architectures. (Throughout this chapter the word "console" will—unabashedly and in the absence of any other generic term—be used to describe an audio control system used by an operator, regardless of its size or processing type.)

Far greater than any supposed analogue/digital divide, the real changes have been in how the audio is controlled; the rationalization of control surfaces is now normal and indeed welcomed (such heresy was brutally fought-over not long since), while the surface—whether it be glass or knobs-and-faders—is entirely a separate matter to that which actually processes the audio.

A range of described console and audio workstation arrangements (architectures) gives enough clues to analyze how any encountered system actually operates.

Description of circuitry and techniques is less theoretically driven than practically derived, with no apologies given for blow-by-blow analyses of salient elements of commercial audio system designs. It is hoped that this will augment and lend perspective to earlier descriptions of typical circuit blocks. An overview of digital signal processing as applied to consoles will, as deeply as it is possible to go before nasty equations arise, give an insight into how these things work; far from voodoo, in some respects architecturally there are strong similarities to analogue solutions. Similar to the way that a real analog console design is dissected and explained in the following pages, a real digital console design is broken down for overview and analysis.

All the fancy processed audio is pointless if you can't use it. This means moving it around. Audio transport is undergoing a revolution, and an overview of schemes from humble AES3 to the ascendant AoIP (Audio over Internet Protocol) are discussed.

Traditional analogue consoles spectacularly evolved and matured operationally in the same era as similarly burgeoning analog technology, incidentally driven by a music scene exploding in a brilliance arguably not seen before or since. Consoles and their constituent signal-processing elements have nowadays migrated into and are being emulated digitally. Yes, a huge amount of the engineering of digital mixing consoles is in accurately recreating foibles inherited from their analog ancestors, good or ill. It behooves one to understand why things ended up the way they did, so that one can optimally progress into the new domain. It's called learning from history, only in this case the history is still very much alive, is in use daily, and still has its teeth.

Console Development

The establishment of consoles was a slow and gradual process. Similarly, systems—or preorganized arrangements of devices—evolved slowly, too. In most audio work the two are now considered as almost synonymous; the greatest departure from this is the inclusion of a console and recorder as parts of a system. But even then, there is no doubting that the console is the heart and substance of the system.

The history of consoles reaches back to the time when the recording process was purely mechanical, Fig. 29-1, followed by its electrical analog, Fig. 29-2, which included a source transducer (in this instance a microphone), a means of gain (an amplifier), and an output transducer (a disk-cutting head). It doesn't take a staggering amount of imagination to extend this system to embrace other applications: public address, acoustic enhancement of natural sound by electronic means, Fig. 29-3; disk replay, Fig. 29-4; and broadcasting by replacement of a simple electromechanical transducer by a radio transmitter, Fig. 29-5. The objective of the system is to facilitate the transfer of a signal from one source—be it a simple transducer or another system—to a destination.

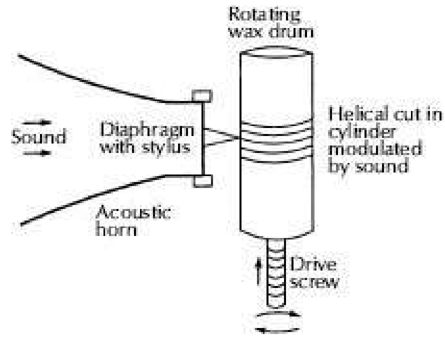


Figure 29-1. Mechanical recording or early drum recording.

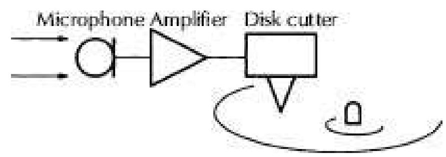


Figure 29-2. Electrical recording (disk cutter driven by electricity)

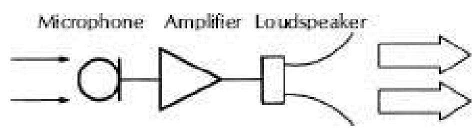


Figure 29-3. Public address system.

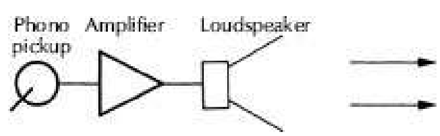


Figure 29-4. Public address system.

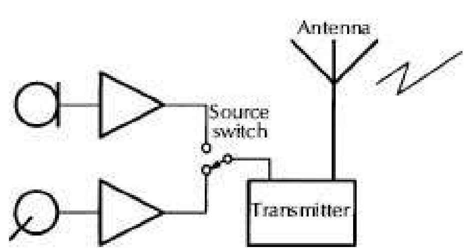


Figure 29-5. Simplified broadcast system.

Of course things get a bit more involved than that, and to demonstrate this complexity, the evolution of what is probably the most important subsystem to our industry—the recorder, mono, stereo, or multitrack—will be used to explain how it, almost single-handedly, made everything as complicated as it is today. Disks were permanent. You got it right or you didn't. Tape at least gave the chance of one more take.

Mixing in the early days of system development was surprisingly easily achieved—just connecting the outputs of the various source input amplifiers together did it perfectly adequately. It's important to understand that the technology of the day facilitated this simplicity far more, paradoxically, than today's gear. Tube amplifiers, such as were then used, needed to be terminated at their outputs by a specific impedance for proper operation, which for reasons discussed later was universally 600 Ω floating balanced. By simply connecting amplifier outputs together, a mix of sources was achieved, provided each of the source amplifiers saw 600 Ω. It was only a very minor step for interspersed networks to become constant-impedance variable attenuators, usually in the form of rotary controls. The pot (from potentiometer) or fader was born. The ability to create a balance of sundry sources for the chosen destination is perhaps the most recognized feature of the console and its system. Convention and common sense rule this as the main signal path, and other paths are subsidiary or auxiliary to it.

29.2 Auxiliary Paths

29.2.1 Monitoring

Take the example of Fig. 29-6, where a single microphone is being laid on a recorder. It's operationally necessary for the system operator to hear the signal going to the recorder with headphones or a control-room monitor loudspeaker. To facilitate this requirement, a parallel feed is taken off the machine input to the operator's monitor. Monitoring is perhaps the most important of the auxiliary signal paths; upon it is based the qualitative decisions of the nature of the signal in the main path. It is the reference.

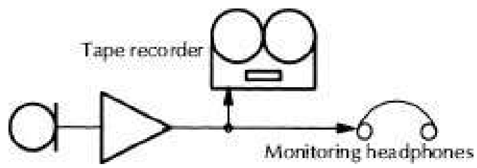


Figure 29-6. Simple microphone-to-recorder monitoring (source only).

Fig. 29-7 applies a small extension to the basic monitoring path in the form of a source/replay switch, enabling operators to hear the aftermath of their efforts. If the recorder has separate record and play signal paths, they can even toggle between the two while actually recording for immediate quality assessment. The monitoring section is born.

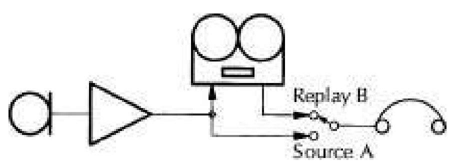


Figure 29-7. Recorder monitoring (source and tape).

29.2.2 Prefade Listen and Audition

When a multiple-source system is established (similar to Fig. 29-8), another monitoring requirement, prefade listen (PFL) and audition, is required. Imagine the case of a radio broadcaster, where the sources consist of disk replay units and microphones; it's an obvious necessity to be able to listen to a source prior to its being put on air to check that:

1. The microphone is set at the correct position, level, or even working!
2. The required section of a recording is cued up or ready to play.

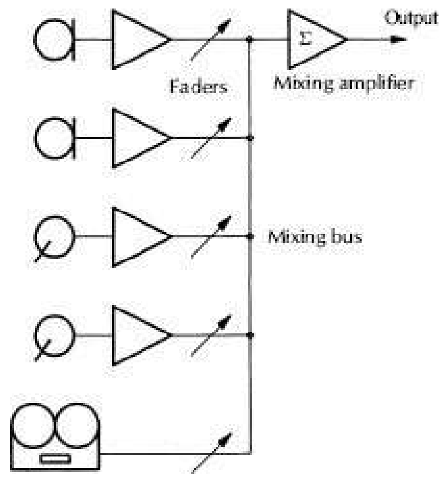


Figure 29-8. Multisource mixer.

There are two basic methods of arranging this prehear function, as shown in Fig. 29-9. They owe their existence primarily to slightly different operating practices on opposite sides of the Atlantic. The first, Fig. 29-9A, involves switching the signal immediately prior to the fader on the selected source path into the monitoring chain. This is called prefade listen (PFL). A useful but not immediately obvious virtue of this arrangement is that it is possible to listen to a channel's contribution to a mix of which it is part without disturbing that mix. It is, therefore, a nondestructive monitoring function. The alternative method shown in Fig. 29-9B consists of removing the required channel (postfader) from the mix and placing it onto a second parallel mix facility, commonly called audition or rehearse; it is possible in this mix to emulate exactly what would happen in the real mix without upsetting the presently active mix. A disadvantage of this method is the inability to use the function when the channel is live because it disrupts that source and prevents it from going to the mix. It is a destructive monitoring technique. Each method has its virtues, though, and most modern consoles use both techniques to varying extent. That said, all but really small American broadcast consoles nowadays employ a PFL-type cue function, the audition/rehearse bus being arranged to be a secondary mix bus independent of but in the same vein as the program bus. The name often remains as an echo of its original function. Postfader monitoring, however, lives on in large production consoles as in-place stereo monitoring, described later.

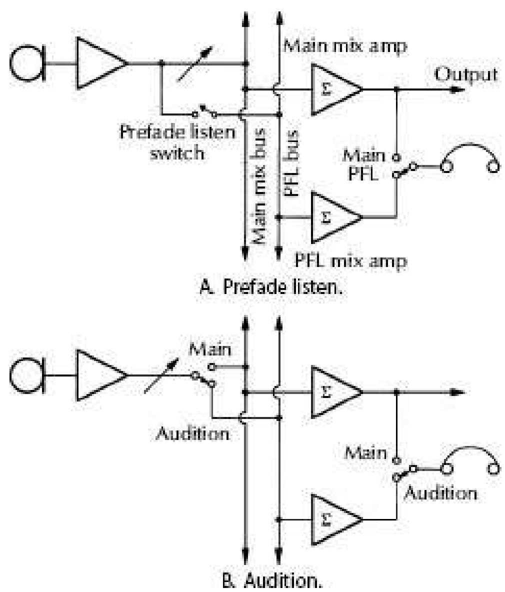


Figure 29-9. Prefade listen (PFL) and audition.

29.2.3 Overdubbing and Foldback (Cue)

While broadcasting lends itself to explaining the need for individual channel monitoring, original material generation onto a recorder serves best to explain another crucial auxiliary signal path.

It didn't take long before studios were using more than one recorder in a technique known as overdubbing. Briefly, this involved recording a backing track (for instance a rhythm section) on one machine and then playing that back while vocalists sang or soloists played along with it; the whole was mixed together and recorded on a second machine (bounced) as shown in Fig. 29-10. This could be carried on until the subsequent machine-to-machine generation losses became too objectionable (although that never seemed to bother many early producers!). (Generational losses occur because the tape machines of the era were less than perfect, what came out of them being noticeably ropier than what went in!) Naturally, it was essential that the musicians in the studio were able to hear via loudspeakers or headphones that to which they were supposedly playing along; this is where foldback (cue) comes in. In its simplest form, it could be a straight derivative of the main mix output, since this output has basically everything necessary in it. This system, however, has a few shortcomings due primarily to conflicts between what the final mix is intended to be and what the artist(s) needs to hear to perform satisfactorily. A prime example of this dilemma is in the recording of the backup vocalists sections; usually they take a fairly minor part in a mix, being balanced well down. Contrary to this is the need of the vocalists to not only hear the track played back to them but to hear themselves sufficiently well—usually enhanced—to pitch and phrase themselves effectively. These conditions are next to impossible using a final mix. A solution lies in Fig. 29-11 where a separate balance of the relevant sources is taken and fed separately to the performers, giving them what they most need, a foldback mix. The takeoff for the foldback feeds is almost invariably prefader so that the artist's balance remains unaffected regardless of what modifications may be necessary for the main mix.

29.2.4 Echo, Reverberation, and Effects Send

The move (regrettable as it may seem) from natural performing acoustic environments to the more cultured, drier, closer mic'ed techniques brought with it many problems attendant to the advantages. How do you make a sound seem as though it was recorded in a great concert hall if it was done in a small studio? Reverberant chambers were an initial answer, being relatively small rooms acoustically treated to have an extended reverberation time (bathroom effect). Driven obliquely at one end or corner by a loudspeaker(s) and sensed by a microphone(s) at the other end, which is amplified and balanced into to the main mix, a fairly convincing large room reverberant effect can be achieved. Simplistically, all that's needed to feed the loudspeaker in this room is a derivation of the main mix, but similar to the problems with foldback mixes, artistic judgments dictate something more complex. Some instruments and sounds benefit greatly from being dry (most of a drum kit, for example), while others—vocals, in particular—sound quite dry, cold, and uninteresting. A means of adjusting the relative amounts of artificial reverberation due to various sources would be beneficial. Fig. 29-12 shows a small console system complete with an echo send mix bus (echo in this sense including reverberation); the echo return is brought back into the main mix just as any additional source would be. Echo feeds are nearly always taken postfader, keeping the reverberation content directly proportional (once set) to the corresponding dry signal in the mix regardless of the main channel fader setting.

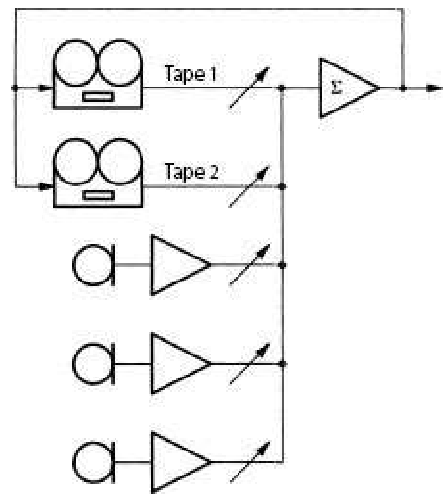


Figure 29-10. Overdubbing/bouncing, a previous microphone mixer recorded on tape 1 may be played back along with a further microphone mixer onto tape 2 and vice versa.

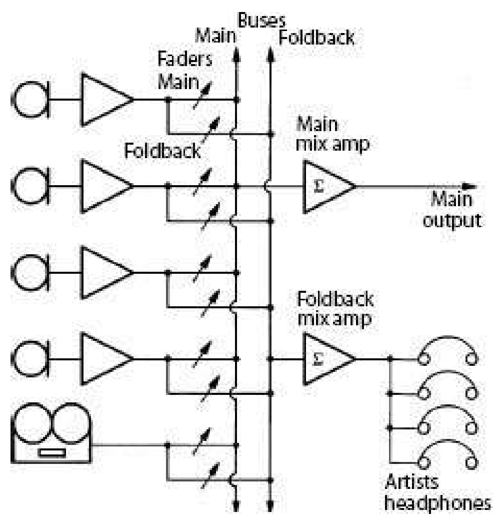


Figure 29-11. Foldback mix.

Today, any number of foldbacks and effects sends are in use as toys (effects boxes) proliferate; long gone are the days when one of each feed was sufficient.

29.2.5 Communications (Talkback)

An often mentally mislaid but crucial console auxiliary path is talkback; that is, the ability of the console operator/producer to talk to various people involved in the recording. The primary need for talkback is to be able to communicate with the studio area that is necessarily acoustically separate from the control/monitoring room. Since there are already foldback feeds going to the studio area for performer cues, it makes sense to talk down these feeds, which is called talk to foldback (talk to studio), known in broadcast as IFB (Interruptible Fold Back). Another useful function in this vein is slate. This curiously named facility allows the operator to talk into the main mix output and thus onto the recorder for track and take identification purposes, sometimes in conjunction with a videoed clapper-board (which used to be made of slate).

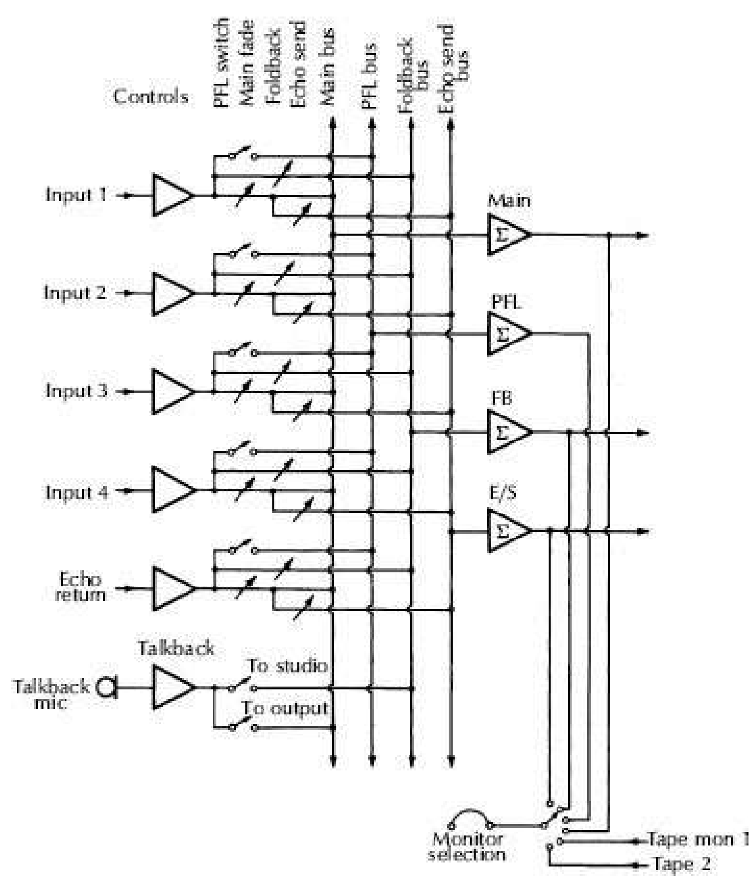


Figure 29-12. Small mixer system showing auxiliary functions.

29.2.6 Combined Auxiliaries

In summary, a usable console has to have several signal paths in addition to the main mix path. These include overall and presource monitoring, prefader adjustable foldback feeds, postfader artificial reverberation feeds, and communication (talkback) feeds as shown in Fig. 29-12.

29.3 Stereo Consoles

Stereo predated multitrack recording. Technically the required console techniques were not very far removed from those just described. Assuming the same bouncing (machine-to-machine overlay and transfer system described earlier in the section on overdubbing), stereo just means two of everything in the main signal path.

29.3.1 Panning

Panning is the technique of positioning a single monophonic source within a stereophonic image. It isn't true stereo; true stereo can only be achieved from coincidentally aligned microphones. Instead, it is panned mono. Simply, the ear is deceived by pure level differences between the left and right paths of a stereo pair into perceiving differing image position; fortunately for the entire industry, this is a trick that works rather well and is quite simply realized.

Complementary attenuators (one increasing and one reducing attenuation, with rotation) feeding the L and R mix paths from a mono source is the most common method. Fig. 29-13A illustrates this system. The pan pot is usually inserted after the source fader.

An alternative arrangement is shown in Fig. 29-13B. Here the pan pot is inserted prior to the fader; a ganged matched fader is required with this method. This arrangement can be useful when stereo PFL is required, although there are other ways of achieving stereo in-place monitoring for sources that will be described later.

29.3.2 Auxiliaries for Stereo

Auxiliary paths remain largely untouched by the upgrade to stereo of the main mix path; the monitoring section stays just the same in systemic function (but obviously with two paths instead of one to cope with stereo feeds). Both the prefader foldback and PFL takeoffs are still in mono. The postfader echo-send feed is usually taken out before the main path pan pot, so they remain mono, but the returns pass through their own pan pots such that the reverberant image may also be spatially determined in the mix. It's become normal practice to make echo-send feeds stereo in their own right, Fig. 29-14, via their own pan pots' mixing to two outputs. Many reverberation rooms, plates, and boxes are capable of supporting a diffuse stereo field. The purpose of this is to excite the reverberant chamber (or plate or springs or little black box) spatially, conjuring a more solid and credible reverberative effect in the main mix. If a panned echo-send output isn't available, it's common to use a pair of separate postfader feeds and juggle the levels between them.

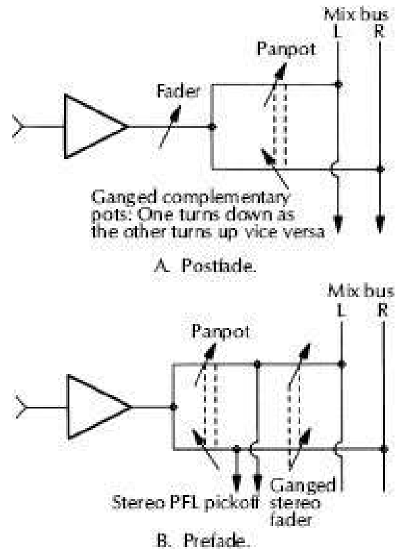


Figure 29-13. Panpots.

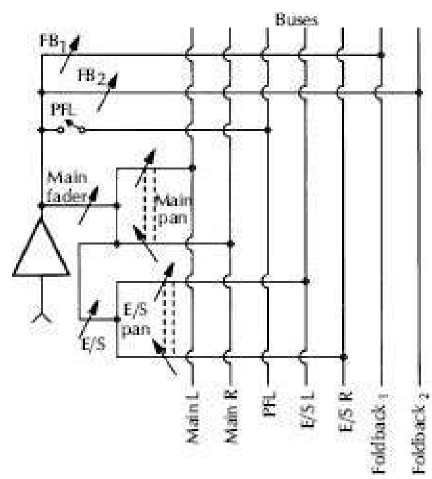


Figure 29-14. Channel feeds showing foldback and stereo echo-send feeds.

29.3.3 Multiple Effect Feeds

There is a whole gamut of electronic toys applied to mixdown to achieve specific sounds: Harmonizers™, delays, flangers, phasers, automatic panners, artificial reverberators of various sorts, and so on, many generic versions of which may be built in to the console. These all need to be fed from their own effects mix paths. Similarly, studio foldback mixes have grown more profuse with changing music and increasing musician sophistication and awareness of studio techniques; consequently, the number of auxiliary mixes within modern consoles has risen dramatically. A rationalization of this is to make those auxiliary mixes multipurpose, usually by allowing them to be switchable between prefader and postfader feeds on their appropriate sources. A smaller number of buses are needed in this way; during the recording process the emphasis is on many foldback (prefade) feeds for the musicians in the studio and maybe one or two toys to spice up the monitoring. On the other hand, during the overdubbing and mixdown phases very few foldbacks (if any) are needed but every bus will be set to postfade and laden with effects. Additionally, in broadcast it is commonly necessary to talk back down some or all of these mixes individually (interruptible foldback, IFB) as well as individual DCOs (Direct Channel Outputs). Large modern consoles for live applications often have many stereo foldback auxiliaries, driven by the trend toward the (usually wireless) in-ear headphones beloved by performers.

Toys, or sometimes more prosaic signal processing such as compression, are also commonly introduced as inserts within a channel's signal path, usually pre-fader. In DAWs where toys are rampant and easily introduced by way of plug-ins, many processes are often cascaded within a channel input source.

29.4 Dawning of Multitrack

Multitrack operation is when a number of separate parts of the recording are laid onto separate tracks on a recording machine and subsequently remixed down onto another machine (be it mono or stereo) or even interim bounced onto spare tracks on the same machine.

Stereo recording using two-track tape technology seemed to many to be the zenith of professional audio. Many will argue the point even today. There is inescapable evidence of the validity of that opinion in that some of the finest stereo recordings, especially of classical and jazz works, were done using fundamental microphone techniques straight onto two track. Even in the field of pop records where things were bounced mercilessly the final master still represents the first generation of the last overdub. (In retrospect that is an advantage over contemporary multitracking where the master is at best the second generation of everything.) (Remember, tape generations meant serious quality loss, unlike digital bounces today.)

Multitrack soon reared its head(s?) in the early 1960s—initially as three track and four track across 1 inch tape; there are those who regard that as the zenith. More tracks reduced the number of intermachine bounces, but they still added up! Three tracks afforded a great advantage over two tracks for modern music producers at the time. Two-track recordings were always hampered by the need to make sure that all the earlier things done in a bouncing sequence were right to begin with; there was no chance of subsequently altering them. Three-track recordings, typically in a Track/Vocals/The Rest format, took a little of that pressure away. Already producers and performers were taking advantage of the multilayered production approach to take the heat out of recording; it was no longer necessary for everyone from lead vocalist to third trianglist to be present all at once for a momentous occasion. Bits could be done one at a time. The extension to this given by multitrack is simple to see: the more tracks, the smaller those bits need be and the fewer things needed to be incontrovertibly mixed. Putting off the day of reckoning—the final mixdown—is one of the strongest appeals of multitrack. This, indeed, has led to a curious polarization in the business; tracking, the laying down of individual tracks, is typically done in entirely different studios or environments to mixing (and often purely in analogue). And remixing, the construction of yet different mixes from the same basic tracks for specific genres such as dance mixes, has spun off into yet another subindustry. So much for making spontaneous music.

29.5 Grouping and the Monitoring Section

Each signal source in the console needs some routing to determine the recorder track on which it is going to end up. It's a situation that hardly existed previously, since it was pretty sure that the mono or stereo output of the console was going to go straight to the respective mono/stereo inputs on the recorder(s). There were on a stereo console just two groups where all the sources were summed together; for multitrack as many groups as there are tape tracks are switch selectable from the sources—any source to any machine track. The alternative hard way, patching everything across on a jackfield, was and is exceedingly tedious, messy, expensive, and error prone.

Four-track recording set the mold for console design for many years. The monitoring section evolved. Fig. 29-15 can be compared to the simpler back end of a stereo mixer in Fig. 29-12. The main difference can be seen as the addition of an entirely separate mixer within the console just to handle the multitrack monitoring.

Fortunately, it's a fairly bare-bones mixer; it's all at high signal levels, and little, if any, gain is required except as makeup gain in the monitor mix bus.

While all these tracks are being laid, it's necessary to hear what has been done previously in the control room and studio. In the same way that source/return listen of stereo machines was needed, so each individual track of a multitrack needed similar treatment. It grew, though. Initially, as the number of tracks per machine increased, the number of mixer groups increased correspondingly. Each group had its own A/B switch relating to that individual console track output and the associated machine return, with its own level and pan controls feeding an altogether separate stereo monitor mix. This new monitor mix appeared as another source on the main monitor selector. This, alas, was insufficient. Foldback prefade mix feeds no longer became a luxury but a necessity, since the desk stereo output or a derivation thereof could no longer be relied upon to be even roughly what the artist needed to hear. There was no proper console stereo output at any time other than mixdown. Foldback feeds were added to the monitor system on each group. Effect sends were also added, just to help the monitoring sound pretty.

The monster has split itself amoeba-like into two entirely separate signal-processing systems: the main mixer and a monitor mixer. A curious situation occurs: the mix used for monitoring during the original multitrack recording had to be transferred over to the main system entirely at some time for mixdown. Ordinarily, tape-machine returns were not only brought back into the monitoring section but also tied to high-level line inputs on the main mixer section. The remix took place using those channels into the main stereo mix bus.

Perhaps the first major rationalization (which occurred long after many conventional X-input, 24-group, 24-monitoring consoles had been made) was a result of the realization that few people actually needed 24-group faders sitting there full up, collecting dust. Losing them instantly avoided a normally unnecessary gain-variable stage in the signal path, which, if maladjusted, could upset noise or headroom performance.

Individual channel outputs together with a much smaller number of stereo mixing subgroups—usually four or eight pairs—which could again be routed to any of the multitracks, proved easily as flexible. But still there was duplication of monitor buses and main stereo mixing buses both with their attendant effects and fold-back feeds rarely being used simultaneously. At last the dawning of the realization that the pair, that is, the monitoring and stereo mastering buses, could be one and the same thing. In-line monitoring recording systems had come to fitful fruition. The in-line console includes all of a recording channel's processing and all of a machine return's monitoring controls within one channel strip; it allowed efficient sharing of controls, processing, and mixes between those paths, maximizing their full utilization through the tracking, overdubbing, and mixing phases of a production. It also did away with the separate multitrack monitoring section, the mixer within the mixer that nearly doubled the physical width of conventional consoles. In-line consoles were somewhat more complex to learn, but overall the advantages were huge with few compromises.

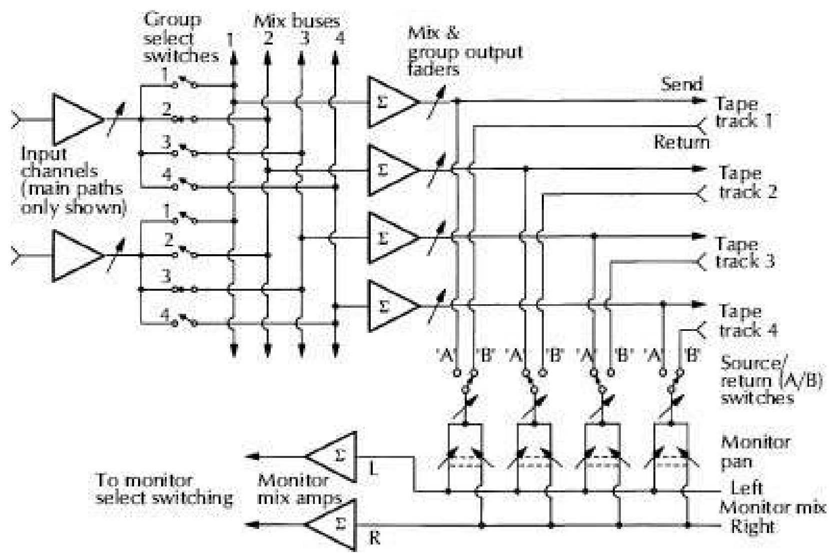


Figure 29-15. Four-track monitoring.

We all have to be thankful for the cranks and visionaries along the way (often the same) who have manipulated or shocked the industry into grudgingly lurching back into step with technology's capabilities. These developmental milestones represent significant plateaus of thinking that form the basis of today's console concepts. In-line is a classic example.

Subgrouping and Output Matrices

Particularly in live applications, e.g., sound reinforcement or broadcasting, the ability to make a subgroup of related sources—say drum mics, bass, guitar, keys, backing vocals, lead vocal (each of which can have many sources themselves)—and then rebalance them together is a valuable addition. (This means that instead of having to gingerly pull down the 10 mics on a kit without destroying the previously hard-won balance, a single fader on that subgroup can be moved instead.) These are real subgroups, so called because a real mix of real audio sources is created, rather than a similar overall result happening by way of a VCA subgroup (described fully later) in which only the fader movements are tied. An output is available just containing the subgroup member sources, useful if processing (EQ, dynamics, etc.) is required over them exclusive to other sources such as auxiliary sends for the addition of effects solely to the subgroup and remixing. This latter is a particularly powerful use for these subgroups; feeding them as sources into a downstream mixer, often called a matrix mixer, from which an often large number of matrix output mixes are created.

Again using sound reinforcement as an operational example, the many performers on stage all need to hear both themselves and the rest of the performers either in monitor speakers or in personal earpieces; the trouble is, the balance that each of these people needs is typically entirely different! Using the individual remix capability on each matrix output fed by the earlier-created subgroups, many different mixes of the same few subgroups are possible, hopefully resulting in a calm stage.

29.6 Console Design Developments

Two distinct considerations interplay in determining the ability of a console to fulfill a given application. These two—the system and the electronics—have entirely differing parameters that need to be defined but are, nevertheless, completely indivisible.

The electronics, as much as being designed to perform required functions, must be very carefully designed not to be a major influence on the sound of the console. Most causes of sonic disturbance can be attributed or predicted, and still dubious circuit configurations can be avoided altogether. There is a groundswell of designing sonic character back into analogue electronics; this after generations of striving for accuracy and neutrality is a touch alarming. The good news is that consoles (unless otherwise eccentrically contrived) are still expected to be neutral, the color being acquired by the gallon in external toys or plug-ins. To that end, unless specifically stated, the electronics described here are intended to be neutral sounding. To the shock of some purists, commonly available integrated circuit operational amplifiers are generally used throughout the designs in this chapter. The reasons why (other than the obvious convenience), together with the reasons why they acquired a bad reputation, are treated in depth in section 29.7.

Operational amplifiers (op-amps) revolutionized the concepts and systems capability of full-performance audio consoles. Their use allowed system elements to be thought of, designed, and implemented as building blocks. This simplified matters considerably, but it also raised the worry that console design could be relegated to a do-it-by-numbers routine. Device and circuit topology idiosyncrasies, subtleties, and the entirely separate science of getting heaps of individual system elements to behave successfully as a total console precluded this!

Fortunately for the console industry, the large proportion of the current console manufacturers started off in life as small groups of musicians and studio engineers furtively constructing mixers for their own ends, resulting in grass-roots system design owing everything to immediate operational needs. Continuing in this vein in production, the manufacturers are listening to and, most importantly, relating to customer needs because they've played this game for themselves.

Once, systems and mixers as such didn't exist. All the bits of electronics used in the control room sat there with all their inputs and outputs accessible by way of a jackfield for the prosperous or by small screwdriver and sore knees for those who weren't.

Mixing sources was accomplished by directly paralleling amplifier outputs (possible because all the old tube gear was designed with a particular termination impedance in mind, usually arranged to be a conventional balanced 600 Ω) and either hoping or arranging that the destination had enough gain to make up accrued paralleling losses. Crude as that may seem today from an engineering viewpoint, it has a sheen of pure elegance. An amplifier was just that, a box that had a balanced 600 Ω source and termination impedance. It might also have an alternative bridging (> 10kΩ) input terminal and a selectable amount of gain offering universal application from microphone amplifiers through mixing amplifiers to headphone amplifiers. To do more things, more boxes were added. Equalizers and limiters, a treasured few if there were any, were similarly universally applicable. Variable-level control was again attained by true balanced 600Ω source and termination, via studded rotary attenuators. The utter beauty of the systemless studio was that anything could go to anywhere via anything else and be mixed or distributed at any point on the way. (Many, many decades passed until we could say the same thing again.)

Soon enough amplifiers were hardwired to attenuators and designated specifically a microphone amplifier, and a system had been created. Some of these together with a mixing gain makeup amplifier were thrown in a box. The mixer was born.

It was downhill therefrom, with ever-increasing numbers of system elements being tied together in increasingly knotted manners in order to maintain some kind of flexibility. Perversely, a system can be defined as a means of reducing the ultimate versatility of its constituent parts.

Once a mixer was accepted as a system element itself, the problem set in further. There was no need to provide for convenient connection of its internal interconnections to the outside world, so the balancing transformers disappeared, and more economic alternatives to the stud attenuators operating at more convenient internal impedances evolved. By a more positive token, the electronics were gradually becoming optimized for the specific functions to which they were designated, such as the microphone amplifier and the mixing amplifier. (The question nagged as whether a universal amplifier, by now all but obsolete, could be optimized for all the varying requirements; it wasn't, and didn't.) Still, at least all the inputs and outputs of the mixer were conventional. This held true until the slow demise of vacuum tubes in professional audio.

29.6.1 Transistors

Transistors were justifiably unpopular for a long time since they placed numerous limitations on design, forced changes to design habits, and until long after their introduction really weren't very good. The headroom was severely limited because of the low supply voltages that could be applied to the early devices. They were noisy. They were prone to thermal instability. They weren't terribly or consistently gainy, all of which made designing them in dodgy. (The "Germanium Period" is one most in electronics choose to forget; silicon was a savior.) The lower operating impedances and differing modes to tubes took some getting used to and, when they saturated, they actually clipped rather than gracefully bending (a characteristic of tubes that people had known, loved, and frequently taken advantage of—even now). To realize a reasonably low stage distortion, many transistors in compound configurations using heavy amounts of negative feedback were used—a far cry from a single tube stage operating virtually wide open with little feedback. This gave rise to a peculiar phenomenon that sounded as if it hailed from science fiction—zero impedance.

The heavy negative voltage feedback employed around transistor circuits could be made to render the output of an amplifier insensitive to varying load impedances; they would deliver the same output voltage level almost regardless of their termination impedance. This eliminated termination problems with the attendant worry of compensating in level for differing load hookups. With the exception of long line feeds, 600 Ω terminations were as good as dead. High-level balanced inputs were now almost exclusively bridging; they had a sufficiently high impedance (usually >10kΩ) not to disturb the level of the source to which they were tacked on. For better or worse, it has become the conventional studio interconnection technology. It took a while for a distinction and separate level specification for the two technologies to be accepted.

29.6.2 Level Specifications

The original transmission line level specification referred to a power level of 1 mW regardless of impedance. This was 0dBm. It was a universal specification applicable to any signal of any frequency being transmitted along any length of wire for any purpose at any rated impedance, and it is used extensively in radio-frequency work and other things entirely unrelated to audio. The dBm definition is sacred and can't be changed. Zero dBm in a 600 Ω load works out to 0. 775Vrms; this was adopted de facto as the reference for use in general audio work. With zero impedance technology, although the working voltage is specified, the impedance isn't. It can be anything, but the power (as measured in dBm) necessarily varies as a result; for instance, 0.775Vrms across a 100 Ω load is +7.78dBm while across 10kΩ it would be minus 12.22dBm. But it's still 0.775Vrms.

The reference level for zero impedance thinking is a voltage, and the one chosen is the familiar 0.775Vrms with which everyone was historically used to dealing. That voltage is distinguished as 0dBu. Some have tried to impose a universal reference based around a voltage level of 1V called the dBV for audio, which is easily divided by 10 but has proved sufficiently confusing to anyone brought up on the dBm that it is now all but dead.

But wait! There's more! The ubiquitous VU meter when implemented as intended imposes a nominal system level of +4dBm (0VU +4dBm at 600 Ω), and in territories and market segments where the VU reigned, +4dBm (and latterly +4dBu) is still a common reference. And try as one might, it is impossible to ignore that there is more semipro recording and audio gear in use than real audio equipment and that generally uses the domestic level of −10dBu as a nominal reference. Glad that's all cleared up, then.

29.7 Operational Amplifiers in Consoles

Consoles utilizing integrated circuit operational amplifiers (IC op-amps) have suffered from a curious syndrome, collecting in earlier days a (sometimes deserved) dreadful reputation, which has stuck. This section is an attempt to explain the history, shortcomings, and attributes of IC op-amps from conception to present day, to point out how some shortcomings are overcome. It is also an example that this, along with most other technology, is well understood and quantified, the concepts if not the details having been defined many years ago, Fig. 29-16.

When ICs first came out (the Fairchild μΛ709, e.g., they were expensive, prone to oscillate, and had no short-circuit output protection.

At this stage in the game, discrete transistor circuitry ruled supreme in pro-audio while considerable vacuum-tube gear was still in use. Techniques expanded and ICs were tamed sufficiently to remain operationally stable, but little high-frequency loop gain remained to guarantee enough feedback to adequately reduce high-frequency distortion. Also, they were very noisy. Although their parameters could be set up to be acceptable for any set application and gain setting, the very nature of control in consoles is variable, so the devices almost inevitably ended up operating away from their optimum.

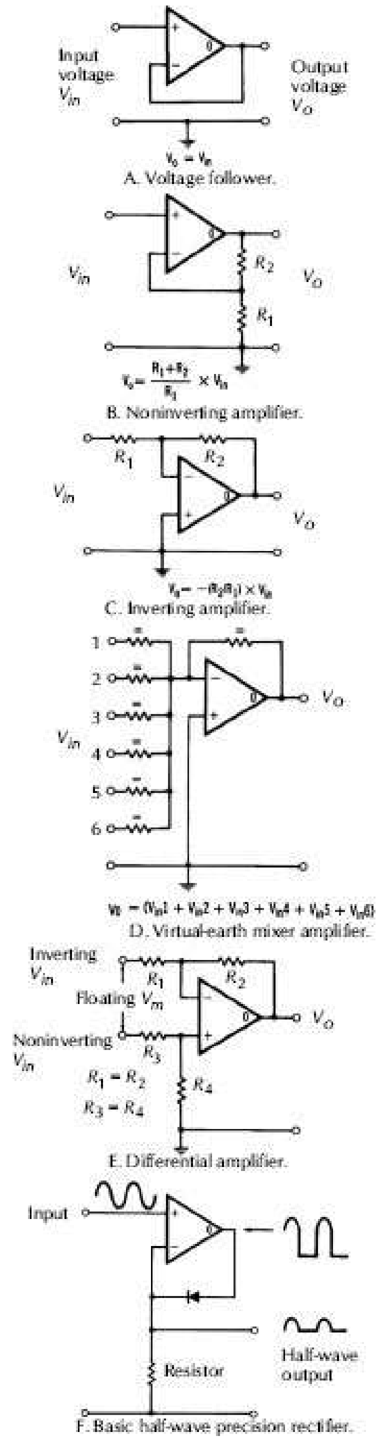


Figure 29-16. Basic op-amp configurations.

A new term entered the audio design vocabulary: compensation. Compensation is the brutal slowing down of the amplifier in order to stop rampant, screaming instability. Essentially it was accomplished by defining the bandwidth of the overall loop around the amplifier or a particular gain stage within the amplifier—or both. And typically robbing the device of its promise.

Hot on the heels of the μA709 came the now much loved and despised, but always revered μA741. Best known in its plastic encapsulated eight-pin dual-in-line configuration, it still took our industry many years to catch on to the fact that here existed a seemingly almost vice-free op-amp. Well, at least, it was free of some of the 709's vices. It was heavily internally compensated to nominally guarantee stability, but the penalty for this was rapidly disappearing open loop gain with increasing frequency. There was just enough gain left to squeeze 20dB of broadband gain safely over a 20 kHz bandwidth. Some IC manufacturers came up with good μA741s, usably quiet and free of the grosser output offset voltage problems that plagued earlier devices. The μA741 was also output-protected to the extent of being short-circuit proof, a relief to all.

Subsequent generations of op-amps to the μA709 included the μA748 (the uncompensated sister to the μA741) and the 301, again, some versions being excellent for this class of device. That the μA748 and μA301 were user compensated did allow for more optimal parameter setting and in most circuits only required one capacitor to achieve this (as opposed to the necessary two resistor-capacitor networks for the μA709).

Although on the surface this appeared to be of great convenience to the designer, it disguised the fact that far superior bandwidth and phase-margin performance could be obtained by carefully considering the nature of the compensation network. Rather than just a simple capacitor of sufficient value to hold the amplifier stable (which also turned the internal compensated transistor into a

Miller integrator doing absolutely nothing for the speed of the device), a more complex network such as a two-pole resistance-capacitance network, Fig. 29-17E, improved matters greatly.

External feed forward, while in use as an inverting or virtual-earth mixing stage, also enabled a dramatic increase in bandwidth and speed over the more conventional compensation arrangements, as shown in Fig. 29-17.

29.7.1 Slew-Rate Limitations

All these early devices had one great failing that was leaped on vigorously by the hi-fi fraternity and audio engineers alike. Slew rate is the speed (measured usually in volts per microsecond, V/Fs) at which an amplifier output shifts when a step source of extremely high speed is applied to the input. All the early-generation op-amps had slew rates on the order of 0.5 V/Fs, but no one really understood it or its implications or effects then, and it was not the issue it is now.

Why is slew rate a problem? If the audio signal that the device is attempting to pass has a rise-time that exceeds the amplifier's slew rate, then obviously distortion is created—the amplifier simply cannot react quickly enough to follow the audio. Slew is an issue at high frequencies (rapid transitions) and high levels (a quiet signal is moving at fewer V/Fs than the same signal louder). This by happy accident means that a lot of program material that does not contain high amplitudes at high frequencies can be passed by slow, low-slew-rate devices with impunity. Unfortunately, a lot of sources found in recording studios or on stages don't fit that bill; they're loud and have serious high-frequency content. Making matters worse, some later musical genres such as EDM specialize in exactly this.

The speed limitation was nearly always in the differential and dc level-shifting stages of the devices. It is quite difficult to fabricate on an IC wafer ideal classes of transistors in configurations necessary to improve matters without compromising other device characteristics (such as input bias current, which affects both input impedance and offset performance).

Feed forward, in which a proportion of the unslewed input signal is fed around the relatively slow-responding lateral pnp stages, improving slew rate and bandwidth appreciably, was used to great effect in devices such as the LM318; a slew rate of some 70V/Fs achieved by this technique. It was in this area of slew rate, combined with a significantly improved noise performance (again another parameter suffering from difficulty in fabricating appropriate devices in a relatively dirty wafer), that the next major breakthrough occurred in devices commonly used for audio applications—the Harris 911. Although improved, the slew rate was still not fast and was also asymmetrical (+5 and −2V/μs).

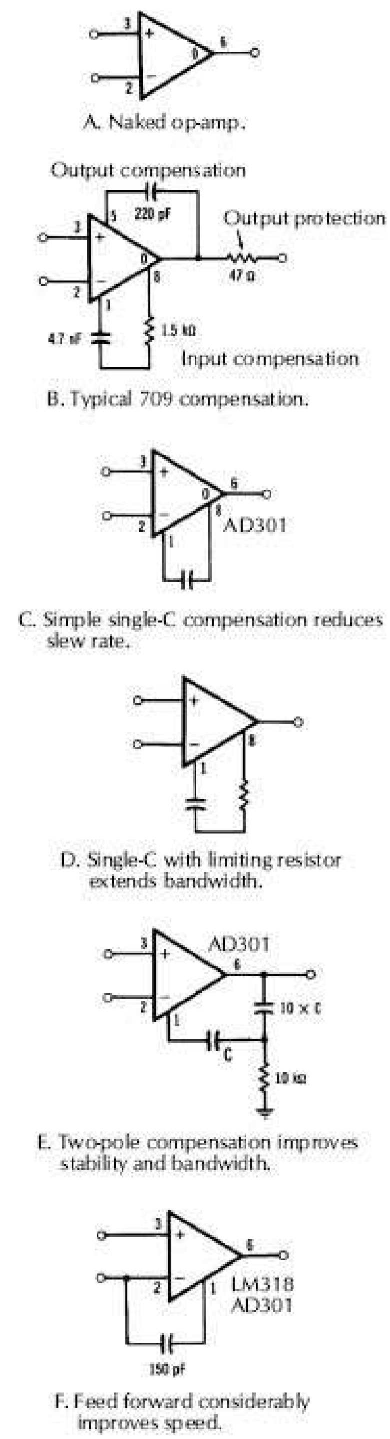


Figure 29-17. Various op-amp compensation techniques.

29.7.2 Bipolar Field Effect Transistors (BiFETs)

A breed of op-amps called BiFETs, or bipolar field effect transistors, emerged. These devices have a closely matched and trimmed field effect transistor input differential pair (hence, the typically unimaginably high 10MΩ input impedance) and a reasonably fast 13V/μs structure. These devices are typified by the TLO series from Texas Instruments, Inc. and devices such as the LF356 family from National Semiconductor Corp. Selected versions can, when source impedance is optimized, give noise figures better than 4dB at audio frequencies, which is thoroughly remarkable for units costing very little more than a μA741. There are better, quieter and faster devices available nowadays of course, but the BiFets were really the first generation of devices used in pro-audio whose quirks and limitations needed little dancing around nor woefully proscribe system performance, and are non-horrible choices even today.

The speed of the devices was achieved by the replacement of the conventional bipolar transistor differential input and level-shifting circuitry with FET configurations. Incidentally, the intrinsic noise characteristic of these FET front ends is significantly different from that of bipolars and seems perceptually less objectionable.

Eventually devices designed specifically and optimized totally for inclusion in high-quality audio equipment. With a quoted noise figure of better than 1 dB at audio, a slew rate of 13V/μs, and the ability to drive a 600 Ω termination at up to +20dBm, the Signetics Corp. NE5534 (or TDA1034) was in the vanguard of these; many nice devices have since followed.

These are all somewhat more expensive than the BiFET types, but none is prohibitively so, unless the target design is extremely cost sensitive. Today’s designer is spoiled by the ability to choose appropriate devices for each application almost regardless of cost, and as will be shown, sometimes the less grand and glorious parts are sometimes the better choice.

Noise in any competently designed and operated console can be attributed mostly to two sources:

1. Mixing amplifiers with an appreciable number of sources and, hence, a lot of makeup gain.
2. The input stage, especially a microphone amplifier with a fair amount of gain in it.

Once a background noise level is established from the front-end stage (at a level obviously dependent on the amount of gain employed there), the difference in noise contribution further down the line between an amplifier with a typical unity gain noise of −20dBu and one of −15dBu is for the vast majority of considerations totally insignificant. Concentration on these two hot spots will define the noise performance of an entire console.

In circumstances where extremely low system noise floors are actually necessary (rather than just deemed a good idea) and where such a noise level isn't being totally swamped by the source (which it usually is), then devices like the 5534 make sense elsewhere. Not so much that they are that much quieter within themselves but that their substantial output driving capability allows circuit impedances to be reduced, resulting in a worthwhile difference to noise floor. It's nice to know that there is also maybe a chance of chucking enough current at capacitors in filters for them to work properly at high frequencies and high levels. Using the 5534 as a microphone amplifier far outweighs the hassle of a similarly performing discrete transistor design.

Every design case demands a long cool look to determine what sort of device makes most sense; there is no one-fix cure-all technique, or device. For the most part, the designs here are based on TLO-class or 5534-class devices, determined mostly by whether low-noise, high-output drive capability or high-input impedance driving criterion. Modern devices (for a mature technology, new op-amps do seem to pop up almost weekly) that fulfill these specific or other niche parameters would of course be applicable.

29.7.3 Discrete Operational Amplifiers

The JE990, designed by Deane Jensen of Jensen Transformers and manufactured by Hardy Co. of Evanston, Illinois, is an example of an encapsulated discrete amplifier module. Many fascinating solutions to op-amp internal-design problems (some of which even IC designers evidently haven't realized existed) are implemented in this design whose features demanded a total reappraisal of contemporary audio circuit design and philosophy. Optimum input source impedance (normally about 10EΩ with most IC and discrete amplifiers) was reduced to about 1 kΩ by the use of an IC multiparallel input transistor differential pair. Small inductors in the emitters provide isolation from potential high-frequency instability due to the gain-band-width characteristic of the first differential stage shifting with varying source impedances. Unity-gain noise is a quoted staggeringly low –133.7dBu, while the output is capable of delivering full voltage swing into a 75 Ω load. This permits the use of exterior circuit elements of far lower impedance, reducing thermal noise generation. This elegant device inevitably carries an elegant price tag. Its many attributes pointed to the direction for amplifier design; it is a master-class embodied in that regard. Overall it outclasses any devices available in IC form and also, to the author's knowledge, of any universal discrete circuitry elements used to date in console manufacture.

That said, actual use of the ‘990 or such discrete modules is a philosophical rather than technological choice; there are ready means of achieving overall system performances that equal those possible with the ‘990 with considerably less cost and size. They shine if minimum elements in the signal path is a goal.

29.7.4 Instability

An unexpected thrill facing designers as they upgraded to newer, much faster devices was the tendency for all their previously designed circuits to erupt in masses of low-level instabilities even in what had been perfectly tame boards.

Layout anomalies, such as track proximity, were a major contributor toward the stability problems, so new layouts had to be generated with a whole new set of conditions added to the already hazardous game of analog card design. However, given adequate card layout environments, the real roots to this problem are with the devices themselves and a lack of appreciation of the relationship between their internal configurations and the outside world. Everyone who had been brought up designing around 741s had become too used to treating them in a somewhat cavalier fashion and for good reason. It was very hard work to make them misbehave or even show a hint of oscillation. People got used to treating ICs as plug-in blocks of gain with little consideration for the fact that inside was a real, live collection of electronic bits that still had all the problems real electronics always had. The reason the 741 was relatively impervious to user-inflicted problems is analogous to the fact that it’s quite difficult to get anything that is bound, gagged, and set in molasses to not behave itself.

Mistake number one with the new devices was believing that they were unity gain stable because the data sheets said so. What that really means is "does not burst into oscillation at unity gain (under these circumstances…)," which is not the same thing at all.

29.7.5 Phase Margin

It is important to maintain as large a margin as possible between the internally structured gain-bandwidth roll-off set for open loop and the roll-off around the external circuitry determining the closed loop gain. This is to preserve sufficient phase margin at all frequencies for which the circuit has gain. Failure to do this can result in the feedback being shifted in phase sufficiently to become reverse phase to that intended (positive feedback) with oscillation resulting. Even if the phase isn't shifted quite that far, the feedback tends toward positive and damped ringing when transients hit the circuit ensues. Also, these resonance effects can be extremely high in frequency, typically many megahertz, so any radio signal that gets as far as the circuitry will absolutely adore an amplifier that is critically resonant at its frequency! A reasonable phase margin to aim for at all gain frequencies is better than 45°. In practice, a compromise between desired circuit bandwidth traded off against the need to tighten that bandwidth for the sake of phase margins can be fairly easily reached with the newer devices, provided the need to do so is recognized.

There seem to be two schools of thought on bandwidth versus stability phase margin. First there are the Pragmatists, who close down the bandwidth of an amplifier as rapidly as possible outside the required passband, maximizing stability phase margin and RF neutrality. Then there are the Purists, who maintain circuit gain as far out and as high as possible, walking the tightrope of stability—usually in deference to the in-band phase linearity.

The normal, easiest, and most flexible way to determine the closed loop roll-off of a circuit is by means of a feedback phase-leading capacitor across the main output-to-inverting-input feedback resistor. A typical arrangement is shown in Fig. 29-18. Generally, the need to properly define the bandwidth of a gain block by just such a means automatically takes care of the matter, although it's dangerous design practice to assume that the two requirements—phase-margin determination and bandwidth limitation—are always mutually satisfiable.

A fairly common eroder of phase margin and progenitor of instability is stray capacitance from the inverting input of the amplifier to ground. This capacitance, a combination of internal device, pinout, and printed-circuit layout proximity capacitances, reacts against the feedback impedance to increase the closed loop gain at high frequencies. In normal circuits, even the typical 5 pF or so is enough to tilt up the closed loop gain parameters, threatening stability. Far worse is the situation where the inverting input is extended quite some distance along wiring, and worse yet, a bus—as in a virtual-earth mixing amplifier—hundreds, and sometimes thousands, of picofarads may be lurking out there. It can arise that despite a sizable time constant being present in the feedback leg, none of the expected high-frequency roll-off occurs since it is merely compensating for the gain hike created by bus capacitance. Ensuring required response and phase characteristics using any virtual-earth mixer can only be done properly with at least two orders of compensation around the mix-amp and with the finished system up and running completely, since any additional sources modify the impedance presented by the bus.

To define just how much this unwanted gain can rise, a small limiting resistor may be added as close to the amplifier inverting input terminal as possible; this is at the expense of the virtual-earth point now having a minimum impedance based on the value of that resistor. The resistor, incidentally, is also a measure of protection against any radio-frequency signals on the bus being rectified by the input stage junctions. Better yet, a small (real!) inductance in series with the summing amplifier input provides another means of out-of-band gain reduction and RF immunity. The size of this inductance is a matter of judgment in context. The danger is that this inductance in conjunction with the bus capacitance forms a resonance, which, if too far within the gain-bandwidth of the amplifier will nicely define the frequency at which it will oscillate; too little inductance will have next-to-no effect on stability.

29.7.6 Time-Domain Effects

There is invariably a finite time taken for a signal presented at the input of any amplifier to show an effect at the output of the amplifier—the so-called transit time. Every tiniest capacitance and consequent time constant in the internal circuitry of the amplifier make this inevitable; electronics takes time to do things. This transit time becomes an appreciably greater proportion of the wavelength of the wanted signal as the frequency increases, and as such it has to be taken into account. Fig. 29-19 shows how the fixed transit time becomes more relevant to increasing signal frequency. Ultimately, of course, the transit time will become half the time necessary for a wavelength of the signal frequency. At that stage what emerges from the amplifier will be a half wavelength or 180°

out of phase. Before this point, its detraction from phase margin with increasing frequency can start to cause serious problems; at this ultimate state, though, the negative feedback on which the amplifier depends for predictable performance is now completely upside down. Now it's positive feedback. Now the amplifier oscillates.

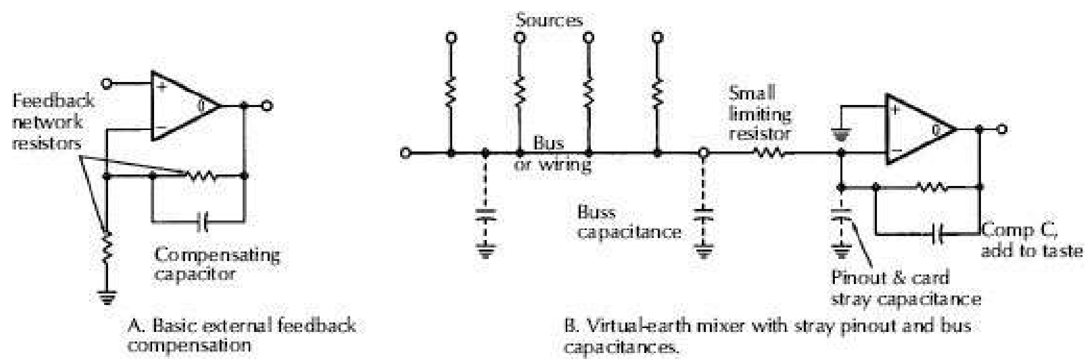


Figure 29-18. Feedback phase-leading stability compensation.

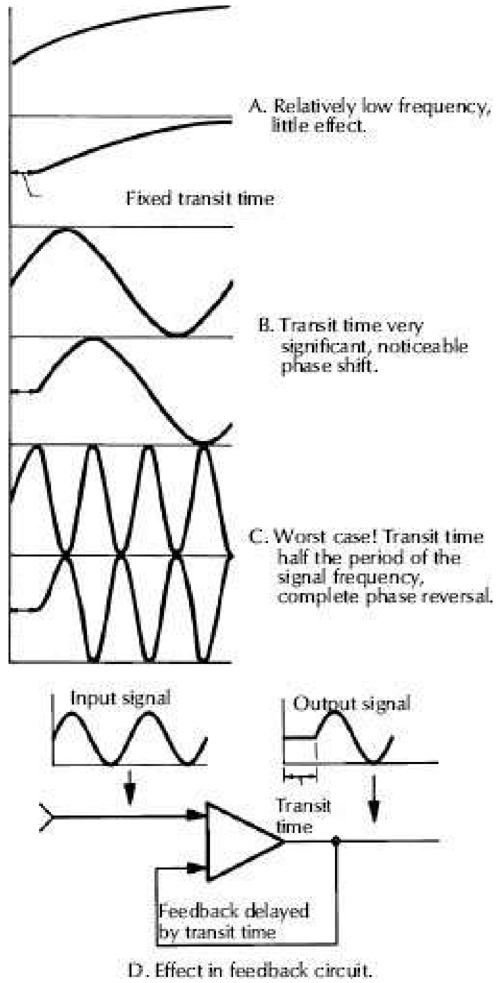


Figure 29-19. Transit time effects with increasing signal frequency.

29.7.7 Transient Intermodulation Distortion (TID)

The TID effect, if not fallout from and overwhelmed by the effects of insufficient slew rate, is due to amplifier transit times. Not surprisingly, as is nearly always the case with fad problems (as was TID during the 1970s), TID has been known and appreciated for as long as there have been negative feedback amplifier circuits—the twenties. It is and always has been totally predictable.

TID is a direct result of the servo nature of an amplifier with a large amount of negative feedback. The feedback is intended to provide a correction signal derived as a difference between the amplifier output and the applied input signal. It is a simple concept: any difference between what goes in and what comes out is error in the amplifier. All we need do is subtract the error. However, it is not so simple. Since there exists a time delay in the amplifier, the circuit has to wait for that amount of time before its correction signal arrives. The output during this time is uncontrolled and just flies off wildly in the general direction the input tells it to. Once the correction arrives, the amplifier has to wait again to find out how accurate that correction was and so forth, see-sawing on and on until the amplifier output settles. Fortunately, this all takes place rapidly (depending on the amplifier external circuitry), but it still represents a discrepancy between input and output. It is an effect peculiar to amplifiers with large amounts of negative feedback (typical of most contemporary circuitry), frequently displaying itself quite audibly—especially in power amplifiers where transit time is quite long with the usual huge, slow output devices. A larger and sonically more awful consequence is when such amplifiers clip; unlatching from clip (of course unaided by the now broken feedback loop) can be slow, ragged and ugly.

Amplifiers that rely on their own basic linearity—such as tube amplifiers—rather than on a servo-type nonlinearity correction system, are often held to be subjectively smoother. A whole subindustry thriving on the virtues of feedbackless circuitry has evolved. Nowadays, though, with device speeds improved as they are, settling times are becoming insignificant in relation to the signal transients with which they are expected to cope, pushing the frequency area at which TID could manifest itself far, far beyond expected audio excitation.

29.7.8 Output Impedance

A lot of devices, particularly the TLO series of BiFETs, have a quite significant open-loop output impedance. This is because the IC designers obviously considered that instead of an active output current-limiting circuit (standard on most op-amps up until then), a simple resistor would suffice. Although this built-in output impedance—by virtue of the enormous amount of negative feedback used—is normally reduced to virtually zero at the output terminal, it is still present and included as part of the feedback path, Fig. 29-20. Any reactive load at the output is going to materially affect the feedback phase and phase margin.

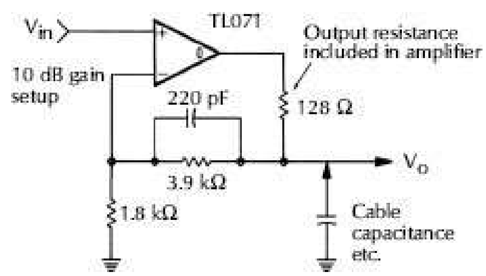


Figure 29-20. Output impedance as part of the feedback loop.

Any capacitance from the output to ground will form a feedback phase-lagging network. This shifts the phase inexorably toward the point where the total amplifier and network phase shift reaches 180° at the inverting input (that's a full 360° total), and the circuit oscillates. The frequency at which it oscillates is inversely relative to the capacitance value. It isn't unusual, with small values, to find oscillations right at the edge of the high-frequency sensitivity of an oscilloscope. Hanging a long piece of wire on the amplifier output (especially shielded cable with its high shield to inner capacitance) is a surefire guarantee of instability for this very reason. It has the added complication that there is a measure of inductance there, too.

It is conceivable that a long cable might start to look like a mismatched tuned stub at a frequency where the amplifier still has some gain, creating a creditably good, stable RF generator.

What this extra resistance-capacitance output circuit is in effect doing is to add dramatically to the transit time of the amplifier where actually the termination problem is creating far more delay than could possibly exist within the device itself. That the cures for the two ills are similar shouldn't be a surprise. Fortunately, a simple fix for this instability is to buffer away the load from the output feedback termination with a small resistor of typically 33–150 Ω. This usually does it, but at the expense of head room loss due to the attenuation from the buffer resistor against the load termination. Provided the load is greater than about 2kΩ, which it would really have to be in order to prevent getting close to current drive saturation in the IC output stage, this head room loss should be well less than 1dB. A better way is to buffer off with a small inductance, giving increasing isolation with frequency; a phase-shifting characteristic opposite to that of the (normally) capacitive load provides a total termination that is phase constant at the higher frequencies. At the lower audio frequencies, of course, the inductive reactance is very low, and the load sees the very low dynamic output impedance of the amplifier. The buffering inductance becomes virtually transparent.

Both of these techniques also provide a measure of protection against the possibility of RF signals finding their way into the amplifier by means of rectification in the output stage or inverting input. Very often output stages are more prone to RF field detection than inputs.

Some devices with a quite low output impedance before applied feedback (i.e. those with unbuffered, complementary emitter-follower output stages) are not likely to be fazed as much by these effects (pun totally intentional) but it is just as well to design in these considerations habitually. Emergency replacement, device upgrades or IC internal design changes can evoke this problem unintentionally.

29.7.9 Compensating Op-Amps

Op-amps generally have a couple of pins dedicated for compensation, which can be taken as a less than subtle message from the manufacturer that their product isn't stable under certain conditions of usage and needs external kludging. Usually this is at low closed loop gain where the bandwidth is at its most extreme. The classic solution is to shrink the bandwidth of the amplifier by slowing the amplifier down. Among other things, this wrecks the slew rate that's been handsomely paid for.

The most ordinary means of slowing down the devices is to slug an internal gain stage, leaving the other stages intact. On the bright side, if it is this internal gain stage around which the external compensation capacitor is hung that is tending toward instability, the capacitor should cure it. Sadly, it rarely is that stage. If a previous stage, say, the input differential amplifier, is unstable, all the capacitor will do is slow up the amplifier and reduce the slew rate to the extent that the oscillation is no longer visible at the output. It does not cure the instability. It's still in there, hiding. Often the only external manifestations are supposed dc offset voltages that won't go away and a poor-sounding amplifier.

There is a moral to this tale of compensation: don't use op-amps that require compensation if at all avoidable. Stability should be ensured by the circuit as a whole, and if speed is to be preserved, the op-amp should not be used below the gain at which it's happily stable. Compensation achieves stability by masking a symptom and not by tackling the cause.

The previous precautions, in addition to the feedback phase-leading capacitor, are now required circuit practice for using the newer, fast devices in many op-amp configurations. It should be said here that because there is no facility for implementing phase leading around the standard voltage-follower configuration and that this is the most critical configuration for stability, it is not a preferred circuit element. The manufacturer will have designed the IC to be just stable enough at unity gain to be able to say so unblushingly, but with probably little real-world margin to spare. Hanging a compensation capacitor across the appropriate pins will slow up the slew rate and not necessarily make the whole amplifier any less unstable. It is better not to tempt fate.

29.7.10 Input Saturation

The use of a standard voltage follower implies that in order to maintain the same system head room in that stage, the input has to rise and fall to the same potentials that the output is expected to. It can't. In most op-amps, especially those with bipolar inputs, the differential input stages saturate or bottom significantly before the power supply rails are reached and certainly before the output swing capability is attained. This is but one reason the temptingly simple voltage follower configuration isn't recommended. Limited input common-mode range means that the follower not only will cease to follow but will also spend a considerable amount of time in unlatching from one swing extreme or the other. Once an amplifier internal stage has latched, the feedback loop is broken; the stage has no assistance from the servomechanism to unstick itself. Once the loop is reestablished, it has to settle again as if from a hefty transient before it can resume following. Basically, this is an ugly scene. Uglier yet is the propensity of some devices when the input common-mode range has bottomed for the output to lunge to the opposite rail. Talk about "sonic character".

IC manufacturers commonly specify the common-mode input voltage range and it is precisely this limit that would be exceeded in use as a follower. For reference they are: ±13V for the 5534, ±11.5V for an LM318, and +15V to −12 V for a typical BiFET. All fall far short of the power supply maxima. There are devices available whose input common-mode range encompasses one or both supply rails, but they might not be appropriate for other reasons.

Provided enough gain is built around the amplifier to prevent these common-mode limits from being reached (in other words, the output saturates well before the inputs do) there should be no latching hangups; the feedback network also provides some substance to hang closed-loop compensation around in addition to enabling the full output voltage swing of the amplifier to be utilized.

Similar settling-time problems occur any time any stage is driven into clipping, but given the high power-supply voltages and consequent large head-room common today, clipping should be rare.

29.7.11 Front-End Instability

Altogether the most obscure potential instability-causing effect relates directly to the behavior of the input stage in bipolar front-end op-amps. The gain-bandwidth characteristic of the input differential stage is greatly dependent on the impedance presented to the input, the gain-bandwidth increasing with reducing source impedance. There is the possibility that given an already critical circumstance, the erosion in phase margin due to this effect can cause overall instability. This instability can be mitigated by limiting the gain-bandwidth excursion by means of a resistor (typically 1 kΩ) and/or some inductance in series with the input. Ordinarily, this would have little effect on circuit performance but may, especially in microphone amplifiers, detract from noise performance. Noise performance is largely dependent on the amplifier being fed from a specific source impedance, and 1 kΩ would be a sizable proportion. However, it's usually fairly easy to arrange in the design stage such that the IC doesn't have a zero impedance at either of its inputs.

Fortunately, because of the far greater isolation between the FET gates and their channels, this is a problem that FET-input op-amps do not have. A similar approach to that proposed for output isolation, i.e., an inductor rather than a resistor, in series with the affected input seems, on the surface, an equally good idea. The impedance of the inductors would be low at audio frequencies (so not affecting noise criteria significantly) and high at radio frequencies where the low source impedance phenomenon does its work. Unless the value is critically defined, an inductor of sufficient value to provide a usefully high reactance at RF also could be self-resonant with circuit stray and its own winding capacitances at a frequency probably still within the gain-bandwidth capability of the amplifier. Takes a bit of care.

Those who have experienced design with discrete circuitry will not be surprised that this source impedance instability effect is also the reason emitter followers are the most instability prone of the three basic transistor amplifier configurations. The cure is the same. Not only does the series impedance limit the source impedance before zero, it also acts together with any pinout and base-emitter capacitance as a low-pass filter helping to negate further external phase shift that may detract from stability. This base source-impedance instability is quite insidious in that it can either contribute to instability of the amplifier loop if it is already critical or it can be a totally independent instability local to the affected devices with nothing whatsoever to do with the characteristics of the external loop.

29.7.12 Band Limiting

One of the first great superficially appealing results of using the enormous feedback inherent to op-amps at the relatively low gain requirements of the audio world was a close approach to dc-to-light frequency response. The author remembers well the hysterical peals of laughter as the response of a new mixer was measured as still 0 dB right to the end of the testing ranges of the oscillator and the badly disguised puzzled looks and worried glances when we put real audio through and actually listened to it.

Many audio signals, especially live ones from microphones, analog tape-machine returns with a high vestigial bias content, keyboards, and a range of other sources, have a fair amount of ultrasonics present. If an analog remix of a digital recording takes place (it happens) many digital-to-analog (D/A) converters have an embarrassment of out-of-band noise that is of no program relevance whatsoever. A good microphone is going to hear all manner of stuff in a space: TV/computer scan whistles, motion alarms, switch-mode power supply or light-dimmer inductor screechings just for a start, none of which can be pretended to be musical. Depending on how good a following A/D convertor implementation may be, some of these may well get aliased down into the audible frequencies to less-than-subtle effect.

There is a proverb: the wider the window, the more muck flies in. Returning focus here just to analog signal processing, it would be perfectly all right if the following circuitry were capable of dealing with signals much higher than the audio band; sadly at the time (and to a lesser degree even now) that is not so. The root of the difficulty is the worsening open loop gain of the individual op-amps; as it drops off at 6dB/octave with increasing frequency, there remains less closed loop feedback available to maintain the op-amp's linearity. In other words, the circuitry becomes less and less linear as the frequency increases and the feedback dwindles.

Fig. 29-21 is representative of the open loop (no feedback) input-output transfer characteristic of an op-amp, i.e., what comes out in relation to what goes in. Not at all linear. In fact, rather nasty. (Incidentally, most big power amps have similar curves.) The good in-band linearity and low distortion of op-amps come from the application of monstrous amounts of negative feedback. Take the case of a noninverting 741-type amp with 40dB of gain around it, Fig. 29-22. At 100Hz there can be 60dB of feedback, which is great nonlinearities are being corrected by roughly the tune of 1000:1! However, the open loop gain plummets above this frequency, leaving a still respectable 40dB of feedback at 1kHz (100:1). (This figure of 40dB is widely regarded as the lowest amount of feedback for good performance from an op-amp.) At 10kHz it's down to 20dB; it is 14dB at 20 kHz; and at an ultrasonic 40 kHz, there is a bare 8dB! There is still gain, though, and the amplifier is quite capable of supporting and amplifying a signal up at those frequencies; it's just not very good at it.

Harmonic distortion of ultrasonics that would be generated by passing through a transfer function like Fig. 29-21 is unimportant; the frequencies would be even more ultrasonic. The problem lies in the intermodulation of two or more signals, products of which more often than not fall into the audible band; even reciprocal mixing with noise results in in-band noise products. A whole slew of intermodulation products are produced. It is no wonder that early op-amps sounded bad.

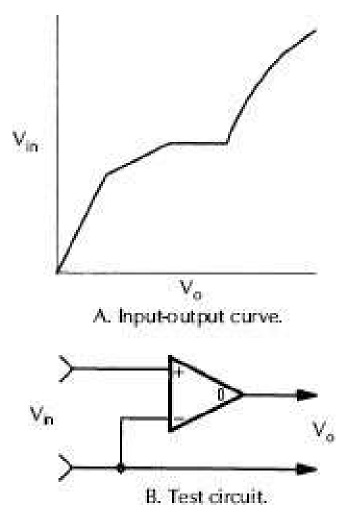


Figure 29-21. Operational amplifier open-loop gain curve typical of a bipolar device.

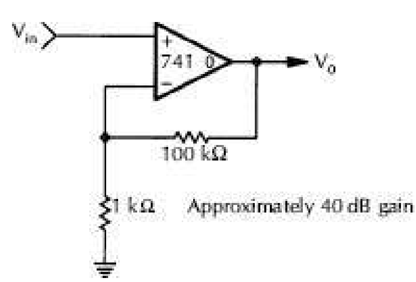


Figure 29-22. A 741 with 40dB gain.

So much for the expected result of improved transient response through having a wide-open frequency response. As is now obvious with hindsight, deliberately limiting the input frequency response of the mixer to a little more than the audio band can result in an amazing cleanup of the sound. By removing a lot of the inaudible signals that cross modulate within themselves and with in-band signals, the cause of much of the lack of transparency and mush that had become the trademark of early-generation IC op-amp consoles is eliminated.

Despite improved devices with greater open loop gains at far greater bandwidths, this approach remains valid today. By band limiting the program signal to reduce inaudible signals as early in the chain as possible, there is far less chance of their generating unwanted audible products. A front-end low-pass filter, operating in conjunction with all the other low-pass effects of feedback compensation arrangements throughout the console, should provide adequate minimization of these products in modern devices.

Purist arguments about the undesirability of any deliberate filtering seem rather futile in a world of real devices and real human physiology. There is a paucity of actual science indicating any human hearing above 20 kHz, so merely from that standpoint providing no in-band phase shifts are inflicted in the process, band-limiting to 20kHz removes all manner of decidedly awful and embarrassing stuff. All but a very few (specialist and expensive) transducers don’t hear/reproduce above 20kHz well, and final signal destinations—like most things digital, airchain, tape, are otherwise inherently band limited. However, with 96 kHz (and even 192 kHz) digital sampling threatening to become mainstream, widening the window to at least utilize some of the fabulously hard-won bandwidth may be in order in systems where such is likely. It has yet to be scientifically proven that claimed sonic benefit from increased sample-rates is from increased bandwidth rather than the many more prosaic reasons such as improved overall performance of the more modern designs used to implement them. Band limiting, to whatever sane degree, is a particularly powerful tool for obviating funny noises and lack of sonic transparency, and its use shouldn't be abdicated without a fight.

29.7.13 Slew-Rate Effects

Slew-rate limiting occurs when the fastest signal rise time the amplifier is expected to pass exceeds the speed of the fastest stage in the amplifier; the input transient becomes slurred to as fast (or slow) as the amplifier's capability. It is a level-dependent effect; at low levels the input signal's transient may be well within the amplifier slew envelope and escape unmutilated, but as the input gets larger the transient's slope can equal or exceed that of the amplifier.

Slewing gives rise to intermodulation effects that are dependent upon both frequency and signal level. The louder and faster the input transient, the worse the damage. A common subjective result of this limiting is for the high end of a drum kit to change in character of sound with differing levels of the lower-frequency instruments on which it is riding. Another favorite is the "disappearing snare drum" in which, again, the sound radically alters with changing level.

29.7.14 Device Idiosyncrasies and the Future

Many circuits rely somewhat on the extremely high input impedances of the BiFET devices and their very low required input bias currents. Using bipolars everywhere may result in unavoidably generated output offset voltages that could manifest themselves in extreme instances as switch clunks and scratchy pots. Also, the feedback phase-leading compensation may or may not be appropriate for devices other than BiFETs, especially with some bipolars with less than ideal internal poles. If there's a temptation to use more conventional bipolar devices, particularly those in multiple packages, it is also worthwhile examining their characteristics when inputs or outputs are taken above or below the power supply potentials. If the device structure under such circumstances is unprotected and turns into a silicon-controlled rectifier that deftly shorts the power supply with a bang, you are possibly better off using something else. In short, if a device is chosen specifically for an application and support circuitry designed for it, throwing in another device type in the expectation of improvement is usually nonproductive and often a step back. Op-amps and their surrounding components should be regarded holistically.

The proliferation of amplifier elements mushroomed in recent years with the availability of compact and extremely low-cost IC op-amps and manufacturing techniques, with increasingly complex functional blocks becoming increasingly commonplace. If, in order to improve their electrical and sonic characteristics, it would mean an increase in size and cost of well over an order of magnitude, would they still be quite as popular? In the good old days of tubes, it was not through any lack of expertise that equalizers even of today's complexity did not exist; it was just the size and cost would have made even the reckless shudder. Also, it is to be noted, they were not really thought necessary.

By way of history repeating itself, though, the astounding complexity of many digital audio algorithms, e.g., the use of as many as nine biquads to achieve not a whole EQ, just a single section, which alone would require a mere 27 op-amps to emulate, makes concerns about analog technology overkill seem a touch quaint.

29.8 Grounding

A human working visualization of anything electronic soon becomes impossible without a mental image of the solid, infinite, immovable, dependable ground. It has many other names too: earth, 0 V, reference, chassis, frame, deck, and so on, each of differing interpretation but all, ultimately, alluding to the great immovable reference.

Electrons could not care less about all this. They just go charging about as potentials dictate; any circuit will work perfectly well referred to nothing but itself. (Satellites, cars, and flashlights work, don’t they?) Ground in these instances is but an intellectual convenience.

Interconnection of a number of circuit elements to form a system necessarily means a reference to be used between them. To a large degree, it’s possible to obviate a reference even then by the use of differential or balanced interfacing, unless, of course, power supplies are shared.

So, having established that ground is seemingly only a mental crutch, why is it the most crucial aspect of system design and implementation?

29.8.1 Wire

Fig. 29-23A shows a typical, ordinary, long, thin length of metal known more commonly as wire and occasionally as printed-circuit track. However short it is, it will have resistance, which means that a voltage will develop across it as soon as any current goes down it. Similarly, it has inductance and a magnetic field will develop around it. If it is in proximity to anything, it will also have capacitance to it.

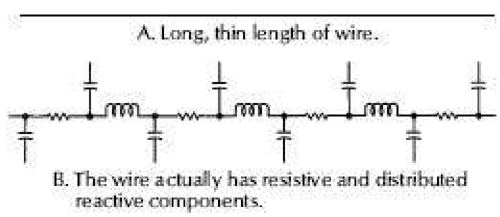


Figure 29-23. What is a length of wire?

So Fig. 29-23A actually looks more like Fig. 29-23B with resistive and distributed reactive components. Admittedly, these values are small and seem of little significance at audio frequencies, but clues have already been laid (particularly in section 29.7 on op-amps) that believing the world ends at 20kHz is not so much myopic as naive.

A radio engineer looking at Fig. 29-23B would mumble things like “transmission line,” “resonance,” or “bandpass filter,” maybe even “antenna.” RF technology and thinking may seem abstruse and irrelevant to audio design until it is considered that active devices commonly used nowadays have bandwidths often dozens, sometimes hundreds of megahertz wide. An even more frightening realization is the enormous quantity of RF energy present in the air as a consequence of our technological being; never mind the gigawatts of broadcasting bombarding us, the proliferation of walkie-talkies, cellular phones, and business radio all beg mischief in our systems. It even comes from other continents; the aggregate field strength of international broadcasters clutched loosely around 6 MHz, 7MHz, 10MHz, and 15 MHz is truly phenomenal.

A more obscure collection of equivalents is shown in Fig. 29-24. Fig. 29-24A represents a wire into a bipolar transistor input; Fig. 29-24B shows a wire from a conventional complementary output stage; and, for reference sake, Fig. 29-24C shows a basic crystal-set radio receiver. It may seem quaint, but for the presence today of wildly more volts per meter RF field energy compared to the heyday of wireless it works just the same. In all the three circumstances, radio frequencies collected and delivered by the antenna are rectified (hence, demodulated and rendered audible) by a diode (the base-emitter junctions in Figs. 29-24A and B). As contrary as it may seem for demodulation to occur at an amplifier output, it is perhaps the most common detection mechanism with the demodulated product finding its way back to the amplifier input by means of the conveniently provided bypassed negative feedback leg.

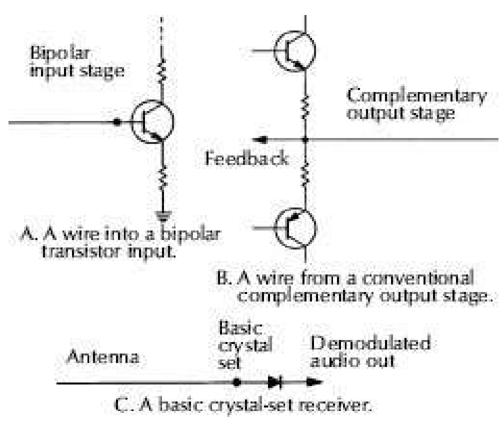


Figure 29-24. A collection of equivalents.

Making our length of wire fatter and thicker has the effect of lowering the resistance and inductance while increasing capacitance (greater surface area exposed to things nearby). So, although the resonant frequency of the wire stays about the same, the dynamic impedance (hence, Q) reduces. Although in general this is deemed a good thing, in some instances it can merely serve to improve the matching and coupling of the RF source to the resonance.

Carried to an extreme, even a console frame constitutes a big fat resonant tank at a surprisingly low (mid-VHF) frequency while frame resistance, however heavily it may be constructed, cannot be disregarded and cannot be treated as a universal ground path.

For the purposes of practical design, these considerations perhaps become a little better defined. The reactive elements of capacitance and inductance with the attendant effects of resonance and filtering are concerned with less obvious aspects (such as electronic stability and proneness to radio demodulation), while resistance gives rise to most of the horrors usually lumped under the collective term grounding problems.

29.8.2 Earth Ground

The closest most of us get to earth is the fat pin on an ac power plug. Fortunately for most purposes, it is adequate, provided just the one point is used as the reference. Other points are likely to have slightly differing potentials due to dissimilar routing and resistances. Compared to a technical earth ground, e.g., a copper water pipe or, alternatively, a fortune in copper pipe hammered into the earth, conventional earth grounds can have a surprisingly high potential, a volt or two, considering it is principally a safety facility not ordinarily carrying current. Any potential implies resistance in the earth path, which is bad news about something intended as a reference while also detracting from the safety aspect.

Practically, though, it does not matter too much if everything is waving up and down a bit provided everything, including even unrelated things in proximity, are waving up and down in the same manner. The potential is usually small, meaning that the ground impedance is reasonably low to the extent it may be considered insignificant.

29.8.3 Why Ground Anything to Earth?

With all our component system parts tied together by a reference ground and everything working as expected, the question arises as to why it is necessary to refer our ground to earth. If the internal grounding is completely correct, our system will operate perfectly, quietly, and tamely regardless of to what potential (with respect to earth) it is tied. If not tied, it will derive its own potential by virtue of resistive leakages, inductive coupling, and capacitance to things in its environment. For an independently powered system, i.e., batteries, these leakages and couplings will be of very high impedance and, hence, easily swamped by human body impedance to earth.

If, as is most often the case, most of the system is powered off the ac lines, this floating ground potential becomes of far lower impedance and consequently is much more capable of dragging current through a human load. That's you or me. (It's the current that kills, not the voltage.) A telltale sign is a burring, tingling feeling as you drag a finger across exposed metalwork on something that is deriving its own ground potential.

The mechanism for this lower impedance is fairly straightforward. Power transformers are wound with the optimum transfer of energy at 50–60Hz and very high flashover voltages (several kilovolts) in mind; the finer points of transformers such as leakage inductance, interwinding, and winding imbalance capacitance are all but disregarded.

Being far greater in scale than ordinary ambient reactive couplings, they primarily dictate the floating ground potential to be anything up to 230Vac above ground or whatever the power lines happen to be locally.

It used to be that some units were fitted with bypass capacitors from each supply leg to chassis ground, partly in the fond hope that this would help prevent any nasty noises on the ac mains from entering the hallowed sanctum of audio within. Ungrounded, this guaranteed the chassis floating at half the supply rail from a fairly low impedance. Ouch. But it gets worse. With the near universality of switch-mode power supplies, and with nearly everything containing digital electronics to some degree or other, there is the imperative of ensuring any nasties that the supply/digits generates don't find their way out of the box and up the power lead! Some reversal of fortunes there. This is required not so much from the altruistic desire to not pollute but that the box likely wouldn't pass emissions testing (FCC Part 15, CE and the like) and wouldn't be able to be sold. Typical supply-side filtering as a minimum in such supplies is a pi-filter of common-mode inductors and parallel capacitors—including capacitors to chassis ground.

The result is that if the chassis is not directly earthed, it rides at (in the case of both lines having tied capacitors) half the line voltage. The capacitor values grossly swamp transformer and ambient leakages and give the chassis floating potential an uncomfortably (literally) low impedance. The chassis tingle changes from “Mmm—interesting” to vile oaths with attendant flailing limbs.

A system composed of many separately powered units will almost certainly hum, buzz, and sound generally uneasy if not earthed, which is seemingly in direct contradiction to the earlier statement that “the system will operate perfectly regardless of what potential it is tied to.” Being tied to a lot of different self-generated potentials at a lot of different points along a system path is definitely not in the recipe.

Each different power transformer will have different amounts and permutations of leakage and, hence, propagate different potentials and degrees of power-line-borne noise into our otherwise perfect grounding path. Assorted ground potentials mean assorted ground currents, meaning assorted noises.

Tying the entire grounding path to earth is the best shot at “swampout” of leakage impedances. A connection to a (nearly) zero impedance makes nonsense of most other potential-creating paths, most of which have reactances in the kilohms.

Regardless of earth termination in such a multisupply circumstance, significant currents exist along the ground reference lines. The resultant interelement noise and hum voltages (developed across the inevitable line resistances) quickly become intolerable in unbalanced systems. Any wobbling of the ground reference becomes directly imposed upon the desired signal.

Balanced, or pure differential, transmission helps to obviate these perturbances by rendering them common mode in a system that is (theoretically) only sensitive to differential information. In reality, practical transformers can afford a good 70–80dB common-mode isolation at low audio frequencies. They deteriorate in this respect at 6 dB/octave with increasing frequency up to the winding resonance frequencies unless considerable effort is made to fake a more accurate balance externally. Although transformer balancing does effect a dramatic improvement in noise levels, it is far greater for fundamental hum (50–60Hz) than it is for other power-line-borne noise. This explains why in tricky systems, lighting dimmer buzz, motor spike noise, or any source with a high-frequency energy or transient content is so persistent.

The golden rule is to treat the grounding of any balanced system as if it were unbalanced. This minimizes the inevitable reference ground currents.

There is one overarching good reason only glanced by earlier for grounding to earth. The consequences of a piece of the gears' chassis becoming inadvertently at the power-line potential are obvious. We would much rather see death to a fuse or breaker than to one of us.

29.8.4 Console Internal Grounding

Let us assume that the grounding for the studio control room is all sensible and that our console has a solid earth termination. What about the intraconsole grounding paths? For most console builders this is perhaps the ultimate unbalanced signal path.

Conventional amplifier stages rely on a voltage difference between their input and reference in order to produce a corresponding output voltage (referred, naturally, to the reference of the input). If the input is held steady while the reference is wobbled, a corresponding (amplified) inverted wobble will appear at the output.

It is plain that any signal the reference sees that is not also common to the input, e.g., ground noise, will get amplified and summed into the output just as effectively as if it were applied to the proper input. The obvious (and startlingly often overlooked) regimen to render extraneous noise unimportant is to ensure that the point at which an amplifier source is referred is tied directly to the reference, while that amplifier output is only taken in conjunction with the reference. Successive stages daisy chain similarly—source reference to destination reference, and so on. This philosophy is called ground follows signal.

29.8.4.1 Ground Follows Signal

"Ground follows signal" is a classic maxim and one that has dictated the system design of nearly every console built. It was particularly true in the era of discrete semiconductor design, where ground was often not only audio ground but also the 0 V power-supply return; ideally the audio and supply grounds should be separate. As an added complication, the power-supply positive lines, being heavily regulated and coupled to ground, were an equal nightmare as they too became part of the grounding path. This could be fairly simply avoided by spacing each circuit element away from the supply line by an impedance considerably greater than that offered by the proper ground path—achieved by either separately regulating or simply decoupling by a series resistor, parallel capacitor network, as shown in Fig. 29-25. This actually gives the lie to the notion that single-rail supply systems are easier than differential rail arrangements; to do them properly results in almost indistinguishable numbers of parts and degrees of effort.

Accelerating technology has for once actually made life a bit simpler—specifically, the trend toward IC op-amps with their required differential (+Ve and −Ve) power supply. This, thankfully, removes electronic operating current from the audio system ground. While the excellent power supply noise rejection ratio of most popular op-amps helps tremendously, out-of-band supply noise rejection of these parts deteriorates roughly in step with open loop gain; it is good design practice to decouple immediately at the device power pins with say 47nF to ground, and ensure that the supplies are decoupled for in-band (audio) nearby.

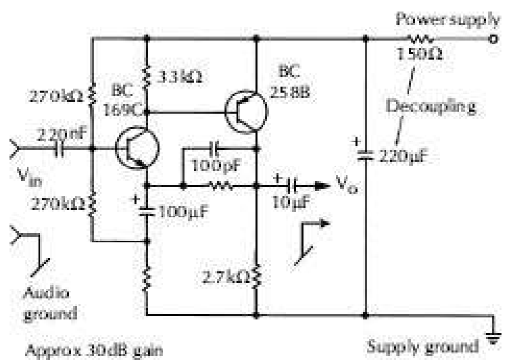


Figure 29-25. Power-supply decoupling: A typical discrete amplifier with the power supply isolated from audio ground.

Nevertheless, correct grounding paths still apply; the removal of supply current just exposes and highlights audio ground subtleties.

Unfortunately, although op-amps have simplified matters in one respect, their ease of use and versatility have been largely responsible for the creation of enormous systems with so many stages, break points, mix buses, and distribution networks that the simple daisy chaining of ground follows signal becomes unwieldy, if not unworkable. Alternate grounding schemes, such as star grounding where every ground path and reference is taken to a central ground or earth, play increasingly important roles.

In practice, a necessary compromise between these two prime systems occurs in most console thinking. Daisy chain applies mostly to on card electronics (e.g., a channel or group), while systems switching and routing rely on star connections.

29.8.4.2 Ground Current Summing

A principal grounding-related manifestation is crosstalk, or the appearance in a signal path of things that belong elsewhere. Other than airborne proximity-related reactive crosstalk, the most unwanted visitations are by the common-impedance or resistive ground path mechanism. In Fig. 29-26A, R_1 represents the load of an amplifier output (whether it's the 10kD of a fader or a 600 Ω line termination is immaterial for the present). The resistor R_G represents a small amount of ground path wiring, loss resistance, and so on. It is quite apparent that the bottom end of the termination is spaced a little way from reference ground by the wiring resistance, and the combination forms a classic potentiometer network. The fake ground has a signal voltage present of the amplifier output voltage attenuated by R_1 into R_G.

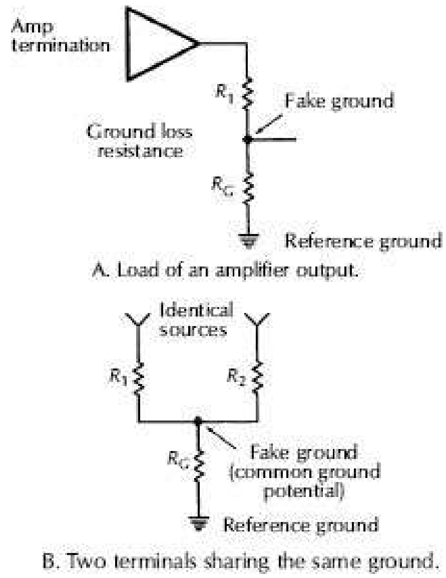


Figure 29-26. Ground current summing.

Practically, with a 600 Ω termination (R_1) and a ground loss (R_G) of 0.6 Ω, the fake ground will have a signal voltage some 60B down. The use of the fake ground as a reference for any other circuitry is a surefire guarantee of injecting −60dB worth of crosstalk into it.

Two identical terminations sharing the same fake ground, Fig. 29-26B, happily inject a small proportion into each other by generating a common potential across the ground loss R_G.

Should the second termination be far higher in impedance (the 10kΩ of a fader), its contribution to the common fake ground potential will be far less (−86dB) since the ground impedance is much smaller in relation to the source. Correspondingly, though, this higher impedance termination is more prone to be crosstalked into from the lower impedance contributors to the common ground.

29.8.4.3 Typical Grounding Problems

Here is a fairly unusual (but definitely not unknown) grounding anomaly resulting from inattention to the grounding paths. In Fig.

29-27 A2 is a line amp feeding a termination of 600 Ω into a lossy ground of 0.6 Ω resulting in a fake ground potential 60dB below the output of the amp. An earlier stage in the chain A1 (in this example, a microphone amplifier, with a considerable amount of gain) has its feedback leg (amplifier reference) tied to the same fake ground. Its input ground reference (here lies the problem) is taken from a separate bus supposedly to provide a nice, clean ground. This it does admirably, the bus being tied straight to reference ground and having no sources of great substance going to it.

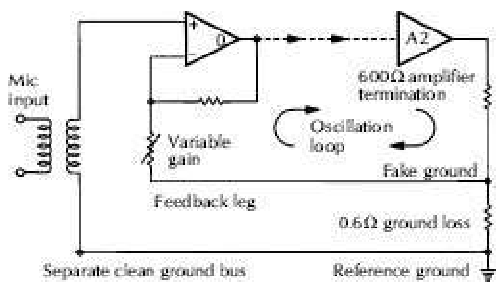


Figure 29-27. Feedback and oscillation via poor grounding.

Any signal present on the fake ground is duly amplified by the microphone amplifier (in its inverting mode) and is attenuated at the line amplifier output back into the fake ground. Naturally, as soon as the microphone amplifier gain exceeds the output attenuation, the entire chain bursts into oscillation.

A very similar mechanism was responsible for an owner's criticism of his well-known console that whenever he attempted to use the track routing on any channel modules, the sound of that channel discernibly altered. It was found that ordinarily nothing in the channel drew much current; all ground impedance requirements were quite light. Light, until the track routing line amp with its load of routing resistors and a terminated output transformer was accessed, demanding a relatively large ground current. This output stage current shared the only ground access point of the module (two paralleled connector pins) with all the rest of the module electronics, with the notable exception of the microphone and line input transformer ground returns. The resultant feedback, although nowhere near enough to promote oscillation, did by virtue of the phase shifting of the output transformer at both high and low frequencies result in distinct coloration.

A purist answer to these fake and loop problems is to choose one grounding point for the entire console and to take every reference and ground return directly to it through separate ground wires. Good luck with that.

A few less than minor problems would ensue. The enormous number of ground lines would soon outstrip the capacity of the module connectors, and the mass of wiring would cause apoplexy from the wiremen and aggravate an already critical world shortage of copper. Fortunately, a working compromise suggests itself based on separating the different classes of ground requirements by impedance. Bucket grounding refers to tying fairly high-impedance sources to a common ground point, bus, or line (since the ratio of their impedances is so great that resultant fake ground potentials can hopefully be made low enough to ignore). Anything that is likely to draw current (any kind of output or line amplifier stage) should go directly to ground, will not pass through any bus, and will not collect shared ground paths on the way to the bucket.

Any ground bus will have a measure of resistance and must, therefore, be fake to a certain degree. If we do our sums right, ground bus signal levels can be kept acceptably low, below –100dBu. Smugly, we can expect to ignore figures like that until we (almost inevitably) amplify them up.

A radical reduction in common-ground effects such as these results from reversion to old-fashioned balanced interconnection, but the underlying issues are neglected at peril.

29.8.4.4 Ground Noise in Virtual-Earth Mixers

A virtual-earth mix-amp unavoidably amplifies ground noise. Fig. 29-28A tells the story. For instance, a multitrack mix-amp can typically have 32 sources applied to it; the through gain from any source is unity (assuming the source resistors equal the feedback resistor), but the real electronic gain of the circuit is 33 or a touch over 30dB. Redrawing the circuit slightly in Fig. 29-28B shows exactly what this 30dB is amplifying. Consider as a clue that which is directly applied to the noninverting input of the op-amp—the ground! True, it is amplifying the noise due to the resistors and the internal noise mechanisms of the device, but for our argument here, it is amplifying ground. In any reasonably sized console, providing no sources are grossly out of proportion to the majority, ground noise is pretty random and noisy in character. The result is that, on being amplified up, it serves to make the mix-amp apparently much noisier than would be expected from calculation. In suspect systems it has been found to be the predominant noise source. It is no accident that the real electronic gain of a mix-amp is also known as its noise gain.

It is truly astonishing what attention to virtual-earth mixer grounding can have on bus noise figures. For mix-amps, practical noise performance has little to do with the device employed and nearly everything to do with grounding. Only when that’s sorted can the niceties of the electronics itself be addressed.

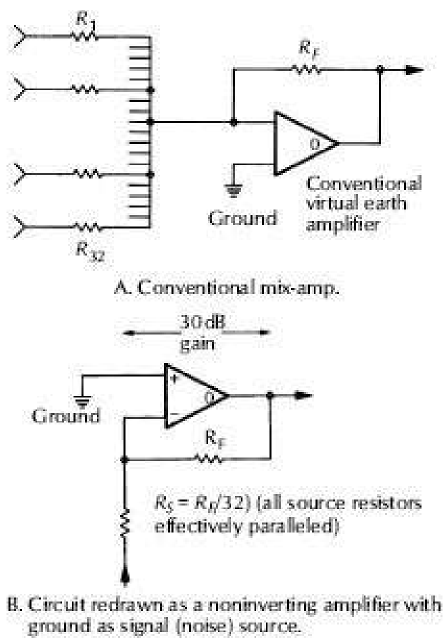


Figure 29-28. Virtual-earth mix amplifier as amplifier of ground-borne noise.

29.8.4.5 Reactive Ground Effects

Noise generation due to grounds is not limited to the resistance predominant in the ground wiring at audio frequencies. At radio frequencies (well within the band-widths of modern op-amps) even fairly short ground wires and buses can have very significant reactances, dramatically raising the effective ground impedance. This not so much reduces the isolation between the various stages as directly couples them together. All the inherent RF noise and instabilities of the stages become intermodulated (by the nonlinearity of the device at those frequencies) to make their presence felt as yet more audible and measurable noise.

A good “shock horror” extreme example, though described in simplistic theoretical terms, manifests itself sometimes dramatically in practice and can be called the standing on one leg effect.

The box in Fig. 29-29 represents a device that relies on a wire to be connected to the ground mass. It looks all right, and so it is, apart from the fact that at certain radio frequencies the wire is electrically three wavelength or an odd multiple of three wavelength. In accordance with transmission-line theory our innocuous bit of wire turns into a tuned line transforming the zero impedance of the ground to an infinite impedance at the other end. The result is that the device is totally decoupled from ground at those frequencies. Practical consequences of this, of course, vary, from instability at very high frequencies on cards with long supply and ground leads to painful, unreasonable susceptibility to RF in otherwise wholesome items of equipment.

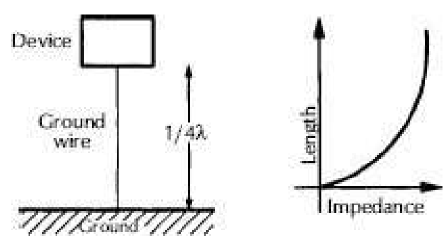


Figure 29-29. Standing-on-one-leg effect.

29.9 Signal Switching and Routing

Signal routing within the channel and other areas of the system is a touchy affair that has always been an area of much discontent for console designers, especially since the advent of in-line consoles and remotable and assignable systems. There are always standard relays, but these have lost, justifiably, a lot of appeal in the light of other technologies.

29.9.1 Relays

Unless they are of the expensive miniature IC package variety, relays tend to be big, heavy, eventually unreliable, mechanically noisy, a nuisance to implement electronically, and expensive. They also demand support circuitry such as back-emf protection diodes and drive transistors for a realistically operable system. The coils, being inductive in nature, draw a surprisingly large instantaneous on current and release an equally surprisingly large amount of back-emf energy when deactivated. Both of these—through mutual-inductance coupling, dubious common ground paths (even as far back as the master ground termination in separated supply systems), soft power supplies, and even mechanical microphonic effects—tend to impinge themselves on audio signal paths as clicks, splats, pings, and other assorted bumps. Of course, it's possible to have silent relay switching. However, after designing in separate ground unrelated power supplies of considerable heft, spatially separating the relays from the audio (preferably on another card), working out the drive interfaces, and liberally sprinkling the whole issue with diodes, resistors, and capacitors to tame the spiky transients, the circuit becomes very complicated.

Certain routing applications do implicitly require relays and their lack of concern about the amount of dc and either common-mode or differential signals of absurd quantities that may accompany the audio in balanced networks. Such circumstances are to be found anywhere a telephone line is used.

This is almost specifically a broadcaster's concern, where many external high-quality sources appear down phone lines and need to be routed before hitting either the internal distribution amplifier system of the station or perhaps even a console line input directly. Outside source selection, as it's called, does not fortunately have the same splat-elimination constraints as intraconsole switching, since the signal is nearly always of high level, balanced, and riding with at least a little dc (which will unavoidably click upon switching); most importantly the selector is very unlikely to be switched while actually live on air.

29.9.2 Electronic Switching

The wish list for an audio switch is simple:

1. It has an infinite off impedance.
2. It has a zero on impedance.
3. It has a control signal that is isolated from and does not impinge on the through signal path.
4. It costs nothing.

In the real world, of course, some leeway has to be given, but, fortunately, the tradeoffs are more in subtleties than in these basics.

Transistors are out of the picture right away despite their high on-off impedance ratios, because they are essentially unidirectional in current flow, and the control port (the base) is actually half of the signal path as well. In certain circumstances they have been used in the place of relays as a soft output muting clamp as in Fig. 29-30A.

Diodes, Fig. 29-30B, are used extensively for signal routing in RF equipment, with the required signal riding on a dc bias larger than the signal swing (so it can't over-swing itself off) that also overcomes the diode forward voltage drop making it a low-impedance path; it's turned off by a correspondingly large back bias. Considering that in some audio design cases getting rid of a handful of microvolts dc can be an ordeal, somehow a few dozen volts hurling about lacks a certain appeal. Typically, switching diodes for RF with a PIN structure are chosen; their very small capacitances are also considerably less parametric with respect to varying reverse bias, so minimizing automodulation and consequent distortions.

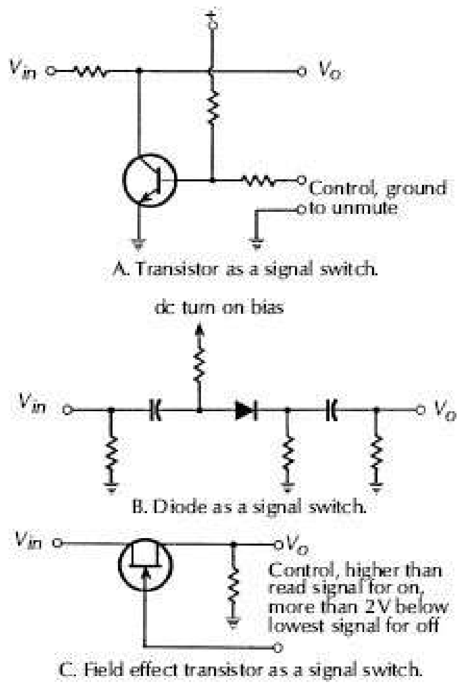


Figure 29-30. Solid state devices as simple switches.

FETs have been and still are used extensively for signal switching. They again have a high on-off ratio, and the control port (the gate) is of extremely high impedance and well isolated from the signal path, but the gate on-off voltage levels are a bit awkward for interfacing with logic control signals. They also define signal head room through the switch, based on the gate on-off biasing voltage range. It is bidirectional, its channel path being essentially just a voltage-controlled resistor, but the on resistance tends to vary with the varying audio voltage across it (auto modulation); distortion in the more basic FET switching configurations can be a problem. However, they are workable, Fig. 29-30C.

29.9.3 MOSFETs and CMOS

Closely related to FETs are metal-oxide-semiconductor field effect transistors (MOSFETs). They have a different chemical structure and physical construction but have essentially similar characteristics with the exceptions that the gate is of even higher impedance, and the control voltage swing required is easier to deal with. Complementary MOSFET (CMOS) elements, connected back to back to form close to ideal bidirectional analog transmission gates, are manufactured in all manner of variations and packages by IC manufacturers. At extremes of performance minor control-port breakthrough (charge injection) can rear its head and be taken into consideration.

Early versions of CMOS transmission gates had some rather untoward vices. They were raw CMOS elements, and one of their main attributes, the extremely high impedances in their off states and of their control ports, made them liable to destruction by normal amounts of static electricity. Also, they tended to latch up easily if any of the MOS junctions inadvertently got reverse biased into conduction (this happened easily if the signal voltage passing through a gate even momentarily exceeded the supply voltage). Most present devices are now gate protected to prevent static blatting, and the worst that happens with the audio signal exceeding the switch supply voltage by a small amount is that the switch breaks over (i.e., conducts audio momentarily). It does not result in the fatal consequences it once did.

Perhaps the best-known and most-used switch of this kind is the 4016 (and its younger brother the 4066, which is essentially identical but for a lower on resistance). It is a 14-pin dual package containing four independently controllable CMOS transmission gates. Each gate can pass up to the IC's supply voltage (typically 18Vdc) into a load exceeding 10kΩ with a distortion of about 0.4% in rudimentary switching formats. Obviously, both the distortion figure and the head room availability of 18dB above 0.775 V (for an 18Vdc supply) are both woefully inadequate by today's expected console standards. Another less obvious pitfall is the decreasing switch isolation at high frequencies due to leakage capacitance across the gate.

Fig. 29-31A gives a typical representation of the variation of the on resistance of a CMOS transmission gate with signal voltage applied to the gate. This variation in resistance is, of course, the source of the distortion. If we could restrict the signal voltage to within that (linear) bit in the middle, or better still virtually eliminate the signal voltage altogether, our problem would go away.

Placing the switching element right up against a virtual-earth ground point, as in Fig. 29-32A, achieves this signal voltage elimination; the switch now behaves as a two-state resistor. When closed, the on resistance variation, which will be small anyway because of the very low voltage swing across it, will be effectively swamped by the (relatively) much larger series resistance. When open, the off resistance extends the total series resistance to a value approaching infinity. In practice, the on-off ratio is not really adequate. Capacitance across printed circuit tracks and in the device encapsulation itself, combined with common-ground current and other essentially flat-response crosstalk mechanisms, results in a cross-switch leakage characteristic ultimately rising 6dB/octave against frequency. Also, despite the fact that the distortion problem is now largely resolved, there still remains a head room problem when the switch is open. If the source voltage presented to the series resistor exceeds that of the power supply of the CMOS gates, the gate will break over, turning on for that excessive portion of the input waveform.

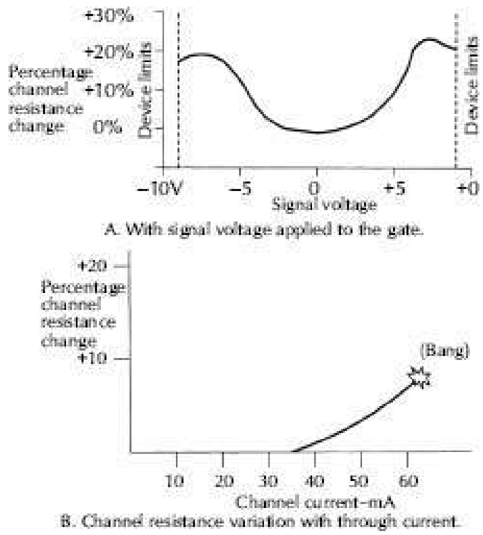


Figure 29-31. Typical CMOS transmission gate linearity.

Attenuating the source signal by the needed amount before it hits the gate skirts this hangup. Unfortunately, this worsens the noise gain of the virtual-ground amplifier by the amount of that attenuation. In Fig. 29-32B dropping an equal-value resistor to the series resistor to ground from its junction with the gate is a working approach. The maximum signal that can be present across the gate when off is now half that previously, which is usually more than enough attenuation to prevent breakover. This 6dB loss is magically made up for in the on mode because the source resistance of the signal into the amplifier is now halved (series resistance effectively in parallel with the dropped resistor).

Incidentally, the crosstalk improves as a consequence by almost 6dB—less signal voltage actually within the chip. For many practical purposes, this switching configuration, with its performance limitations as defined, is quite adequate. For instance, the noise and crosstalk characteristics are a good order of magnitude superior to any analog multitrack recorder, so this element can be a good choice for an inexpensive track assignment routing matrix.

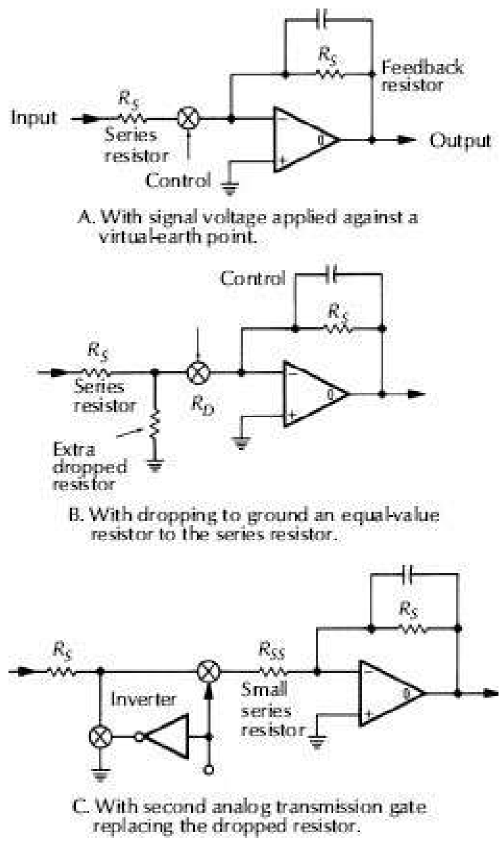


Figure 29-32. Switching arrangements using CMOS transmission gates.

29.9.4 Potentiometric Switching

A refinement of this element—in fact, really an extension of the same principle—is shown in Fig. 29-32C. Here, a second analog transmission gate replaces the dropped resistor and is driven through an inverter from the control line for the original gate, arranging for it to be on when the other is off and vice versa. When the original gate is on, there is very little potential across either of the gates (they're both at virtual ground from the op-amp). Similarly, there is little potential across either of the gates when the second gate is on, since it is tying the series resistor to ground and the open gate is between ground and virtual ground. Crosstalk is dramatically improved when the element is off because any signal present at the series resistor faces the double attenuation of the series resistor tied to ground by the on second gate followed by the off original gate into the virtual-earth input of the op-amp. In the on mode of the element, there is no input attenuation; hence, there is no gain and no extra noise contribution from the amplifier. The only limitation now to the cross-switch leakage characteristic of this switching element is printed-circuit card layout and grounding arrangements. Given a good home, this element is virtually unmeasurable.

It does, however, have one quirk that may preclude its use in some places. Unless a great deal of care is used to arrange complementary on-off switching timing for the two gates, they are both momentarily partially on together during a switching transition. This, for an instant, ties the virtual-earth amp input to ground via the quite low half-on impedances of the two series gates, creating an instantaneous burst of extremely high gain from the amp; this shows as a transient of noise or worse still as a splat if any dc offset is present at the virtual-earth point. It can be minimized, or at least the extent of the transient defined, by a small value resistor (R_{SS}) in series with the input, Fig. 29-32C. This will, of course, increase the signal voltage across the gates and increase the distortion, so a compromise has to be struck to suit the given application. Even so, excessive distortion owing to this has never shown itself to be a problem.

The tiny and essentially meaningless residual nonlinearity remaining in the potentiometric implementation can be improved by cancellation, by connecting an enabled CMOS switching element in the feedback loop of the amplifier together with an equal-value build out resistor.

29.9.5 Minimizing Noise

To reduce the thermal noise contribution as part of the circuit noise performance, the resistances involved in switching should be as low as practically possible consistent with device limitations and the ground current arrangements. The feedback resistor around the virtual-earth stage is limited by the output drive capability of the op-amp, bearing in mind it has to drive its load, too. Fig. 29-31B demonstrates a typical channel resistance variation of a CMOS switching element with through current. It behaves linearly until about 40mA, which actually compares more than favorably with the output drive current capability of most op-amps. (FETs are excellent constant-current sources, self-limiting in nature.) As a rule of thumb then, the resistors used around analog gate switching circuits can be as low as 2.2kΩ without exceeding device limitations; the high-output current capability of the 5534 can be used to good effect here if the drive for lowest possible thermal noise is that important. Generally, ground-borne noise generously provides a noise floor well before this theoretical limit is attained; the whole lowest-impedance question becoming self-defeating eventually. The more current is chucked around, the worse the ground noise becomes.

The 4000 series of CMOS devices, which are very commonly used, have one important feature at odds with general mixer technology—their maximum supply voltages. The earlier 4000A series were limited to a 15Vdc total (as compared to the 30Vdc or 36Vdc total commonly used in console design), while the more recent buffered B series can stand 18Vdc. More recent families (HC, e.g.) nearly always adhere to the common digital electronics supply of 5 V (actually 7 V max). An advantage of the earlier series was that a separately regulated 5 V supply wasn't necessary. Nowadays, though, there is nearly always 5 V or 3.3 V running around for this, that, or the other and it is not a difficulty to create a sufficiently well-regulated and quiet supply for HC switches. Given the virtual-earth switching technique described, this diminution of supply is immaterial.

29.10 Input Amplifier Design

A console is expected to accept signals of wildly varying input level and impedance while producing a uniformly consistent output capable of being deposited in the tightly defined container that is a recording track, or similarly defined output.

Fortunately, industry standards provide at least some clues as to what mixers are likely to have applied to them. Nevertheless, these standards can obviously do nothing to alter the physics of the operation of the assorted transducers and sources in common use; the disparity in the treatment required between a dynamic microphone and a recorder output totally precludes a universal input stage.

Mixer front-end design tends to be a little like working on a grown-up jigsaw puzzle where all the important pieces perversely refuse to fit. It's delightful to discover or cultivate some that fit nicely, like in line-level input stages. This euphoria is chipped away by the problems inherent in other areas, notably microphone input stages.

Optimizing input noise performance in a dynamic microphone preamp is quite an operation, juggling a seemingly endless number of variables. A dynamic microphone may be represented (a little simplistically) as a voltage source in series with a fairly lossy inductance representing a midband impedance anywhere between 150–1000 Ω depending on the model, Fig. 29-33. Being a transducer and, of necessity, mechanical in nature, many complex varying motional impedance effects contribute to the overall scene, as do the effects of matching transformers used in many microphones. For most design purposes, however, this simplistic electrical analog can suffice. The low impedance commonly and conventionally used is primarily to mitigate high-frequency attenuation effects due to inevitable cable capacitance. Despite the fact that the characteristic impedance of microphone cable is not too far removed from that of our typical sources, the runs are so short in wavelength terms that transmission-line "think" is not really applicable and the cable just looks like a distributed capacitance. This, in practical circumstances, amounts to a large value of capacitance that the transducer must drive along with its load. Unfortunately, the impedance is not low enough that it may be treated as a pure voltage source; therefore a tiny signal at a finite impedance must be ferreted out and treated with care for optimum performance.

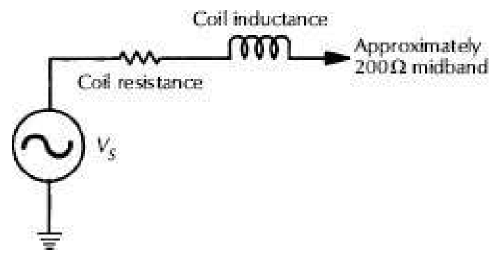


Figure 29-33. Simplistic dynamic microphone model.

29.10.1 Power Transfer and Termination Impedance

Textbooks on electrical theory state that to extract maximum power from a given source the optimum load must be equal in value to the source impedance. In the dynamic microphone, it is of doubtful (if any) value. We've squeezed all the energy possible from the generator, but to what end? Given that most electronic amplifiers of the type useful in low-noise applications are of relatively high-input impedance, i.e., voltage amplifiers, then the terminating resistance that largely defines the load of the microphone would, in fact, dissipate most of our hard-won power. It is the output voltage capability of the source that is of greatest use here, not the power. So, as can be seen in Fig. 29-34, matching source and load impedances does a very effective job of sacrificing 6dB of signal level that then has to be made up in the succeeding amplifier. This does not imply that the noise performance is 6dB worse than possible, since the source impedance as seen by the (assumedly perfect) amplifier is now a parallel of the microphone and its matching load; hence, it is about half the value of either. The thermal noise generation of this combined source is consequently 3 dB less; so although the voltage is down 6dB the noise performance is only degraded 3dB by such a termination. Still, it is better not to throw away a good 3dB before even starting to hassle with the amplifier itself.

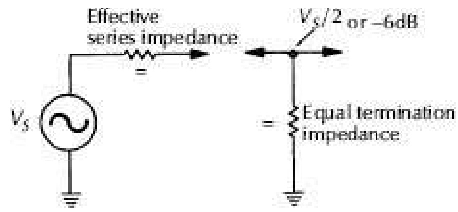


Figure 29-34. Matching—how to lose 6dB.

Another good reason for not terminating with an equal or any fairly low resistance is the effect on microphone response and subjective quality. Having an inductive characteristic, the dynamic microphone capsule has an impedance that steadily rises with frequency, becoming predominant at high audio frequencies where the inductive reactance of the source is large with respect to the coil-winding resistance. When terminated with a relatively low resistance, the complex impedance of the capsule and the termination resistor form a single-order 6dB/octave low-pass filter, gracefully rolling off the high-frequency output of the microphone. Not a useful consequence. Adding insult, if the termination impedance is not flat with frequency—some input configurations roll down at the top and/or bottom—the overall response would tend to follow that impedance, too.

With a fairly hefty cable capacitance, the system is no longer graceful; the complete network now looks like a rather rough second-order filter. There isn't too much to be done about that; regardless of termination method, cable capacitance is here to stay and is always a consideration unless the preamp is remoted to be at or close to the microphone itself.

29.10.2 Optimizing Noise Performance

Amplifiers are not perfect. For noise criteria, the first device that the signal hits in the amplifier is the key one, since the noise it generates usually masks—by a large margin—noise from all succeeding stages.

All practical amplifying devices are subject to a variety of internal noise-generating mechanisms, including thermal noise generation. When measured, these give rise to some important values; namely, the input noise voltage, the input noise current, and the ratio between those two that is in effect the input noise impedance. This becomes all important in a little while. Although there are some very low voltage noise JFETs sometimes used, for the most part, bipolar transistors—either standard or more usually large-geometry and sometimes multiparalleled—are used as front-end devices both in discrete designs and op-amp IC packages in this application so much of the following relates to these.

These noise voltages and currents alter in both individual magnitude and ratio to each other with differing electrical parameters, especially collector current. Predictably, as this current decreases, so does the noise current (most of the noise is due to minor random discontinuities in device currents); the ratio between the noise voltage and current—or noise impedance—may be altered in this fashion.

Thermal noise generation is common to all resistive elements. The amount is related to both the temperature and the bandwidth across which it is measured; an increase in either will increase proportionally the noise power generated. Under identical circumstances, the noise power that is generated by any values of resistance is the same. Differing resistor values merely serve to create differing ratios of noise voltage and noise current; the product of the two always equals the same noise power. This particular noise phenomenon, thermal noise or Johnson noise, is totally unavoidable because the nature of atomic structure is such that when things get hot and bothered, they grind and shuffle about randomly, creating electrical disturbances white in spectra, i.e., equal energy per cycle bandwidth.

Even the real (resistive) part of the complex impedance of a dynamic microphone generates thermal noise; this ensures that there is a rigidly defined minimum noise value that cannot be improved upon.

29.10.3 Noise Figure

The difference between the noise floor defined by thermal noise and the measured noise value of a practical system is known as the noise figure (NF) and is measured in decibels (Noise Figure = System Noise – Theoretical Noise). The noise output from a resistor or the real part of an impedance is calculable and predictable—Herr Boltzmann rules. A direct comparison of the noise voltage measured at the output of an amplifier due to a resistor applied to the amplifier input versus the noise voltage expected of the resistor on its own is possible just by simply subtracting the measured gain of the amplifier. This is a measure of NF.

An interesting effect occurs when, with any given set of electrical parameters set up for the amplifier front-end device, the source resistance is steadily changed in value. A distinct dip in the NF occurs, Fig. 29-35, and the value of the resistor at which this dip occurs changes as the device parameters are changed (collector current primarily). For the usually predominant noise mechanism (thermal noise), a minimum NF occurs with a tiny amount of collector current (5–50 μA) and a high source resistance (50kΩ up). Without diving into the mathematics, the nulling is a balancing of interaction between the external noise source and the internal voltage and current noise generators. Larger geometry devices exhibit similar curves but at lower impedances and higher collector currents, but bottoming out at similar noise figures.

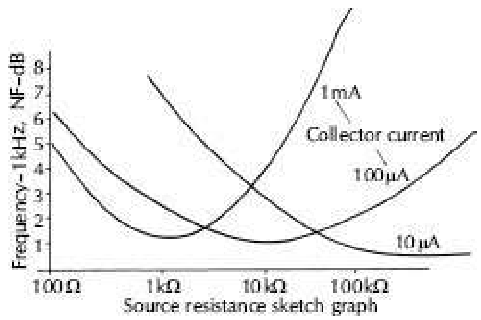


Figure 29-35. Bipolar noise curves (noise figure curves for a good pnp front-end transistor for collector current versus source resistance).

29.10.4 Inverse-Frequency Noise

There is another major noise mechanism inherent to semiconductors. It is the low-frequency (inverse level with respect to frequency) noise—a burbly, bumping-type noise caused by the semiconductor surface generating and recombining sporadic currents—most prevalent in dirty devices but present to a degree in all. It is subjectively apparent and has to be considered. Measured alone, low-frequency noise has its own set of collector current and source resistance nulls, usually far higher in current and lower in resistance than for thermal noise.

Commonly known as 1/F noise—implying its predominance at very low frequencies—it is often specified by way of the frequency at which it is contributing the same amount of noise as the device's effective thermal noise. Below this knee frequency 1/F noise predominates. A good clean device will have a knee frequency below

10Hz; judicious filtering along the signal path can render 1/F noise unimportant within a system, but it remains a serious consideration within each individual amplifier stage.

29.10.5 Optimum Source Impedance (OSI)

A compromise has to be struck. To make a generalization from Fig. 29-35, a 100 μA collector current and a 10kΩ source impedance for a typical low-noise small-signal pnp transistor seem about right. (pnp transistors are commonly used in this area due to slightly better low-frequency performance figures over npn types.) The source resistance value is that at which the device is optimally quiet for audio purposes and is known as the optimum source impedance (OSI). Incidentally, this impedance has absolutely nothing to do with the kind of circuit configuration in which the device may be. Whether it be in a common-base amplifier with an input impedance of 50 Ω or in a totem-pole front end with bootstrapping and a consequential input impedance of over 10MΩ, it doesn't matter. The source impedance for optimum noise performance stays at 10kΩ, or whatever, provided that the collector current is the same in all cases. Optimum source impedance has nothing to do with input impedance.

This optimum source impedance varies depending on the type of input device used. For an FET, the noise figure typically obtainable drops to an amazingly low value but, unfortunately, at a substantially useless impedance of several dozen megohms.

Large geometry transistors with very low base-spreading resistances (small power devices etc.) can have much lower OSIs if not better noise figures, however good small-signal bipolar transistors have OSIs in the region of 5kΩ to 15kΩ whether discrete or as part of an IC amplifier package. Fortunately, these values closely coincide with the source resistance value that provides for optimum flatness of device transfer characteristics. This helps a long way toward best frequency versus phase linearity, which translates to enhanced stability in a typical high negative-feedback amplifier configuration.

Fig. 29-36 shows the effect of altering the source impedance into such an amplifier (using a conventional bipolar transistor input device) on output frequency response. The droop is due to the excessively high source impedance reacting against the device base-emitter, board, and wiring capacitances to form a low-pass filter. The high-frequency kink is a practical effect of the curious mechanism; when a bipolar transistor is fed from an impedance approaching zero, its high-frequency gain-bandwidth characteristic extends dramatically, radically altering the phase margin and, consequently, the stability of an amp designed and compensated for more ordinary operating circumstances. The kink is a resonance within the amplifier loop caused by erosion of phase margin resulting from this mechanism. It is but an uncomfortably short step from oscillation.

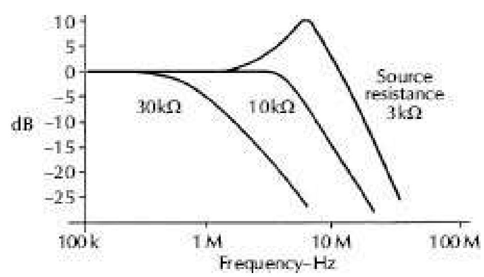


Figure 29-36. Source impedance versus bandwidth gain versus frequency for a typical follower-connected operational amplifier highlighting effects on response of source impedance on input device.

As can be seen from the graph in Fig. 29-36, the response is maximally flat at a source resistance of around 10kΔ about the same value as the OSI for optimum noise performance for the same configuration. A problem to reconcile is that our practical source impedance is nominally 200 Ω for a dynamic microphone, whereas the OSI for the best conventional input devices is around 10kΩ. How do we make the two fit?

29.10.6 Microphone Transformers

There are some horrible stories about how bad transformers are. On the other hand, they have become trendy again. Properly designed and applied, however, they do offer a good to excellent solution to impedance matching and sundry other problems facing input stage design. Simplistically, a transformer is a magnetically soft core around which are two windings, the voltage ratio between the two being equal to the ratio of the number of turns on each. The impedance ratio is the turns ratio squared, e.g., a 10:1 turns ratio corresponds to a 100:1 impedance ratio, because power output cannot exceed power input. If the voltage is stepped up ten times, the output current must be stepped down ten times. Impedance, which is the ratio of voltage to current, is consequently the square of the transformed voltage or current ratio, see Chapter 15 *Audio Transformers*.

Given this, it is a simple matter to calculate the ratio necessary to match the microphone impedance to the amplifier OSI that is realistically achievable. The convention that 200 Ω is a necessary target for source impedance serves well, despite it seeming that the microphone manufacturers didn't get the memo with midband impedances measuring all over the map from 150 Ω to πΩ.

Variations within the normal range of actual microphones make trivial differences in the larger scheme of things though. The assumption that most bipolar input amplifiers have an OSI of between 5kΩ and 15kΩ indicates that the transformer ratio should lie somewhere between 1:5 and 1:8.7.

Many consoles use higher ratios (typically 1:10), maybe in the naive belief that the noise advantage of a step-up input transformer stems from the free gain it affords. Although on a basic level it would seem to make sense that the less electronic gain needed the quieter the system must be, this fallacy is completely belied by the truth that the transformer merely allows you to choose and alter the source impedance for which the amplifier is optimally quiet. Increasing the turns ratio beyond this easily defined optimum can and will actually render the amplifier noisier.

In practice the free gain can be more of a nuisance than a benefit. It is not unusual for microphone inputs to receive transients exceeding +10dBu and mean levels of –10dBu, especially on a rock-and-roll stage or from a drumkit. Even dynamic capsules can deliver frightening levels that can pose headroom problems in the mixer front end. A typical 1:5 transformer has a voltage gain of 14dB (20dB for a 1:10 ratio), which would mean that even with no electronic gain after the transformer, normal mixer operating levels are being approached and possibly exceeded. These circumstances make worrying about a dB or two of noise performance total nonsense to be sure; it just serves to point out that our microphone front end has to be capable, if not perfectly optimized, for elephant herds as well as butterflies.

29.10.6.1 Transformer Characteristics

Transformers have numerous limitations and inadequacies resulting from their physical construction that make their actual performance differ (in some respects radically) from that expected of a theoretical model.

The heart of the transformer is the magnetically pliable material into and out of which energy is induced. Virtually any material—nickel, steel, iron, ferrous derivatives, and substitutes—have the same basic limitations. They saturate at a magnetic level beyond which they are incapable of supporting further excursion, and exhibit hysteresis—a crossover like nonlinearity at low levels responsible for a significantly higher distortion at low levels than anything else likely to be found within a well-designed modern-day signal path.

These two effects at opposite ends of the dynamic spectrum mean that all transformers have a well-defined range within which they must be operated and this range is less than the range of levels the microphone amplifier (mic-amp) is expected to pass. This is especially true at low frequencies, where the core is earliest prone to saturation. Optimization begins here. Is it to be designed for minimum hysteresis (butterflies) or with plenty of material to be tolerant of monstrous (elephantine) signal levels?

Windings are made of wire, which has resistance. Resistance means loss and decreased efficiency and noise performance. By the time there are enough turns on each of the windings to ensure the inductive reactances are high enough not to affect in-band use, winding resistances can no longer be ignored. The resistive losses alone mean that even very good transformers rarely afford a total microphone input stage noise figure much better than 1 dB.

Capacitance exists between things in close proximity and that includes transformer windings—between each other, between adjacent turns and piles in the same winding, and from the windings to ground. In this given instance it is nothing but bad news. Capacitance between windings means unwanted leakage and imperfect isolation, while winding self-capacitance reacts with the winding inductances to form resonances. Resonances, even if far outside the audio band, invite response trouble, and disturb in-band phase linearity. Combinations of these capacitances greatly affect one of the greater advantages of transformers, common-mode rejection (CMR).

29.10.6.2 Common-Mode Rejection

For a transformer to work and transfer wanted information from one winding to another, a current must be made to flow through the primary; this is ideally achieved just by the opposing polarity (differential) signal voltage applied across it. Again ideally, any identical signals on the two ends of the windings (common mode) should not cause any current to flow in the winding (because there is no potential difference across the winding to drive it) and so no signal transfer can be made into the secondary. So much for ideal.

Common-mode rejection is the ability of the transformer to ignore identical signals (in amplitude and phase) on the two input legs and not transfer them across the secondary as differential output information.

Principally, it is imbalanced distribution of capacitance along the length of the two windings, both with respect to each other and to ground that makes CMR less than perfect. Co-winding capacitance has the effect of directly coupling the two wiring masses permitting common-to-differential signal passage, which worsens with increasing frequency at 6dB/octave. Electrostatic shielding (a Faraday shield) between the windings can alleviate co-winding capacitance coupling.

Further CMR worsening can be expected even if the two windings are perfectly balanced with respect to each other, if the primary winding is not end-to-end capacitatively matched with respect to ground. Any common-mode signal from a finite impedance source (almost always the case) when confronted with such a capacitatively unbalanced winding sees it as being just that—unbalanced (becoming more so with increasing frequency). Again, input common-mode signals are transferred across to become output differential information indistinguishable from the wanted input differential source.

Broadcasters particularly are concerned with winding balance, not only on microphone transformers but also on line-output transformers, reasoning that common-differential transference is as likely to occur at a source as at an input.

29.10.6.3 Two Microphone Transformer Model

Fig. 29-37 gives a better idea of what the small signal of a dynamic microphone has to suffer. The winding capacitances (CP and CS) form lovely resonances with the inductances, while the transformed up primary winding resistance (RP) added to the resistance of the secondary winding (RS) merely serves to increase the effective source impedance of the microphone producing loss and resultant inefficiency.

A frequency response of a less-than-ideal transformer fed from a 200 Ω source and measured at high impedance across the secondary looks something like Fig. 29-38, where the low-frequency droop is attributable to one or both of the winding inductive reactances becoming comparable to signal impedances, while the high-frequency peak is an aforementioned secondary winding self-resonance. Usually the primary self-resonance is fairly well damped by the source impedance, but occasionally added cable capacitance can play cruel tricks here, too.

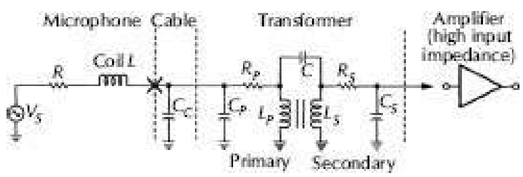


Figure 29-37. Transformer coupling model showing major elements.

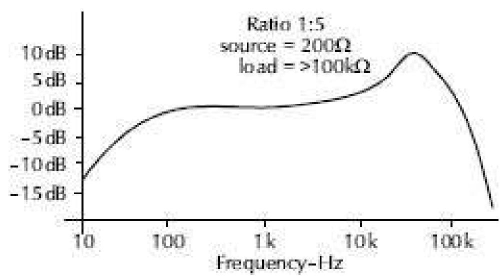


Figure 29-38. Typical transformer transmission response.

The mic-amp itself, as discussed, has a high-input impedance (hundreds of kilohms and up) while its optimum source impedance is defined at around 5–15kΩ.

It's good engineering practice to consider how the circuit behaves when the operating impedances are no longer defined by the microphone, i.e., when it is unplugged. Ordinarily, unless its layout and screening were extraordinarily good, and luck in plentiful supply, the circuit of Fig. 29-37 with the microphone disconnected would probably oscillate, as would any circuit with a high-gain, high input-impedance amp terminated only by the collection of vile resonances and phase-shifting elements that are an open-circuit transformer. An open-circuit impedance-defining resistor (R_o in Fig. 29-39) with a value 10 or 20 times that of the amp OSI, helps tame this. It also marginally tames the secondary resonance.

There are a variety of techniques for dealing with this resonance. They vary from pretending it doesn't exist to actually using it as part of a front-end, low-pass filter to keep ultrasonic garbage out of the electronics. Minimization of the high-frequency bump can be attempted passively, prior to the amp; the taming network in Fig. 29-39 represents a typical approach. Here, a series resistor-capacitor Zobel combination in conjunction with the open-circuit impedance-defining resistor is used. The values are calculated to produce a step-type response, Fig. 29-40, which when combined with the hump at the high-frequency end of the transformer response, produces a more acceptable roll-off characteristic. Naturally, the interreaction between this network and the complex impedance of the transformer is not quite that simple. The network capacitance reacts heavily with the transformer inductance, shifting the resonance frequency in the process. It is this fact that has led to the misconception that the capacitance somehow magically tunes out the resonance.

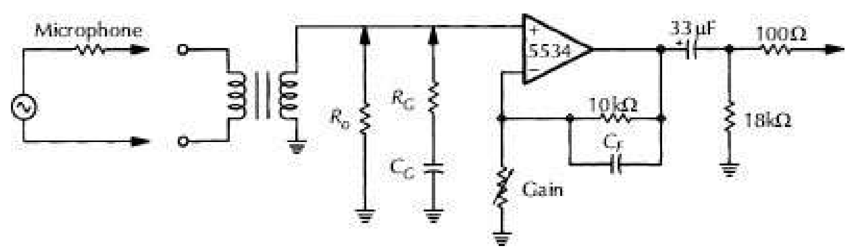


Figure 29-39. Basic microphone preamplifier with compensation components.

Open-circuit stability is dramatically improved, Fig. 29-40. The network takes an even larger slice out of the overall high-frequency response but unfortunately at the cost of a falling in-band termination impedance at the higher frequencies. An alternative approach, assuming the Zobel isn't actually necessary to achieve open-circuit stability, is to perform the response roll-off around a subsequent amplifier stage, or at least buffered away from the input itself. In practice, both passive and active techniques are used simultaneously to achieve stability and optimize in-band phase and frequency responses with the least-wounded high-frequency termination impedance.

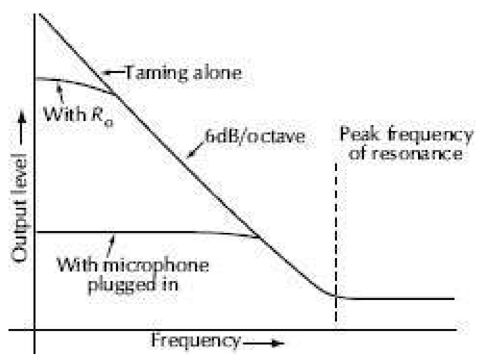


Figure 29-40. Frequency response of taming network.

29.10.6.4 Bandwidth

Providing the compensating high-frequency roll-off around a subsequent amplifier, in the form of exaggerated feedback phase-leading around the mic-amp itself in this case (CF), has the advantage that the noise performance of the combination at higher frequencies remains unimpaired by an impedance mismatch resulting from a passive network.

Problems result in several areas. Compensation around the mic-amp becomes limited when the electronic gain approaches unity, while compensation around a late fixed-gain stage means that all stages prior to it, including the mic-amp, have head room stolen at the frequency of the resonance and to a degree of the magnitude of the resonance. This may or may not be a problem depending on how far the lower side of the resonant curve invades the audio band.

The passive method reduces the magnitude of the resonance. The ultimate low-pass roll-off slope is that of the high-frequency side of the resonance, which more accurately is a lightly damped inductance-capacitance, low-pass, 12dB/octave filter. The active method uses an additional 6dB/octave curve in the compensation making a total of 18dB/octave, but it relies on the resonance being of a manageable degree to begin with. A measure of both techniques is usually required; their balance and relationship are an experimental process to optimize for each different type of transformer.

This enforced filtering is of considerable advantage, helping to keep all sorts of unwanted ultrasonic noise from finding its way into the mixer. It also represents a major advantage of transformer inputs over solid-state varieties.

A further advantageous filtering is the falling source impedance seen by the amplifier at extreme low frequencies. This is due to the fact that the winding inductive reactance reduces with frequency. This is a definite help in combating the generation of excess low-frequency noise in the first amplifier.

29.10.6.5 Common-Mode Transfer

There are two different amplitude response curves to be considered. The first, the normal differential input, has been fairly thoroughly determined. The second, by virtue of its mechanism, relies on imperfections within the main filter element itself (the transformer) rides over and oblivious to our carefully calculated filter responses. Common-mode unrejected signals still appear at the amplifier input as if nothing had happened.

29.10.6.6 Input Impedance

As determined earlier, we would end up with better noise performance and cleaner sounds if the microphone looked into a high, preferably infinite, impedance. Preferences apart, we have already had to define the reflected load (input), impedance by the resistor needed to keep the front-end stable under unplugged conditions (R_o), but at least it is an order of magnitude or so above working impedances, so its effect is small. It does, though, act as part of an attenuator of input signals along with the source impedance and winding losses, Fig. 29-41. This is the major factor responsible for worsening front-end noise performance using transformers. Any attenuation before the optimized amp directly degrades the noise figure, typically between 1 and 6 dB, depending on the transformer.

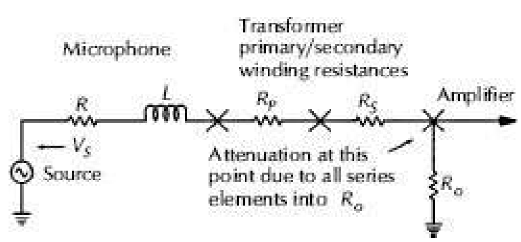


Figure 29-41. Input losses which worsen noise figure.

If the transformer were perfect, it could be assumed that the reflected impedance, as seen by the microphone, would be constant over the audio band. At the low-frequency end, Fig. 29-42, the diminishing inductive reactance of the transformer windings (tending to zero with frequency) becomes a term of greater importance, affecting parallel impedances, attenuation, and accordingly, efficiency. Winding self-capacitances and the passive compensation networks are largely to blame for the high-frequency droop, although the list of contributing mechanisms is nearly endless.

A good rule of thumb is that the midband input impedance should exceed ten times the source impedance, or about 2kΩ for a dynamic microphone. Any wild variation in this impedance is obviously going to result in frequency and phase response aberrations, which are probably the greatest single drawback to transformer front ends. Things aren't quite as bad as they seem; examples of performance shown here have been deliberately of a marginal transformer to highlight the ill-effects, notably in response and impedance flatness; good transformers from good reputable manufacturers such as Jensen, Lundahl, and Sowter generally show much nicer results, but the design criteria to eke the best from them remain nonetheless.

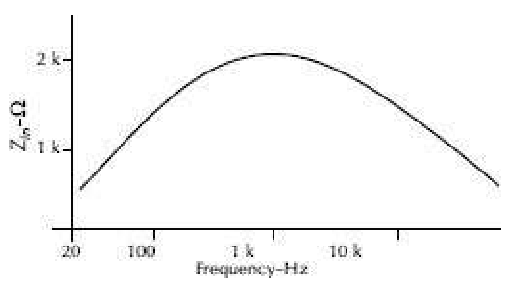


Figure 29-42. Typical input impedance curve.

29.10.6.7 Attenuator Pads

Attenuator pads, regrettably necessary in many instances to preserve head-room and prevent core saturation with elephantine sources, should maintain expected operating impedances when introduced. The transformer primary should still be terminated with a nominal 200Ω, while the microphone should still look at 2kΩ or above. Departure from these will cause the microphone/amplifier combination to sound quite different when the pad is thrown in and out, as would be expected from altering source and load impedances in and around complex filter characteristics. A significant downside to pads is that although the differential (desired) signal is being attenuated to the expected degree, any common-mode signal isn't.

29.10.6.8 Transformerless Front Ends

Bringing the amplifier optimum source impedance down to that of conventional dynamic microphones is possible by means other than transformers. Reducing the ratio of amplifier-inherent voltage and current noises has this effect. Two main techniques, either alone or in concert, are used:

1. Large-geometry devices have innately lower base spreading resistances and noise impedances. Even power-amplifier drivers have been used, e.g., 2N4918, BD538, but these tend to suffer from low transit frequencies (bandwidth) and beta (gain), which can lead to additional complexity in the circuitry.
2. Paralleling multiple identical input devices, and so proportionally increasing the noise current in relation to the noise voltage, reduces the ratio between them, i.e., noise impedance. Even the cheap and humble general-purpose 2N4403 so used is capable of surprisingly good results in this way.

PNP transistors, as mentioned elsewhere, have less surface-recombinant noise / lower base-spreading resistances than NPNs and are favored.

The usual technique is to place two of these large and/or multidevice input front-end amps—preferably accurately matched—ahead of an electronic differential amplifier, as shown in Fig. 29-43. All the amplifier gain is made within the first pair of stages, differentially cross-coupled. This gain arrangement, rather than referring to ground, can afford reasonable common-mode signal rejection. Differential input signals are amplified since the reference for each of the two amplifiers is the other amplifier, tied to an identical signal of opposite polarity.

If the input signals to the two amps are identical in phase and amplitude (common), the references for each of the amplifiers are similarly waving up and down sympathetically to the signal. There is no voltage difference for the individual amplifier to amplify; consequently, there is no gain. For ordinary differential input signals, the amplifiers operate conventionally, their ground reference being a zero voltage point half-way along the gain-determining variable resistor. This point is a cancellation null between the opposite sense polarity swings of the two amplifiers.

These amplifiers feed a conventional electronic differential amplifier running usually at unity gain. In order to maintain stage noise as low as possible, the resistors are made as low in value as the devices can sensibly stand. This arrangement is unmistakably a bastardized instrumentation amplifier—a well-documented circuit configuration; the only thing of remark is the pair of low-impedance optimized front-end stages.

A criticism (rightly) leveled at some implementations of this, including earlier-generation integrated parts, is that the noise performance is significantly worse at low amounts of gain than at high gains, where the (hopefully) optimized input pair reigns. This is down to two mechanisms:

1. The impedances chosen around the op-amp differential amplifier being too high; some resistor values here have been seen as high as 25kΩ, guaranteeing a high invariant noise floor for much of the gain range; bizarrely, some implementations attempt gain around this stage, too. The lesson earlier learned of keeping the circuit impedances as low as circumstances permit is salient here.

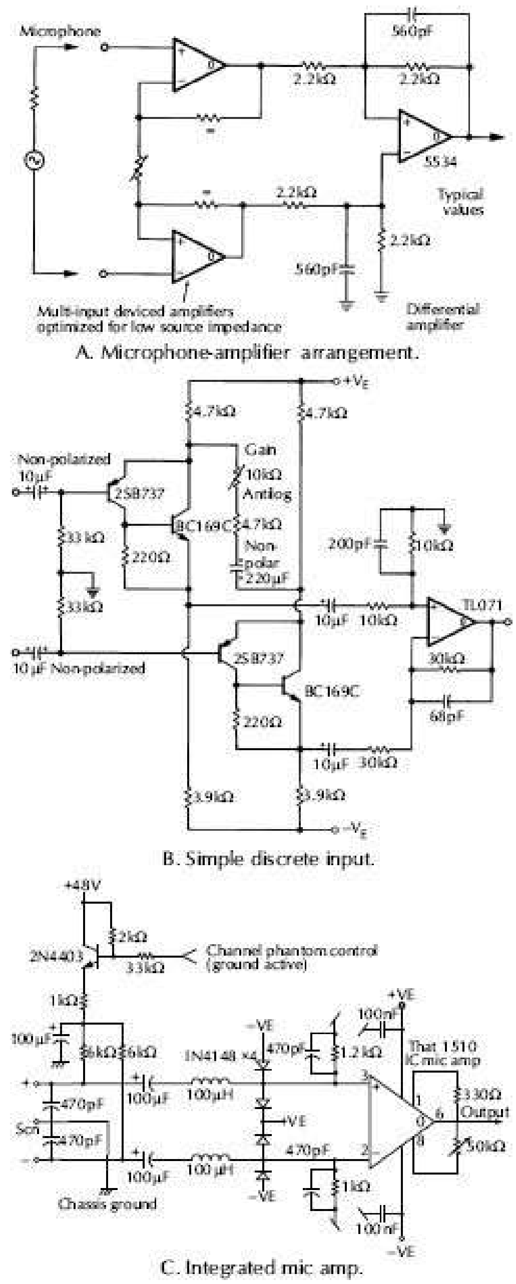


Figure 29-43. Basic transformerless microphone-amplifier arrangement.

2. What can best be described as "The Other Port" syndrome. Connecting the emitters of the input transistor sets is the gain-setting resistance. Each half of the differential amplifier sees effectively half that resistor's value (its center being a canceled

null). The input circuit for noise purposes consists of the input bias resistor (1kΩ), the transistors' base spreading resistance AND half of that gain determining resistance. Missing it is a common oversight, and indeed it's an issue with other topologies, too. Knowing this, it is easy to see how as the gain is reduced by the resistance being made larger, the noise resistance of the input circuit rises, too, inevitably if counterintuitively worsening the noise performance.

There are few reasons now for hand-knitting discrete versions of this input arrangement, and good ones not to, primarily suitable input devices: the much-favored 2SB737 is sadly becoming difficult to procure, and the excellent integrated matched transistor sets such as the THAT 320 are relatively expensive. Relative, that is, to the cost of a purpose-designed IC! There are a few integrated versions of this kind of arrangement offering very acceptable performance in the convenience of a little package; the Burr-Brown INA103 or 163 and the THAT Corporation's 1510 have multiparallel input transistor stages presenting OSIs about perfect for nominal microphone impedance, the latter part taking to heart the op-amp differential stage noise issue, with excellent results. Texas Instruments / Burr-Brown's PGA2500 is a digitally gain-controlled version of this configuration, as is the THATCorporation's 1570/5171 pair; this latter set addresses the Other Port Syndrome, by scaling the feedback and gain-setting resistances simultaneously with gain, so maintaining a very low noise figure irrespective of gain. These digitally programmable gain devices have 1 dB gain steps and inaudible gain-change (zipper) noise, solving a major headache for digital console designers and others desirous of remote control of mic gain. They have proved themselves worthy with ribbon microphones, the ultimate front-end-noise torture test.

With this transformerless input configuration, although potentially offering higher and flatter input impedances than transformer inputs, there are, as always, snags. Common-mode signals directly gobble up headroom in the first pair of stages even if they are operating as followers; that this common-mode stuff is substantially canceled in the following differential amplifier is a bit of a stable-door and bolted-horse routine. There is also the great danger that commonmode signals (in addition to normal differential signals) can exceed the input swing capability of the input devices. At best this will block the input stage, at worst—if the common-mode signal is big enough and at a low enough impedance (think ac line grounding fault here)—serious destruction can result. A transformer input would blithely ignore such.

Radio frequencies simply adore base-emitter junctions, and this configuration has them in abundance. Successfully filtering microphone inputs sufficiently without sacrificing noise performance or input device high-frequency gain (increasing high-frequency distortion and so on) is not a trivial task; it makes the self-filtering properties of an input transformer seem rather appealing.

Fig. 29-43C details the kind of helmet-and-armor with which one has to attire electronic microphone front ends to survive the fray:

- The "π" L/C filters and additional coupling to ground at high frequencies help against RF. The inductors can either be single pieces, one in each leg, or preferably a pair of windings around a common (usually toroidal) core. This has the twofold advantages that the choking effect is concentrated on common-mode signals

—the most common (so to speak) interference method—and that the inductances of the two windings essentially cancel for differential signals, so that there is much less effect of the RF protection impinging on the desired audio.

- The input dc-decoupling capacitors have to be pretty huge in value to maintain low-frequency integrity and at the same time have a high enough voltage rating to handle typically 48 V off-load phantom power voltage.
- The parametric varactor capacitance of the clamping diodes has little to no effect on the audio but are vital to protect the device front end from the very healthy whack as the phantom voltage is turned on or off. Even so, clamped or no, good common-mode rejection or no, phantom ramped or no, these transitions are not exactly subtle. These diodes may help protect the input from other nasties, too, but the aforementioned major ground fault would render all this toast anyway. The good news: such are rare but in the real world of touring sound not unthinkable.

29.10.6.9 Line Level Inputs

The reader is referred to Chapter 15 *Audio Transformers* for Bill Whitlock's excellent further coverage of real-world interfacing.

Transformers—good ones at least—are expensive, large, and heavy. Not so good ones are still relatively expensive, large, and heavy and represent a weak link in a modern signal path, with their low frequency distortion, hysteresis, and high/low frequency/phase response effects. Transformers are best used only where their impedance transformation capabilities, innate filtering, and excellent isolation properties are really needed. And then only really good ones. Most high-level interfacing applications within the confines of a studio environment justify neither their capabilities nor drawbacks.

The search was on for electronic equivalents to transformers for both input and output applications, a moderate degree of success being achieved early on for input stages with classic circuits such as Figs. 29-44 and 29-45. These are simple differential input amplifier and instrument amplifier configurations using op-amps.

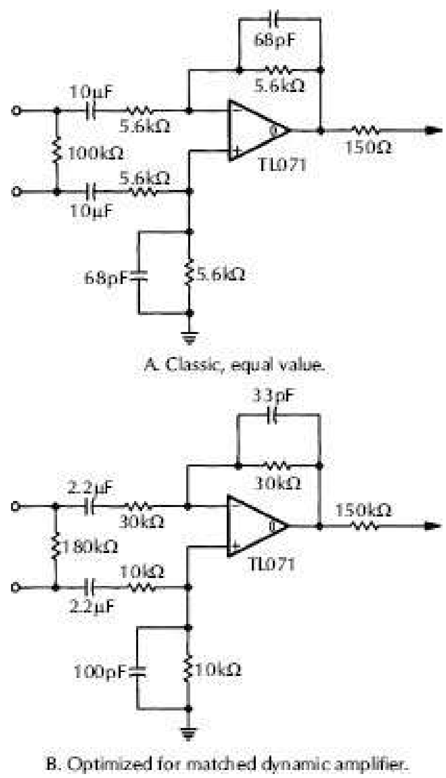


Figure 29-44. Electronic differential input amplifier.

Line inputs are commonly differential amplifiers, similar to the one used in the transformerless mic-amp, but with the resistor values elevated to bring the differential input impedance up to over the 10kΩ required of a bridging termination, Fig. 29-44. The noise of these stages is directly attributable to these resistor values, so the lower resistor values are better. An instrumentation amplifier configuration would seem to offer possibly better performance for noise (the differential amplifier resistor values may be kept small) but it entails the use of undesirable voltage followers (see section 29.7) with potential stability problems, input voltage swing limitations, and unprotected (for RF) input stages. At least with a simple differential amplifier the impedances are comfortably low and the inputs buffered by resistors from the outside world.

The dc-blocking series capacitors must, unfortunately, be large in value to maintain an even input impedance and sensibly flat phase response at the lowest used frequencies. Also, being necessarily unpolarized, they are physically large and expensive. This is a small price to pay, though, for such a simple but important circuit element.

The instrumentation amplifier presents very high, nonground-referred differential and common-mode terminations and has the great advantage that gain may be easily invoked between the two input amps at no cost to the excellent common-mode rejection, Fig. 29-45. Integrated line receivers are typically of this configuration.

A pair of inverting amplifiers, shown in Fig. 29-46, provides a simple, hardy, easily defined differential (but not true floating balanced) input stage. A fascinating circuit known as the Superbal input is depicted in Fig. 29-47; this is a balanced differential virtual-earth amplifier, referred to ground solely by one op-amp input and capable of very good common-mode rejection, limited by the tolerance of the components from which it is constructed. Accepting any lopsided input signal, it delivers a differential output perfectly symmetrical to ground, making it an exceptionally useful input conditioning amplifier.

The capacity of both these circuits to be differential virtual-earth points makes them ideal for use in balanced mixing bus systems.

29.10.6.10 Electronic Balanced Outputs

The simplest balanced outputs configuration is given in Fig. 29-48. This is a pure, no-nonsense, inverter-derived differential feed. For many internal interconnections and especially in differential balanced mixing systems it works well, but it should not be used to connect to the outside world.

Ideally, there must be no discernible difference in characteristics between the output circuit and an ideal transformer. After all, the fate of signals in the real world on a balanced transmission line won't alter in your favor simply because you've chosen not to use a transformer. If transformers are being supplanted it had better be with devices capable of affording similar benefits to the system and its signals. Regardless of applied reverse common-mode potential, the differential output potential must not change. Also the output should be insensitive to any imbalance in termination, even to the extent of shorting one of the legs to ground. This is the floating test. For example, the simple inverter circuit of Fig. 29-48 fails the floating test since, if one leg is shorted to common, the overall output has to drop by one-half (6dB). (The question of what happens to ground noise with a shorted amplifier bucketing current into it will be sidestepped here.) Two basic circuits have emerged as being close approximations to a transformer. Not only are they fairly closely related, but most balanced output topologies are also derived from them. They both depend on cross-coupled positive feedback between the two legs to compensate for termination imbalance.

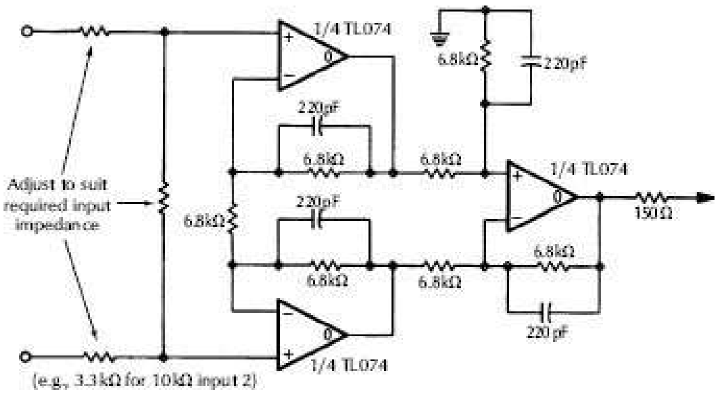


Figure 29-45. Instrumentation amplifier-type line-input stage.

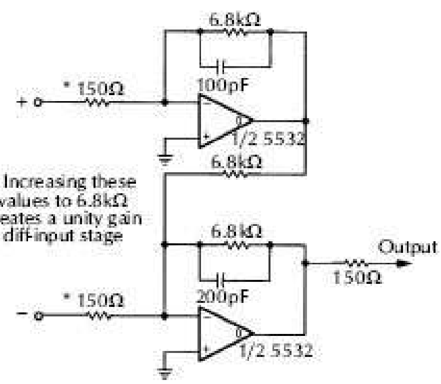


Figure 29-46. Differential mix/input amplifier.

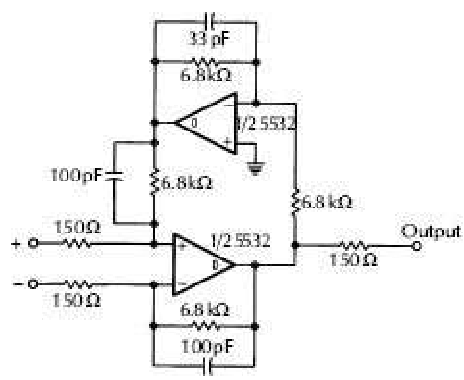


Figure 29-47. Superbal differential mix/input amplifier.

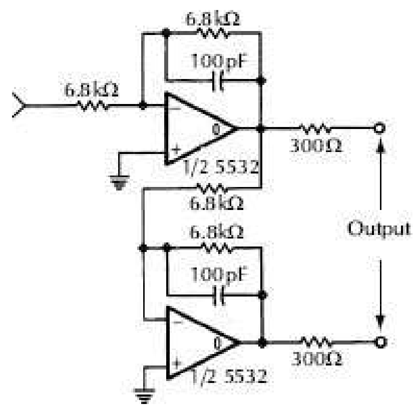


Figure 29-48. Inverter-type differential output.

In Fig. 29-49 a unity-gain inverting stage provides out-of-phase drive for the two legs, each output leg of which is a −6dB gain inverting amplifier with error sensing applied to its reference (positive) inputs. Under normal operation, there is no error-sensing voltage; the two inverse outputs cancel at the midpoint of the equal sense resistors. The two amps invert a differential voltage equal to the unbalanced input voltage appearing between their outputs. (Two −6dB quantities sum to make zero gain.) Take the case of one output, the upper one being shorted to ground. An error potential is derived of such a phase and level on the error-sense line that positive feedback increases the gain of the unshorted amp by 6dB, while matching on the positive input of the shorted one the signal on the negative input, canceling its amplification. Closing the shorted amp down prevents ground-current problems; therefore, any measure of output termination imbalance is reasonably dealt with by this arrangement.

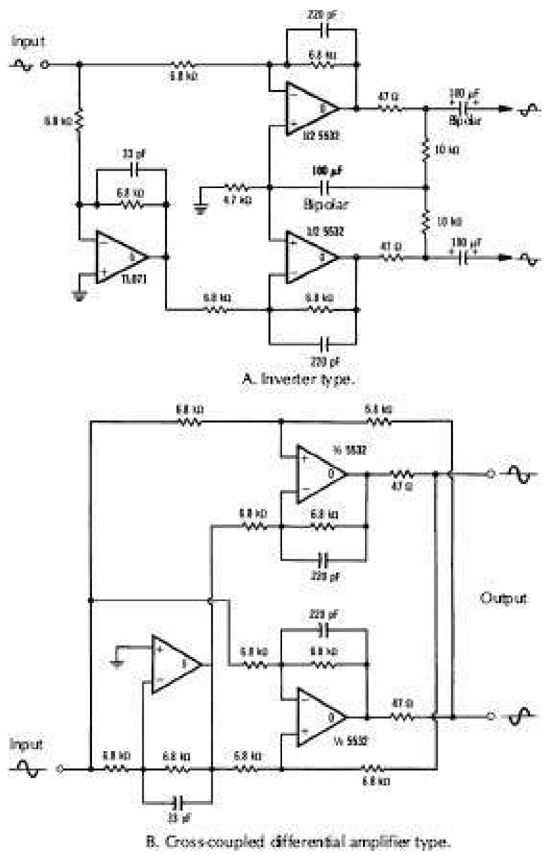


Figure 29-49. Electronic floating differential amplifier stage.

A major problem with any circuit depending on high levels of positive feedback such as these is their potential instability. Both these circuits are right on the edge of instability—they have to be in order to work accurately; a measure of margin has to be given for peace of mind and component tolerances. This backing-off compromise affects primarily common-mode rejection and output level against lopsided terminations. A loss of about 0.5dB in differential output level can be expected when one side is shorted to ground, although tight component tolerances can improve on this. Component tolerance imbalance—even if constructed with 1% resistors—manifests itself also as sometimes quite substantial dc offsets that will likely have to be trimmed to not eat up too much head room.

Curiously, instability tends to show itself as common mode. This fault manifested itself to the author for the first time (too) early one installation morning; a peak program meter (PPM) across such an output read nothing, listening elicited a little bit of hum, but a scope on either leg to ground showed 10Vp-p square waves, driving the tape machine to which the output was connected into shock.

Integrated versions of the cross-coupled circuit such as the SSM 2142 have the great advantage of extremely closely matched/trimmed resistor values, and hence far more predictable performance than discrete versions.

29.10.7 A Practical Microphone-Amplifier Design

Optimizing front-end sound is nothing more than shrewd judgment in juggling the nearly endless electronic operating conditions so that adequate performance is obtained over the wide range of expected and common input signals. Any wrinkles should be arranged to exert influence only under quite extraordinary operational conditions.

The microphone amplifier example described here, Fig. 29-50, is a somewhat developed version of a basic front-end design and is in grave danger of becoming an industry standard.

Initially most striking is the manner in which a single-track potentiometer is used to vary simultaneously the gains of two amplifying elements—the front-end (noninverting) stage and the succeeding inverting amplifier. Since the first stage is (as far as its inputs are concerned) a conventional noninverting amplifier, transformer input coupling is no more problematic than with simpler microphone amplifiers, e.g., Fig. 29-39, a standard generic microphone amplifier.

With maximum gain distributed between two stages, large gain is possible without any danger of running out of adequate steam at high frequencies for feedback purposes in either of the two amplifiers. This, incidentally, also makes for reasonably simple stabilization of the amplifiers, something not easily accomplished with simpler single-amplifier circuits achieving the same gain swing.

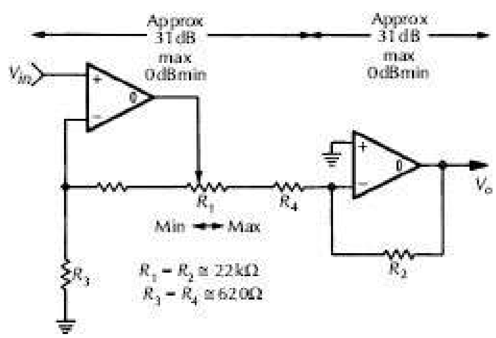


Figure 29-50. Shared-gain two operational-amplifier input stage.

A less obvious advantage to this two-amplifier configuration over a single stage is that dealing with the Other Port Syndrome is inherent; the Other Port in this case is the inverting input of the input amplifier B in the two-amp version it is defined at a fixed low impedance regardless (the 600 Ω resistor R_3) which is small and leaves the noise on that port defined almost solely by the device input noise voltage. The single stage however has its Other Port subject to the gain control value, leaving it with exactly the same worsening noise performance with reducing gain as the electronic front end described earlier.

Other than the obvious simplicity and economy of one-pot gain control, two nice features inherent in the design are interesting from the points of view of system-level architecture and operation.

29.10.7.1 System-Level Architecture

System-level architecture is largely concerned with operating all the elements of a system at the optimum levels and/or gain for noise and head room, i.e., at a comfortable place somewhere between the noise floor and clipping ceiling. Where gain is involved, it's important that the resultant noise be due primarily to the gain stage that has been optimized for noise (or rather lack of it) such that it can then mask all subsequent and hopefully minor contributions. At no point in the gain swing—particularly at minimum gain—should it be necessary to attenuate unwanted residual gain. This amount of attenuation gets directly subtracted from overall system headroom. What good is 24dB of head room everywhere else, if you have only 16dB in the front end?

In this respect circuits similar to Fig. 29-50 score well, and the graphs of Fig. 29-51 show why. Fig. 29-51A represents the gain in dB of a simple noninverting amp varying with the percentage rotation of an appropriately valued linear pot in its feedback leg. This is like the gain/rotation characteristic of the first amp of Fig. 29-50. Similarly, Fig. 29-51B is the gain/rotation plot for a linear pot as the series element in an inverting amp, such as the second gain stage of Fig. 29-50. For the first half of the rotation, the first stage provides all the gain swing and most of the gain; only about 6dB is attributable to the inverting stage at midpoint. Toward the end of the rotation, this position reverses with the front end remaining comparatively static in gain; the extra swing and gain come from the inverting stage. Noise criteria are met, since the first (optimized) stage always has more than enough gain to allow its noise to swamp the second stage, with the exception of minimum gain setting. There it hardly matters anyway because the front-end noise contribution is going to be at a similar level to the overall system noise floor (i.e., really quiet!). The impedances around the second stage largely determine the noise performance of the amplifier, and this is such that it need not be considered in relation to input noise at any sensible gain setting. Head room is satisfactory because no attenuation after the first gain stage is needed for any gain setting. The two gain stages operate nicely complementarily.

An operational advantage can be gleaned from Fig. 29-51C. This is the combined gain/rotation curve for the entire two op-amp circuit. Note that for a very large percentage of rotation around the middle of the gain swing (where it's most often used) the dB gain change per rotation is as good as linear. It gets a bit cramped at the top and bottom, but you can't win them all. For reference a little later on, it may be noted that there are two available resistors (R_2 and R_3) that may be used to modify the gain structure independently of the potentiometer.

29.10.7.2 Input Coupling

As a microphone amplifier, the fairly high optimum source impedance of the op-amp used in Fig. 29-52 (a Signetics NE5534, or AD797) needs to be matched to the target source impedance of some 200 Ω. No apologies are offered for the use of transformer input coupling, especially since they seem to be fashionable again after a long drought. Transformers despite their expense still offer advantages—especially simplicity, impedance step-up, protection, and filtering—over electronic inputs in this application.

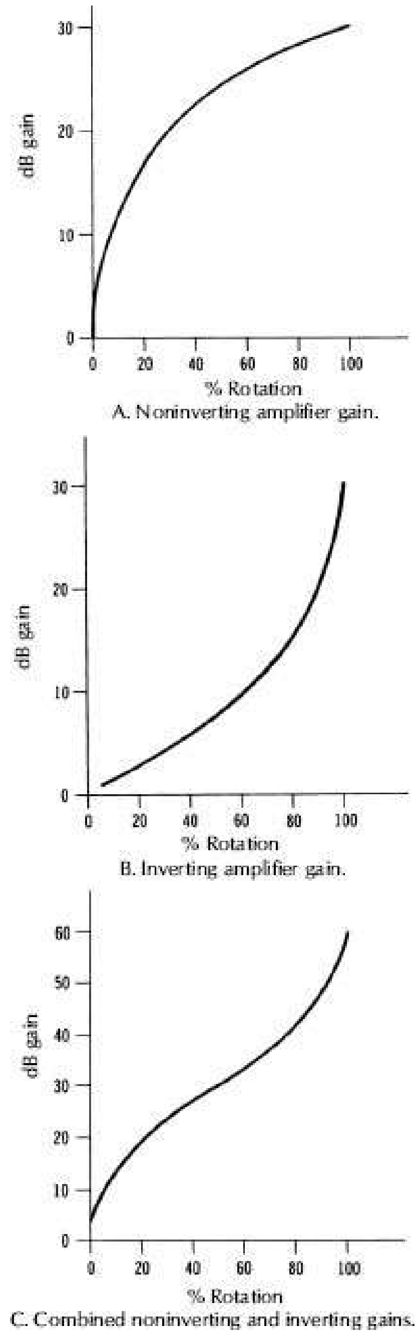


Figure 29-51. Gain versus pot rotation for two op-amp input stage.

Many circuit values (marked with an asterisk in Fig. 29-52, with some in quite unexpected places) are dependent on the specific transformer type in use. Several differing transformers can be very successfully used provided their differing ratios are taken into account in level calculations; a ratio of 1:7 is optimum to match the OSI of the input device employed. Phase and response trimming values will vary significantly. For example, with the Jensen JE-115-K, it is simpler than with the Sowter 3195 around which this circuit was originally developed. Despite the apparent simplicity of the circuit, a lot of effort has gone into defining the front-end bandwidth and straightening out the phase response at audible extremities. Taming the high-frequency transformer resonance in particular is quite tiresome, as earlier described.

On the front of the transformer hang the usual components to make the microphone amplifier useful in this world of capacitor microphones: a 20dB input attenuator and 48 V phantom power via matched 6.8kΩ resistors per leg carried common mode along the microphone line. Further to earlier discussions, the component values in the pad are chosen such that the microphone still sees the same general impedance whether or not the pad is inserted, while the mic-amp still sees about a 200 Ω source to keep all the transformer-based filtering in trim. It is essential the 6.8kΩ phantom resistors are matched; poor matching is a very easy way to wreck carefully won common-mode performance.

29.10.7.3 Line Level Input Facility

A line-in option is brought in via the transformer also. It features far stiffer input attenuation (about 36dB) while simultaneously disabling much of the gain swing of the first amp. The resultant gain swing of 35dB (between −25dBu and +10dBu input level) with a bridging-type input impedance of some 13kΩ should accommodate most things that the microphone input or machine-return input differential amp can’t or won’t. A small equalization network is used in the attenuator to bolster the extreme low-frequency phase response.

An alternative and in many ways preferable line-in arrangement might be to use either a discrete or integrated instrumentation amp-style stage, switched into the second stage of the mic-amp. In Fig. 29-52 this would come in where “from B-check Diff-Amp” is marked. This would avoid the necessity of using the transformer with attenuators, which of course does no favors to the common-mode rejection ratio.

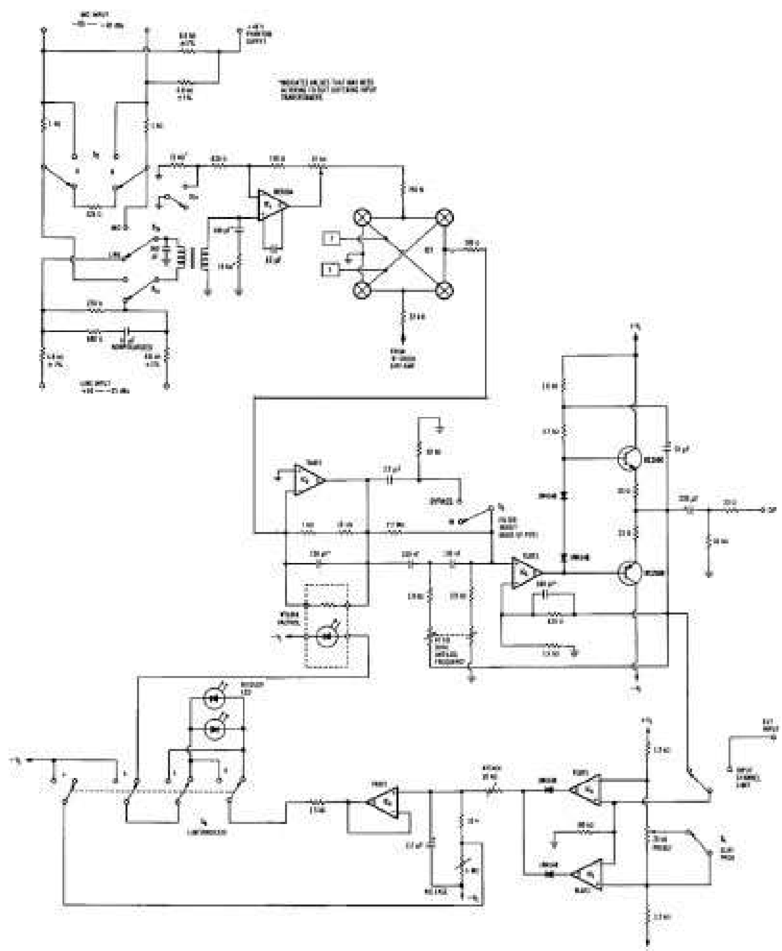


Figure 29-52. Channel input amplifier, high-pass filter, and limiter.

29.10.7.4 Common-Mode Rejection Ratio

Common-mode rejection ratio (CMRR) in the transformer is dependent mostly on the physical construction of its windings. The Sowter, in common with many other transformers, may be in need of compensation by deliberately reactively unbalancing the primary winding to match the inadvertent internal characteristics, Fig. 29-53. Jensen transformers are uncannily good in this respect—no tweaks usually being necessary. There are external circuit influences that can and will upset the maximum obtainable common-mode rejection. The accuracy of the phantom-power resistors is one; any input pad, regardless of accuracy, is another. Assuming any reactive, i.e., rising with frequency, common-mode response has been trimmed out, unequal phantom legs will enforce a lopsided flat common-mode response while true floating input pads instantly reduce the CMRR by nearly the amount of their attenuation. Why? They do this because they only attenuate the differential (wanted) signal and not the common-mode one. A halfway solution is to centrally ground reference the pad. Given all that, less than perfect common-made response shouldn't cause any ill manifestations in a typical recording environment with fairly short input leads. A high electromagnetic field of any sort, or an application with very long leads (or worse yet, a multicore), is far more likely to create problems with untrimmed inputs than with those properly balanced; vulnerability is greatly increased to all types of common-mode problems including noise on the phantom power-supply feed. Indeed, this is a common compounding of faults on a console that exhibits consistently noisy inputs.

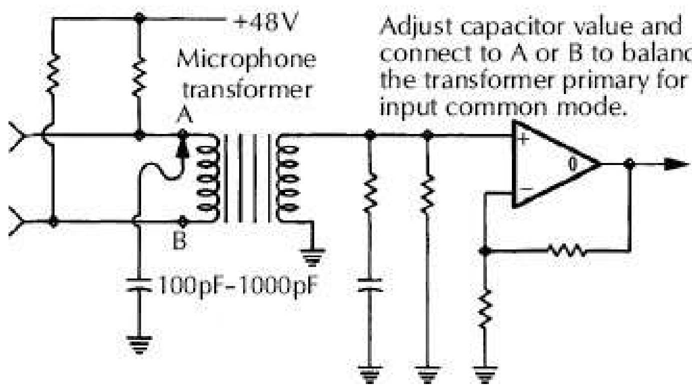


Figure 29-53. Input common-mode "tweak."

29.10.7.5 Minimum-Gain Considerations

A minor compromise is necessary in the first stage to prevent its gasping with exhaustion on extremely high input levels. Ideally, the output of the operational amplifier has to look into an impedance of 600 Ω or greater (this being the lowest impedance into which it can drive full-output voltage swing). Maximum gain state isn't really a problem. If the first stage is overdriven, then the second stage will be some 30dB into clipping; someone might notice!

At minimum gain though, the first stage in Fig. 29-52 is operating almost as a follower with an output load of 770 Ω, with the remaining feedback path to ground. That's safe and easily within the amplifier's driving capability. It would be better though if this small resistance were still smaller because it is contributing a little unwanted thermal noise to the otherwise beautifully optimized front end. The calculated degradation is only minor points of a dB and in practicality is easily lost in the gray mist that always surrounds the marriage of calculation with practical noise measurement.

The idea of using a front-end stage that turns into a follower under operating conditions has proved stable without any obvious trace of ringing within its bandwidth. This is probably because it is only being asked to look into safe, unreactive loads. Things that will make any unstable circuit squeal have not affected it. Among the instruments of torture have been a pulse generator/storage scope and an RF sweep generator/spectrum analyzer. The 10pF compensation capacitor is more an act of conscience than a practical necessity. No compromise comes from its use here, since at maximum gain the first amp is working 30dB below system level (an implied slew rate of nearly 200V/µs!). At minimum gain the incoming signal level is such that it's most likely coming from a line source of certainly much more limited speed than the front end.

Down from the nether world of megahertz, the microphone amplifier is totally stable for audio, even with the microphone unplugged and input unterminated; the input network (of RG and CG) is designed to work in conjunction with the fairly low-input impedance of the 5534 (150kΩ nominal).

29.10.7.6 The Limiter

Elaboration on the simple two op-amp mic-amp element consists of arranging an automatic gain control element in the feedback loop of the second amplifier and following that with a variable turnover frequency high-pass filter, Fig. 29-53.

A photoresistor device has its resistive end strapped across the normal gain-determining feedback resistor. Its resistance drops in value from very high (megohms) in inverse relation to the photodiode current to a limit of around 300 Ω at about 20 mA diode current. This resistance swing in the second amplifier is easily adequate for use in a peak limiter arrangement. The resistance change is close to exponential versus diode current, which could be of use in a gentler compressor, but here as a limiter the resistance change is quite sudden once that point is reached.

The limiter side chain is true symmetrical peak detecting, selectable to be able to pick off from either the high-pass filter output (as an input limiter) or from after the post-equalization breakpoint downstream (as a channel limiter). A positive-going and a negative-going level-detecting comparator are adjustable between clip detection (0.67dB before system head room) or program level (nominally) +8dBu.

A bicolor LED blinks red to indicate limiting in action, and it blinks green when the limiter is disabled to signify that the selected level (clip or program) is being reached or exceeded. In this indicate mode, the limiter integration time constant is deliberately shortened to make the green flashing similar in character to the red flashing in limit.

The difference is due to the nature of servo loops, of which a feedback limiter such as this is an example. In limit, the loop is self-regulating, the gain-control element holding back the audio level so that it's just tickling and topping up the side chain. In indicate, the loop is broken, and there is no such regulation. The green light stays on whenever the threshold is exceeded and tends to hang on while the time-constant capacitor discharges. With even a minor overdrive, this hangover could extend for quite a few seconds; hence, the shortened time constant.

This limiter is not subtle. It's intended for crash protection. The comparators deliver a full-sized, power-supply wallop to the integrator upon threshold, softened a bit by the attack preset in conjunction with the output impedance of the comparators. This rather unusual approach is to help wake up the photoresistor that has a relatively leisurely response time. The combination can be adjusted to be slow enough such that it doesn't clip yet fast enough to prevent an audible snap. Overshoot is generally within 1 dB on normal program, given a release time long enough to prevent pumping.

As a rough guide, if it's intended to use such a limiter for sporadic transient protection, it's best to aim for short attack and release times, bearing in mind that such settings will behave more as a clipper to the lower frequencies. For continual effect use, longer time constants will be less gritting and more buoyant. This side-chain arrangement certainly behaves differently from more conventional FET or voltage-controlled amplifier (VCA) linear proportional systems and needs a slightly different approach in setting up.

29.10.7.7 High-Pass Filters

Constructed around the line output amplifier of the front end in Fig. 29-52 is a second-order high-pass filter. It is a completely ordinary Sallen-Key type filter, arranged to use a dual-gang equal value potentiometer to sweep the 3 dB down turnover frequency from between 20 Hz and 250Hz. A click-stop switch at the low-frequency end (counterclockwise) negates the filter, replacing it with a very large time-constant, single-order dc decoupler. These are both tied to reference in order to minimize clicks. Fortunately, the BiFET op-amp in the filter barely uses any input bias current, so there is little developed offset voltage from that source to worry about.

Being an equal-value filter, the Q or turnover would be very lazy indeed if the feedback were not elevated in level to compensate for the upset resistor ratio. Here a compromise is struck. A low Q gives a very gentle rolloff (which is sonically good), and high Q results in more rapid attenuation beyond the cutoff frequency at the expense of a more disturbed in-band frequency response—pronounced bumps—and frantic temporal and phase responses exhibited as ringing and smeared transients. Luckily, the majority of control-room monitors exhibit far worse characteristics at the low-frequency end.

A maximally flat response midway between the two extremes is chosen by an appropriate amount of elevated feedback (around 4dB). This gain is taken across the filter as a whole, with the second stage of the microphone amplifier arranged to sustain a 4dB loss to compensate. It all works out in the end, with no compromise of head room. With minimum gain set, there is still unity electronic gain front to back. An added convenience of gain is that it provides a better chance of shoring up feedback phase margin, which is quite important in a line amp that may have to drive a lot of heavily capacitative cable. Also, it provides yet another single-order low-pass pole to help smooth out the high-frequency resonance of the microphone transformer.

29.11 Equalizers and Equalization

The term equalization is strictly a misnomer. It was originally utilized to describe the flattening and general correction of the response of systems in which by a matter of course or design had deviated from the original shape, e.g., telephone lines and analog tape machines. (In the latter case, equalization refers to the adjustment tweaks to the preemphasis and deemphasis curves—not necessarily the curves themselves.)

In search of a name for the deliberate modification of amplitude and phase versus frequency responses for taste and the occasional genuine creative effect, the contraction EQ is well understood as both a noun and a verb.

This sonic mutilation uses frequency response curves and shapes in degrees that have grown through an uneasy mixture of operator needs and technical expedience/feasibility. One of today's multiparametric console channel EQs would have needed a rack full of tubes in the fifties and sixties. Funny, they didn't seem to need such EQs then.

The delight (and maybe curse) of IC op-amp design is that active filter (hence, EQ) implementation and techniques have blown wide open, limited only by economics, the largeness of the printed circuit board and the smallness of the user's fingers.

EQ curves can be roughly lumped into three user categories: garbage disposal, trend, and area. High-pass, low-pass, and notch filters that eliminate air-conditioning burble, mic-stand rumble, breath noises, hum, TV monitor line-frequency whistles, and excessive electronically generated noise are obviously in the business of garbage disposal. Fig. 29-54A shows the sorts of responses to be expected from these. Gentle hi-fi-type treble and bass slopes and similar shelving curves establish response trends shown in Fig. 29-54B, while resonance like, bell-shaped lift-and-cut filters manipulate given areas of the overall spectral response, Fig. 29-54D. These are used to depress unwanted or irritating aspects of a sound or, alternatively, to enhance something at or around a given frequency that would otherwise be lacking. As the curves differ, so do the design techniques required.

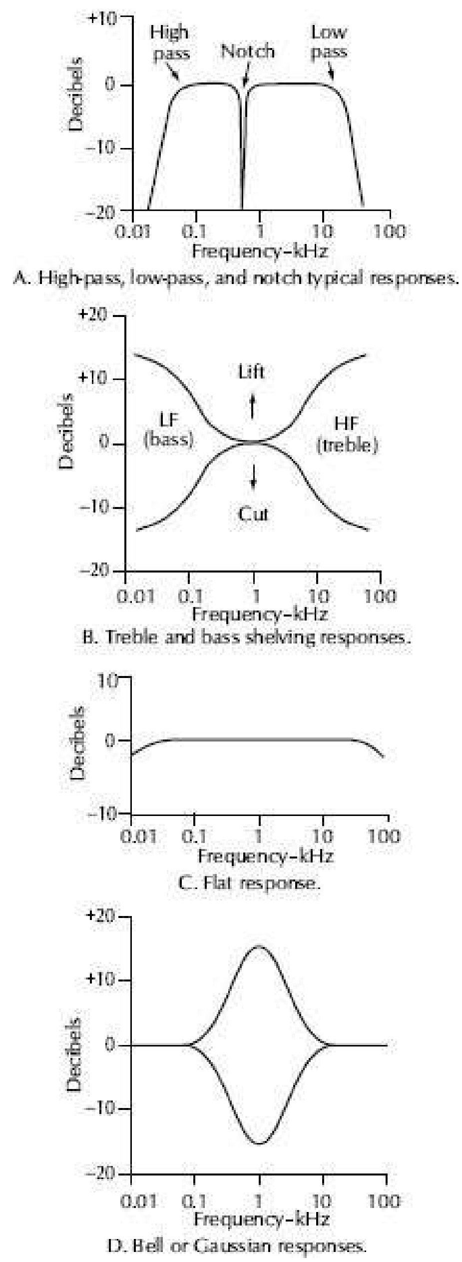


Figure 29-54. EQ responses.

29.11.1 Single-Order Networks

You can't build a house until you have the bricks, so they say. Fig. 29-55 has those bricks in the form of combinations of basic passive components with a rough guide to their input-output voltage transfer functions (essentially the frequency responses).

Assumptions are that the V_{in} source impedance is zero and the V_o termination is infinite impedance.

Capacitative reactance decreases with increasing frequency, working against the resistance to increasingly short the output to ground with increasing frequency in Fig. 29-55A, while in Fig. 29-55B the capacitance steadily isolates the output from the input with reducing frequency (rising reactance).

Inductors have entirely the opposite reactive characteristics. Inductive reactance is directly proportional to frequency, so the curves in Figs. 29-55C and D will be of no surprise at all, being complementary to those involving capacitance.

29.11.2 Single-Order Active Filters

Further useful curves are derived when the passive R, C, and L elements are wrapped around an op-amp in the classic inverting and noninverting amplifier modes, as shown in Figs. 29-55E through L. All the curves in Fig. 29-55 are normalized to unity gain and the same center frequency at which the curve departs significantly from flat.

Standard arithmetic formulas normally consider or obtain a frequency at which the curve has departed 3 dB from flat (the 3 dB down point) being usually also where the phase has been shifted 45°. This is only partially useful in the design of filters for use in practical EQs; the departure point, or turnover frequency, is generally more relevant.

29.11.3 Changing Filter Frequency

With any of these filters, moving the frequency at which the filter bites can be achieved by altering any of the R, L, or C values. Making any value smaller moves the frequency higher, while making the value larger moves the frequency lower.

There are an endless number of combinations of element values to create the same curve at the same frequency. In Fig. 29-55A if the value of the capacitor were reduced (increased in reactance), the filter curve would shift up in frequency. A corresponding proportional increase in the series resistor value would result in the original turnover frequency being restored; we have an identical filter with a different resistor/reactor combination. What does remain the same is the ratio or relationship between the two elements. It is only the filter impedance (the combination of resistance and reactance) that varies.

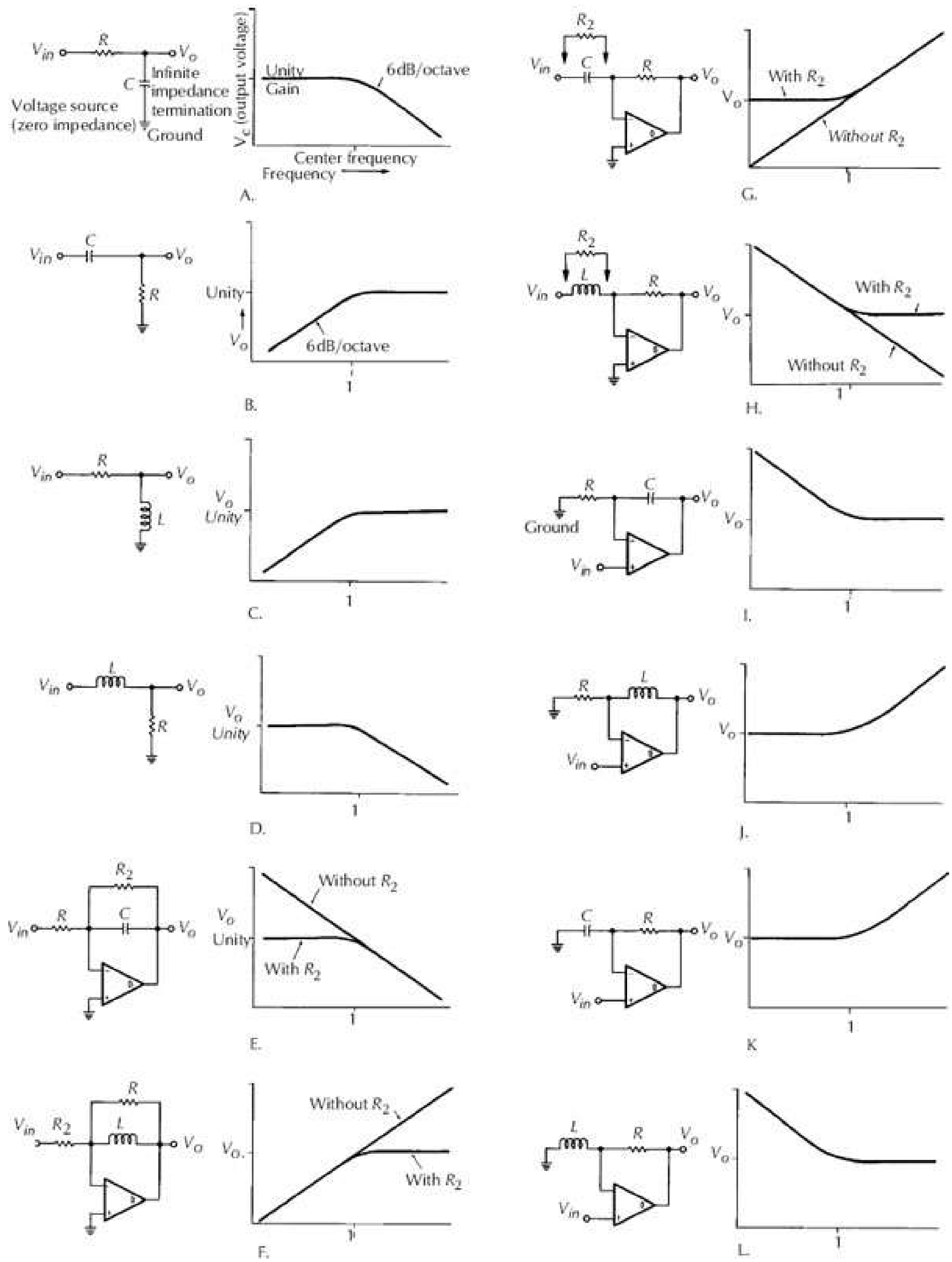


Figure 29-55. Single order filters.

With the exception of a few, the operation of any active filter can eventually be explained by referring to these basic single-order filter characteristics in Fig. 29-55.

There is one particular combination of two reactive elements (capacitance and inductance) that is of prime relevance to the construction of EQs. This, a series-tuned circuit, Fig. 29-56, is where things really become interesting.

29.11.4 Reactance and Phase Shifts

In, for example, the context of a simple resistor/reactor filter, Fig. 29-55A, the reactance not only causes an amplitude shift with frequency but also a related phase shift. A fundamental difference between the two types of reactance (C and L) is the direction of the output voltage (V_o) phase shift with respect to the source (V_{in}). More specifically, the capacitor in Fig. 29-55A causes the output voltage phase to lag farther behind the input as the roll-off progressively bites to a limit of −90° at the maximum roll-off of the curve, while the inductor of Fig. 29-58C imposes an increasing voltage phase-lead as the low-frequency roll-off descends with a limit of +90° at maximum attenuation.

The two reactances, in their pure forms, effect phase shifts of +90° to −90° to an ultimate extent of 180° opposed; they are in exact opposition and out of phase with each other.

Referring again to Fig. 29-56, a slightly different light shines. The two reactances are working in direct opposition to each other with the inductive reactance trying to cancel the capacitative reactance and vice versa. Arithmetically, it is surprisingly simple with the two opposing reactance values directly subtracting from each other; the combination network behaves as a single reactance of the same reactive character as the one predominant in the network.

For example, if for a given frequency, the inductive reactance is a +1.2kΩ (the + indicating the phase shift character of inductance) and the capacitive reactance is −1.5kΩ, then the effective reactance of the entire network is that of a capacitor of −300 Ω reactance.

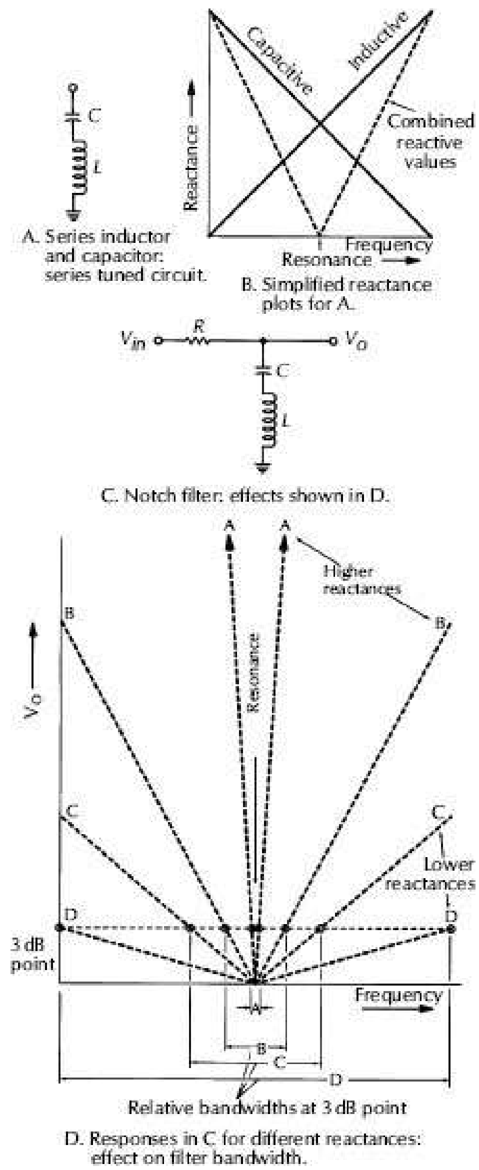


Figure 29-56. Series resonant circuits.

29.11.5 Resonance

Resonance is the strange state where the reactances of both the *L* and *C* are equal. For any inductor-capacitor pair at resonance, the two reactances will be equal. If you subtract two equal numbers, the answer is zero. So, for the series tuned-circuit arrangement of Fig. 29-56A at resonance, there is no impedance. The two reactances have canceled themselves out. It is a short circuit at that one frequency of resonance, disallowing component losses and is, in effect, a frequency-selective short circuit. Either side of that frequency, of course, one or the other of the reactances becomes predominant again.

29.11.6 Resonant Q

Like the single-order networks, there is an infinite number of combinations of *C* and *L* at any given frequency that will achieve resonance, i.e., the two reactances are equal. Similarly, it is the scale of impedance that alters with such value changes; the magnitude and rate of change of reactance on either side of resonance (off tune) hinges on the chosen combination.

At resonance, although the two reactances negate each other, they both still individually have their original values. Off resonance, their actual reactances matter. If each of the reactances is 400 Ω at resonance, then 10% off tune either way they are going to become 440 Ω and 360 Ω, respectively. A 10% change in this instance equates to about a 40 Ω change either way, up or down. Now imagine that a smaller capacitor and a larger inductor were used to obtain the same resonant frequency. Their reactances will be correspondingly larger. If they're five times larger with reactances of 2kΩ each, then at 10% off tune their reactances will become 2.2kΩ and 1.8kΩ or 200 Ω change each. The higher the network impedance, the more dramatic the reactance shift off tune.

On its own, the series-tuned circuit with whatever impedances are involved doesn't amount to much; however, in relation to the outside world, it becomes rather exciting. In Fig. 29-56C the series-tuned circuit is fed via a series resistor with the output being sensed across the tuned circuit. Fig. 26-58D shows input-output curves for three different tuned-circuit impedances based on low, medium, and high reactances with the series resistor kept the same in all cases. The detune slopes are steeper with higher reactance networks than with lower ones. In other words, higher reactance networks have a sharper notch filter effect, less bandwidth, and are said to have a higher Q (quality factor) than lower reactance networks. In all cases, the output sensed voltage would be the same as measured across a single reactance of the appropriate and predominant sort; there is no magic about a series-tuned circuit other than the curious subtractive behavior of the two reactances.

29.11.7 Bandwidth and Q

There are direct relationships between the network reactances, the series resistance, the bandwidth, and the *Q*. *Q* is numerically equal to the ratio of elemental reactance to series resistance in a series-tuned circuit (*Q' X/R*); on a more practical level, the *Q* can also be determined as the ratio of filter center frequency to bandwidth (*Q ' f/BW*). Bandwidth is measured between the 3 dB down points on either side of resonance (and usually where the phase has been shifted ±45°). If a tuned circuit has a center frequency of 1 kHz and 3 dB down points at 900Hz and 1.1kHz (pedantically 905Hz and 1.105kHz), the bandwidth is 200Hz and the network *Q* is 5 (frequency/bandwidth). The greater the *Q*, the smaller the bandwidth.

The filter resonant frequency may be altered by changing either the inductance or capacitance. Q is subject to variation of the resistor or simultaneously juggling the reactances in the inductance-capacitance network, while maintaining the same center frequency.

29.11.8 Creating Inductance

It is most efficient (electrically and financially) in the majority of console-type circuitry for inductance to be simulated or generated artificially by circuits that are the practical implementation of a mathematical conjuring trick. These are known generically as gyrators.

A true gyrator is a four-terminal device that transmutes any reactance or impedance presented at one port into a mirror image form at the other port, Fig. 29-27A.

A capacitor on the input (with its falling reactance versus frequency) creates inductance (with a rising reactance versus frequency) at the output port. The scale of inductive reactance generated may be easily and continuously varied by altering the internal gain-balance structure of the gyrator in Fig. 29-57B by changing the transconductance of the back-to-back amplifiers, creating a continuously variable inductor.

Real inductors have a justifiably bad name for audio design, sharing transformers' less pretty attributes. They are big and heavy and they saturate easily and nastily. Their core hysteresis causes distortion, and they are prone to pickup of nearby electromagnetic fields (principally power line ac hum and RF unless well screened, which makes them even bigger and heavier). The windings and terminations are prone to break. And they are expensive.

It is quite easy to see why it is popular to avoid using real inductors. Naturally, the simulated inductive reactance is only as good as the quality of the capacitative reactance it is modeled upon and the loading effect of the gyrator circuit itself. Degradation of the inductance takes the general form of effective series lossy resistance, the Q of the inductors suffering (Q ’ X/R). Leakage resistance across or through the image capacitor is partially to blame here. Fortunately, for the purposes of normal equalizers, very large Qs are neither necessary nor desirable, so selecting capacitor types to this particular end is hardly necessary.

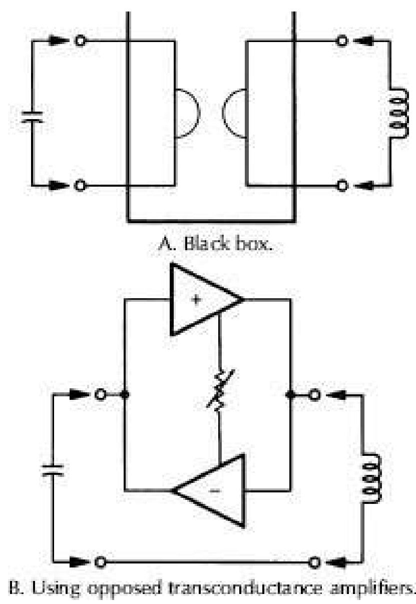


Figure 29-57. Gyrators.

An obvious extension of the continuously variable inductor is the continuously variable bandpass filter formed by adding a capacitor either in series or parallel with the gyrated inductor, forming series- and parallel-tuned circuits to make notch and peak filters, respectively. Although ideal for fixed-frequency filters with the Q of the network or sharpness defined by a resistor in series with the gyrator resonator, the idea falls down when the resonance frequency is moved.

If the frequency is moved higher by altering either the *L* or *C*, the reactances of the element at resonance become lower; consequently, the ratio of the reactances to the fixed-series resistor (this is the ratio that determines the *Q*) becomes smaller, and the bandwidth of the filter becomes broader in response. In order to maintain the same *Q* over the projected frequency variation, the series resistor has to be ganged with the frequency control, which is not easy. Should it be necessary to make the *Q* a variable parameter also, as in a parametric-type EQ section, it would mean devising quite a complex set of interactive variable controls. For this reason parametric-type EQ sections are ordinarily constructed around second-order, active-filter net-works, not individual tuned circuits whether real or gyrated.

29.11.9 Gyrator Types

Let us not write off gyration for functionally variable filters immediately. As we'll see, they form in one way or another the second reactance in many active filters.

True gyrators of the back-to-back transconductance amplifier variety are difficult to make, set up, and use. Fortunately, there are simpler ways of simulating variable reactances—if not pure reactances at least a predictable effect of a reactive/resistive network.

29.11.10 The Totem Pole

Fig. 29-58 performs the magic transformation of the single capacitor C_1 into a simulated inductance between the terminals.

Although emulating quite a pure inductance when set up properly, it is precisely that setting up that is not altogether straightforward. In fact, it is high on a list of circuits most likely to do undesired things.

29.11.11 The Bootstrap

The simplest fake inductor is shown in Fig. 29-60A, with typical values. It relies on a technique called bootstrapping. The principles are shown in Fig. 29-59. A Ω resistor with 1V across it will pass 1 mA. Without changing the source potential of 1V, the bottom end of the resistor is tied to 0.8V. There is 0.2V across the resistor, and so a current of 0.2 mA flows through the resistor. The source (still at 1V) sees 0.2mA flowing away from it, the amount of current it would expect to see going to a 5kΩ resistor value (1V/0.2mA = 5kΩ). It thinks it's looking at a 5kΩ resistor! Continuing this, stuffing a potential of 1 V (not the same source) at the bottom end of the resistor means there is no voltage across the resistor, so there is no current flow. Our original source thinks it's seeing an open circuit (infinite resistance) despite the fact that there is still a definite, real, physical 1 kΩ resistor hanging on it.

This phenomenon holds true with any source voltage, ac or dc, provided the instantaneous bootstrap voltage is the same as the source. Any phase or potential difference creates an instantaneous potential difference across the resistor; current flows and an apparent resistance materializes.

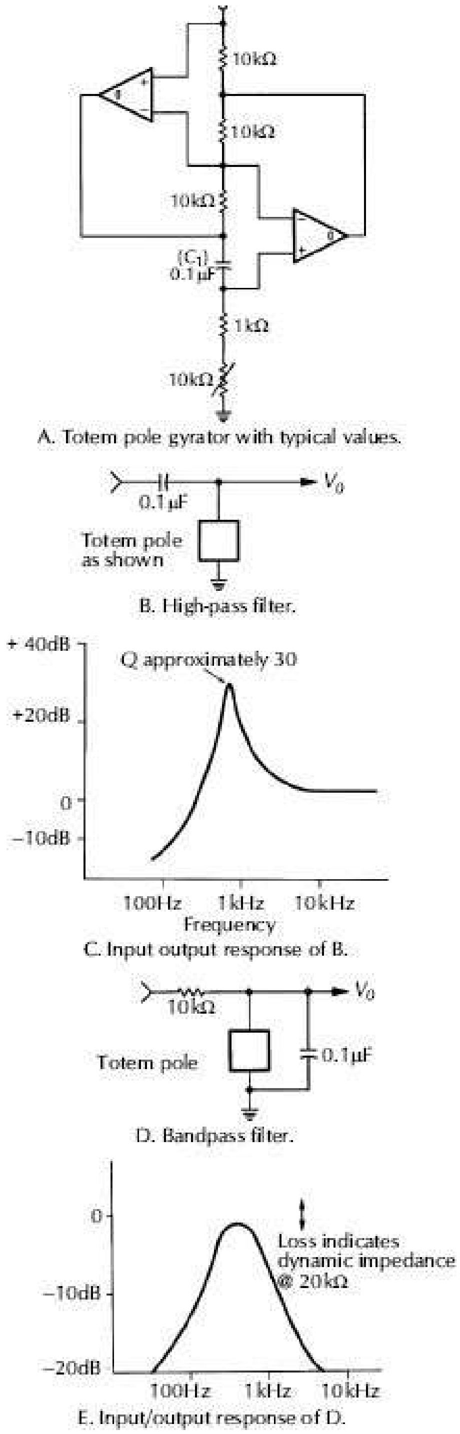


Figure 29-58. Totem-pole gyrators.

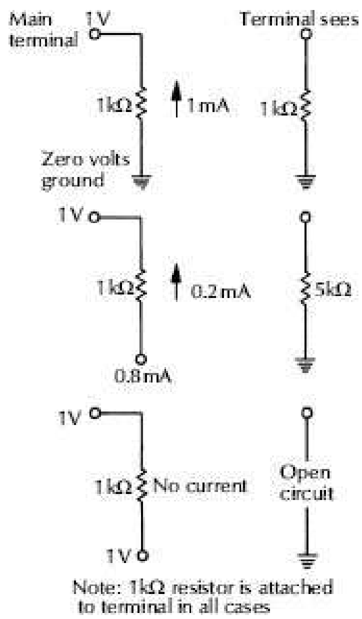


Figure 29-59. Bootstrapping analysis.

This fake inductor works on frequency-dependent bootstrapping, the terminal being almost totally bootstrapped to high impedance via the 150 Ω resistor at high frequencies and the bootstrap voltage reducing (together with its phase being shifted) with falling frequency. At very low frequencies the capacitor behaves as a virtual open circuit. No bootstrap exists, so the terminal is tied to ground via the 150 Ω resistor and the effectively zero output impedance of the voltage follower. The circuit emulates an inductor reasonably well; it has a low-impedance value at low frequencies, increasing with frequency to a relatively high impedance.

A minor failing with this simple circuit is that at high frequencies a parallel impedance (consisting of the variable resistor and capacitor chain) hangs directly from the terminal to ground. Buffering the chain from the terminal by a follower eliminates this, Fig. 29-60C.

Fig. 29-60A creates an analog of an inductor with the losses shown in Fig. 29-60B. The series resistor is the 150 Ω bootstrap resistor; after all, a proper inductive reactance tends to zero at low frequencies, not 150 Ω. The resistor is in series with the faked inductance tending to make it seem somewhat lossy or have a lower *Q* than a perfect inductor. If a fake inductor can be said to have winding resistance, this is it! The R/C network across the lot represents, again, the high-pass filter impedance, which on the addition of the follower disappears to be replaced in Fig. 29-60D by the much greater input impedance of the follower, which is high enough to be discounted.

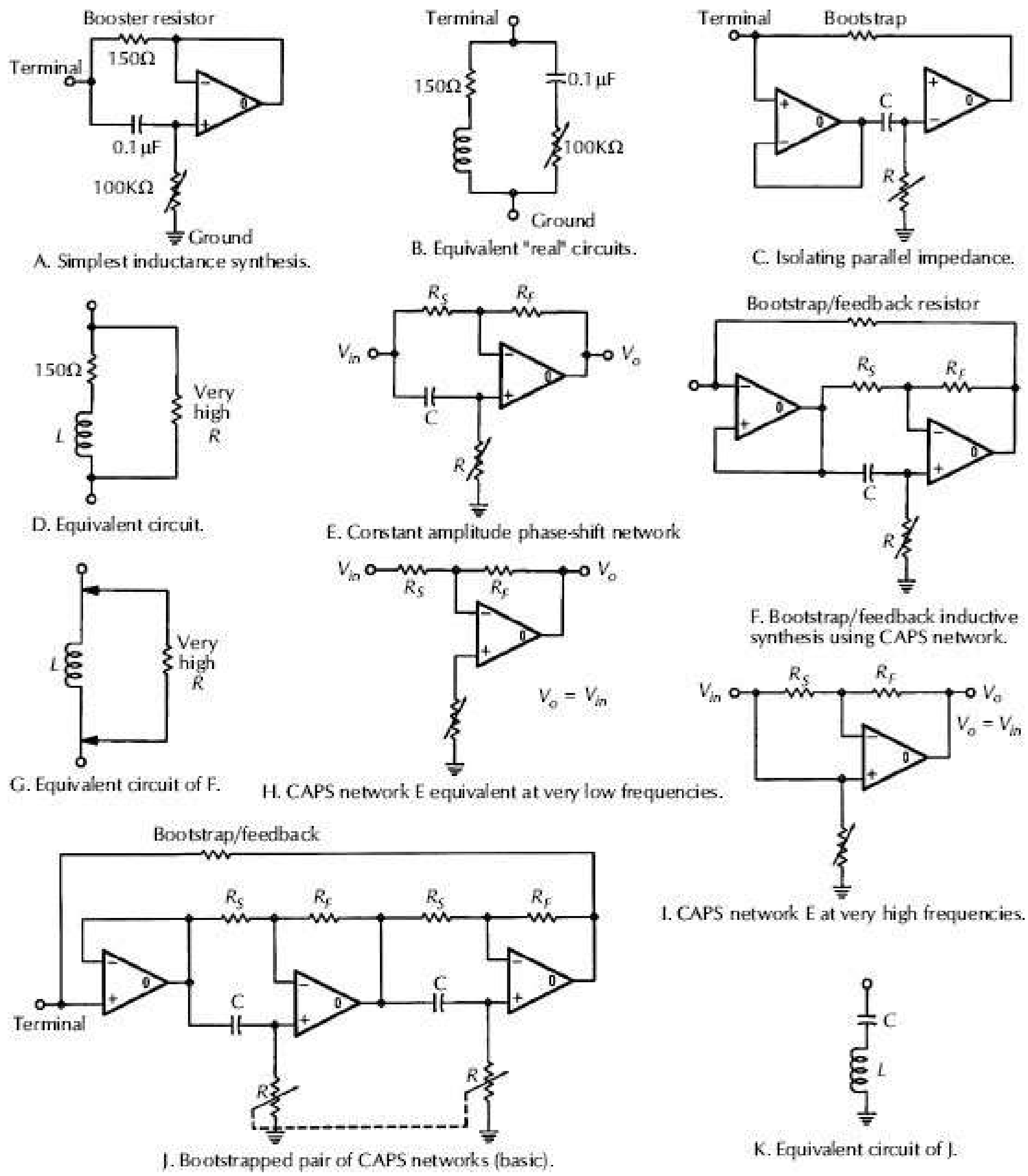


Figure 29-60. Inductive reactance synthesis.

As a short footnote to this gyrator epic, consider what happens to either Fig. 29-63C or F if the high-pass resistance-capacitance filter is replaced by a low-pass filter by swapping R with C. It may seem a bit strange to use circuitry to imitate a capacitor, but imitating a continuously variable capacitor does make sense. Real variable capacitors of the large values needed in EQs (yet easily created by gyrators) simply don't exist otherwise.

29.11.12 Constant-Amplitude Phase-Shift Network

A constant-amplitude phase-shift (CAPS) circuit of previously little real worth (other than for very short time delays or as an allpass filter) is shown in Fig. 29-60E. Bearing more than a little resemblance to a differential amplifier, this circuit can rotate the output phase through 180° with respect to the input, around the frequency primarily determined by the high-pass RC filter. Additionally the input and output amplitude relationship remains constant throughout.

How? This is dealt with in Figs. 29-60G and H where the simplistic assumptions that a capacitor is open circuit at low frequencies and a short at high frequencies show that at low frequencies the circuit operates as a straightforward unity-gain inverting amplifier (–180° phase shift), while at high frequencies it operates as a unity-gain noninverting amplifier (0° shift). The mechanism for the latter mode is interesting. The op-amp is actually operating at gain-of-two noninverting; this is compensated for by the input leg also passing through the still operating unity inverting path, which naturally subtracts to leave unity gain, noninverting.

29.11.13 Simulated Resonance

Detailed up to here are all the variables needed to create single- and second-order filters. Higher-order networks can be made with combinations of the two. Tracking variable capacitors and inductors allows the design of consistent Q bandpass filters irrespective of frequency. This eventually leads to a dawning of understanding in how the much-touted integrator-loop filters such as the state variable actually operate. The clue lies with the 180° phase-shift circuit of Fig. 29-60E. Connecting two such filters (with the variable resistor elements ganged) in series produces a remarkably performing circuit. At any frequency within the design swing, it is possible for the circuit output voltage to be exactly out of phase with the source –180° phase shift). By summing input and output, direct cancellation at that frequency and at no other is achieved. In short, a variable-frequency notch filter with a consistent resonant characteristic results. Alternatively, bootstrapping the input from the output actually changes that input port into something that behaves exactly like a series-tuned circuit to ground, Fig. 29-60J. The circuit is continuously variable in frequency with a consistent *Q* by virtue of the simultaneously tracking simulated inductor and capacitor maintaining exactly the same elemental reactances at whatever the selected operating resonant frequency. This creates the same source resistance, same reactance, same *Q*.

29.11.14 Consistent Q and Constant Bandwidth

The same *Q* definitely does not imply the same filter bandwidth. As the resonant frequency changes, the bandwidth changes proportionally. Bandwidth is, after all, the ratio of frequency to *Q*. Some active filters, such as the multifeedback variety, exhibit a constant bandwidth when the resonant frequency is changed: a 10:1 variation of center frequency, a 10:1 variation of *Q*. This, of course, is rarely useful for real EQ; it is noteworthy though in that the change in *Q* with frequency happens in the opposite sense to that expected from a normal variable tuned circuit. The *Q* sharpens with increasing frequency. It is a perfect example of a constant-bandwidth filter.

29.11.15 Q, EQ, and Music

The near insistence on resonant-type filters being constant in Q when varied in frequency is not through an industry wide collective lack of imagination or desire to keep things tidy. It stems from psychoacoustics, from the way humans react to audible stimuli, and also from the way nature deals with things acoustic.

If something is acoustically resonant, it will need a similar electrical resonance response shape to compensate for, extract, or imitate it in the console. Acoustics are defined by exact analogs of the first- and second-order filters and the time-domain effects that we've been delving into here in EQ.

Differing wall coverings have absorption coefficients paralleled very closely with shelving-type EQ curves. Apertures and partly enclosed spaces big and small act like second-order resonances identical to electrically resonant circuits. The physical size of a room determines the lowest frequency it can support just as a high-pass filter would. Initial and other major room reflections effect precisely the same changes on audio as deliberately introduced electronic delays; the frequency-dependent propagation characteristics of air are emulatable with slope filters.

29.11.16 Bandpass Filter Development

Methods of filtering come thick and fast once the basics are established. The development of a popular bandpass filter arrangement is shown in Fig. 29-61. It starts as two variable passive single-order filters of a common crossover frequency point, ganged so that they track. Reconfigured slightly, Fig. 29-61B, to minimize interaction, they are shown with their drive and sense amplifiers. Wrapping the two networks around an inverting amp isolates them completely from each other, improving the filter shape. The bandpass *Q* is rather low, well under one, leaving it rather limited in scope for practical applications. Positive feedback from the amplifier output back to the noninverting input sharpens the *Q*.

Yes, it does look rather like a Wein Bridge oscillator. Attempting to get the *Q* too high proves the point unquestionably!

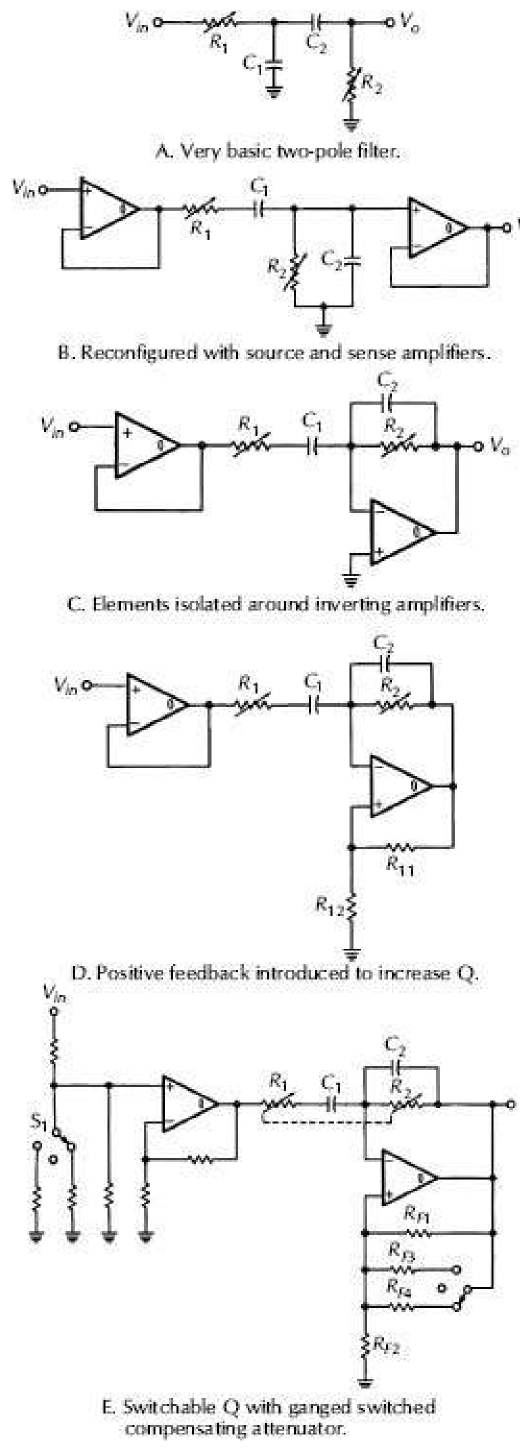


Figure 29-61. Bandpass filter development.

29.11.17 Listening to Q

This raises the problems of excessive *Q*s. Fortunately, extremely high *Q*s (greater than ten) are unnecessary or unusable for EQ purposes. The higher the *Q* becomes, the less actual spectral content of the signal it modifies, so despite the fact that its peak gain or attenuation is the same as a lower Q filter, it seems to do subjectively less. Judicious care is required in using very high Q filters to enhance or trim exactly what is required. Tuning them in is difficult; accidental overkill is easy.

There comes a breakpoint with increasing *Q* where you are not so much listening to the effect of the filter as to the filter itself. Resonant-tuned circuits are essentially ac electrical storage mechanisms, where energy inside the circuit shuffles backward and forward between the two reactive elements until the circuit losses waste it away. The greater the *Q* (and by definition the lower the included losses), the more pronounced this signal storage is.

Think of a high-Q circuit as a bell, which is just an acoustic version of the same thing. If the bell gets booted either physically or by being excited by audible frequencies at its tuned pitch, it will ring until its natural decay. It's the same with a filter. A transient will set it ringing with a decay time related to the filter *Q*. Music containing energy at the filter frequency will set it off just as well; a listener will hear the filter ringing long after the original transient or stimulus has stopped. Despite being good for a laugh, extremely high *Q*s and the resultant pings trailing off into the sunset are of no value whatsoever in a practical EQ. A transient hitting such a filter fires off a virtually identical series of decaying sine waves at the frequency of the filter.

Square waves sent through audio paths are good for kicking resonant ringing off at almost any frequency. It's a convenient means of unearthing inadvertent response bumps, phase problems, and instabilities. The breakpoint—where filter ringing is as audible as signal—is quite low, a *Q* of between five and ten depending on the nature of the program material.

29.11.18 Push or Retard?

It is not too difficult now to appreciate that resonant circuits and oscillators are very close cousins—often indistinguishable, except for maybe an odd component value here and there. There are two fundamental approaches to achieving a resonant bandpass characteristic using active-filter techniques.

The first is to start off with a tame, poorly performing, passive network and then introduce positive feedback to make it predictably (we hope) unstable. The feedback exaggerates the filter character and increases the *Q* to the desired extent. A perfect example of this is the Wein Bridge development of Fig. 29-61. The major disadvantage of such methods is that the *Q* is disproportionately critical with respect to the feedback adjustments, especially if tight *Q*s are attempted.

The second approach is to start off with an oscillator and then retard it until it's tame enough. This is the basis of the state variable, the biquad, and similar related integrator-loop-type active filters.

29.11.19 The Two-Integrator Loop

This, for better or worse, and a variety of reasons not the least of which is low component-count, is by far and away the most popular filter topology used in parametric equalizers. Three inverting amplifiers connected in a loop, as shown in Fig. 29-62, seem a perfectly worthless circuit and, as such, it is. It's there to demonstrate (assuming perfect op-amps) that it is a perfectly stable arrangement. Each stage inverts (180° phase shift), so the first amplifier section receives a perfectly out-of-phase (invert, revert, invert) feedback, canceling any tendency within the loop to drift or wobble. Removing 180° phase shift would result in perfect in-phase positive feedback; the result is an oscillator of unknown frequency determined predominantly by the combined propagation times of the amplifiers.

Arranging for the 180° to be lost only at one specific frequency results in the circuit being rendered unstable at just that one frequency. In other words, it oscillates controllably. Creating the 180° phase loss is left to two of the inverting amps being made into integrators, Fig. 29-62B, so called because they behave as an electrical analog of the mathematical function of integration.

The integrator you may recognize from a single-order filter variation in Fig. 29-59. It's not so much the amplitude response that's useful here as the phase response, which at a given frequency (dictated by the R and C values) reaches −90° with respect to its input. Two successive ganged-value integrators create a 180° shift.

Retarding the loop to stop it from oscillating can be achieved in a variety of ways:

1. Trimming the gain of the remaining inverter. This is unduly critical like the Wein Bridge for *Q* determination.
2. Doping one of the integrator capacitors with a resistor, Fig. 29-62C. This in essence is the biquad filter (after biquadratic, its mathematical determination). The *Q* is largely dependent on the ratio of the capacitive reactance to the parallel resistance; consequently, it varies proportionally with frequency. For fixed-

frequency applications the biquad is easy, docile, and predictable.

3. Phased negative feedback. This is not true negative feedback but taken from the output of the first integrator (90° phase shift). It provides an easily managed *Q* variation, is constant, and is independent of filter frequency, Fig. 29-62C. Forming the basis of the state-variable filter, this has turned out to be 'the active filter most likely to succeed", if the majority of current commercial analog console designs are to be believed.

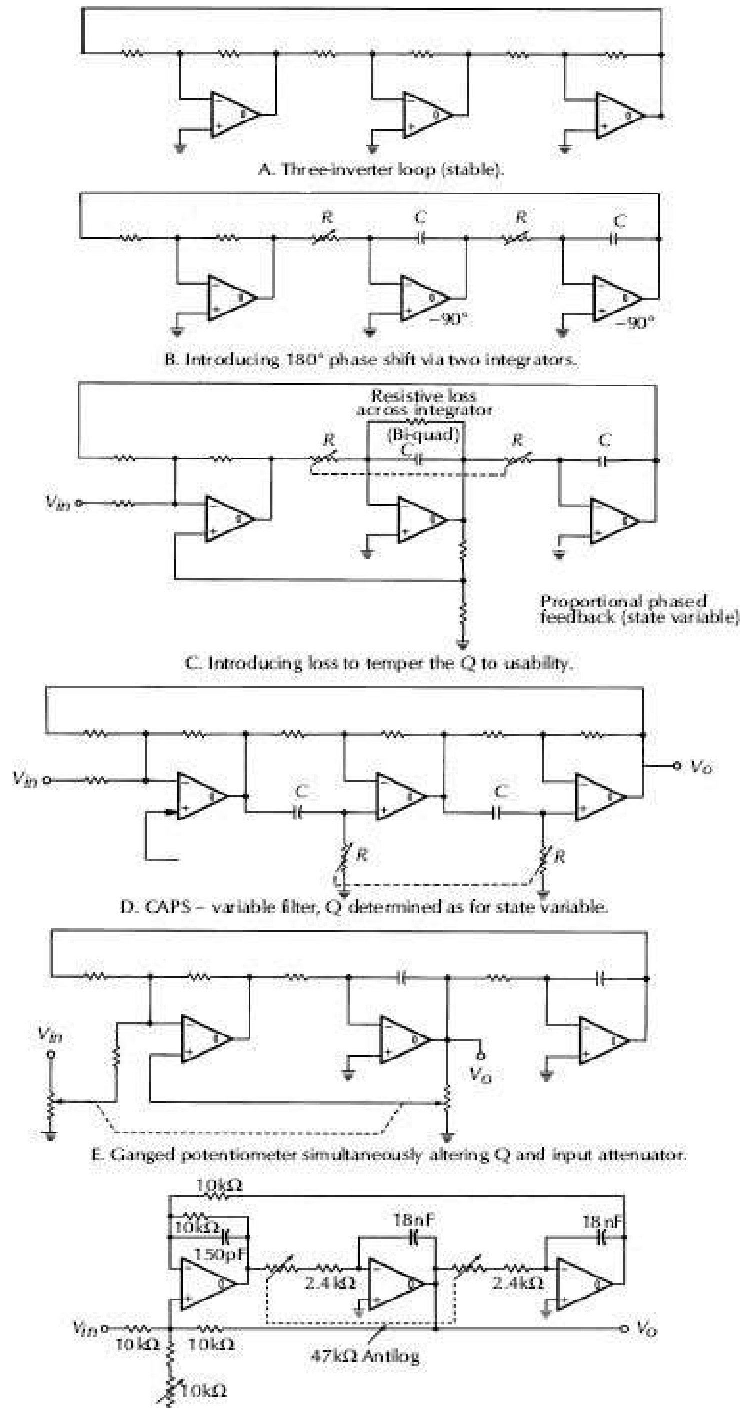
Loop filters, such as described in Fig. 29-62 have a number of inherent problems that are usually glossed over for the sake of the operational simplicity and elegance of the design.

29.11.20 Stability and Noise Characteristics

Each amplifier within the loop has a finite time delay, which together add up to significant phase shifts within the open loop bandwidths of the amplifiers. Some simply add to the delay imparted by the integrators, but the total time discontinuity around the summing amp can promote instability in the multi-megahertz region. If taming one op-amp was a trial, well, here are three. Compensation for this around the summing amplifier can introduce further phase shifts, upsetting the filter performance at high frequencies.

Problems are due to the nature of the integrator arrangement itself. They come to light at the extremes of the feedback capacitive reactance, i.e., at very low and very high frequencies where, respectively, the reactances are virtually open circuit and short circuit.

For typical audio frequency EQ the integration capacitor value can be quite sizable, up to 1 μ F, presenting two further aggravations:



A. Three-inverter loop (stable).

B. Introducing 180° phase shift via two integrators.

C. Introducing loss to temper the Q to usability.

D. CAPS – variable filter, Q determined as for state variable.

E. Ganged potentiometer simultaneously altering Q and input attenuator.

F. State-variable filter with single pot Q adjust.

Figure 29-62. Loop filters.

1. Current limiting. Is the current output capability of the op-amps sufficient to charge such a size capacitor instantaneously? If not, this will result in low maxima of signal frequency and signal level before op-amp slew-rate limitation and current-limited clipping sets in. The amplifier just might not be able to deliver enough current quickly enough. This is common.
2. Finite device output impedance. There will almost certainly be another foible related to the open loop output impedance of the op-amp; this corresponds to a resistor in series with the device output that forms a time constant and a filter with the integrator capacitor, in addition to the intended one. Another time constant means more time delay in the loop, causing a seriously degraded (maybe already critical) stability phase margin. At best it adds a zero to the integrator, reducing the integrator's effectiveness at high frequencies.

Integrators ask a lot of device outputs; not only do they have to cope with a vicious reactive load (with which many op-amps are ill equipped to cope) but they also have to drive other circuitry, such as the next stage. A mad drive to bring circuit impedances down for noise considerations can soon outstrip even the best op-amp's capabilities.

As tame as it may superficially seem, the state variable is not an unconditionally or reliably stable arrangement, with dynamic problems and device limitations potentially degrading its sonic performance. It is an amazement that these filters work as well as they do in many commercial designs. For fun, impulsive and toneburst tests on some of these show the supposed integrator zero-impedance nodes jiving, sometimes substantially. This is a sure indication that the op-amps are struggling with the demands, and the filter not working as intended.

With the exception of inevitable loop effects (usually time related), most of the undesirable things about the state variable can be eliminated or mitigated by replacing the integrators with constant amplitude, phase-shift elements, Fig. 29-62D. This results in what could best be known as a CAPS-variable filter. Here, all the constituent elements are basically stable, and there are provisions for independent device compensation. There is no undefined gain for any of the spectrum, and amplifiers aren't subject to unrealistic loads. This seems to be a far healthier format to start making filters around.

There is another way of looking at the state variable/CAPS-variable filters that will suddenly resolve the previous discussions on gyrators, L and C filters, series-tuned circuits, and so on with the seemingly at-odds approach of active filters.

Resonance depends on the reaction of the two reactances of opposite sense, 180° apart in phase effect. Rather than achieve this in a differential manner, one element +90° with the other −90° at a given frequency, these active filters achieve the total difference by summing same-sense phase shifts (−90° + −90°), i.e., still 180° apart. Two reactive networks are still involved; it is still a second-order effect. At the end of the day, the principal difference is that such loop-type active filters have their median resonance phase displaced by 90° from their input as a result of both reactive effects being in the same sense, as opposed to the nil phase shift at resonance of a real LC network.

29.11.21 Q and Filter Gain

Pretty much every resonant-type active filter has the characteristic of its gain at resonance being at least related and often directly proportional numerically to the *Q* of the filter. This means a filter with a *Q* of 10 usually has a voltage gain of 10 (20dB) at resonance. Naturally, this does not make the building of practical equalizers any easier. Nothing much does. Even specifying a maximum *Q* of 5 (14dB gain) only helps by losing 6dB of boost with respect to a *Q* of 10.

That represents a very sizable chunk of system head room stolen at the filter frequency, which also makes the sum-and-difference matrixing necessary to provide the usual boost-and-cut facilities difficult to configure. The obvious solution is to attenuate the signal going into the filter by the same amount as the gain and *Q* expected of the filter. Arranging a continuously variable Q control that also attenuates the signal source appropriately is not a conspicuously simple task, at least with most filters. Perhaps the most straightforward example is shown in Fig. 29-62C, a state-variable-type filter with an attenuator in the retard network altering the *Q* ganged with an attenuator ahead of the input/summing amplifier. Within reasonable limits this holds the resonant peak output constant over a considerably useful *Q* range. A much neater and more commonly applied solution is shown in Fig. 29-62F: a single potentiometer at the noninverting input of the summing amp that would serve both purposes—filter *Q* and input level—complementarily and simply.

Most other filters are not so obliging in terms of continuously variable Q. Switching between a few values of *Q* while substituting appropriate input attenuation is quite often a practical and operationally acceptable solution, applicable to nearly any filtering technique. Fig. 29-62E illustrates a further development of the Wein Bridge arrangement using this method to provide three alternative *Q*s. The attenuator values are necessarily high in impedance to prevent excessive loading of the source, a factor that in some practical EQ circumstances can be important.

29.11.22 High-Pass Filters

Two basic single-order high-pass filters are shown in Fig. 29-63. The keys, for the purposes of high-pass filtering, are the reduction of inductive reactance to ground with reducing frequency in Fig. 29-63F and the increasing of capacitative reactance with reducing frequency in Fig. 29-63G.

How about combining the two and omitting the resistors as in Fig. 29-63A? As expected, the combining of the two opposing reactances causes an ultimate roll-off twice as fast as for the single orders; however, they have also resulted in a resonance peak at the point of equal reactance. Resonance *Q* is the ratio of elemental reactance to resistance; deliberately introducing loss in the circuit in the form of a termination resistor tames the resonance to leave a nice, flat, in-band response, Fig. 29-63B.

Substituting a basic gyrator or simulated inductance for the real one, Fig. 29-63C naturally works just as well and even better than expected. The filter output can be taken straight from the gyrator amplifier output, eliminating the need to use another amplifier as an output buffer. Further, we can automatically introduce the required amount of loss into the inductor by increasing the value of the bootstrap resistor and get the resonance damping right. (Refer to the discussion of gyrators in section 29.11.9.)

Further yet, we can easily change the turnover frequency of the filter by varying what was the tuning resistor. In doing this, of course, the elemental reactance-to-loss ratio will change, causing damping factor (and so the *Q*) to change with it. The frequency change and required damping change are directly related and in the same sense and may be simultaneously altered with a ganged control—even, if we do our sums right, with the two ganged tracks having the same value!

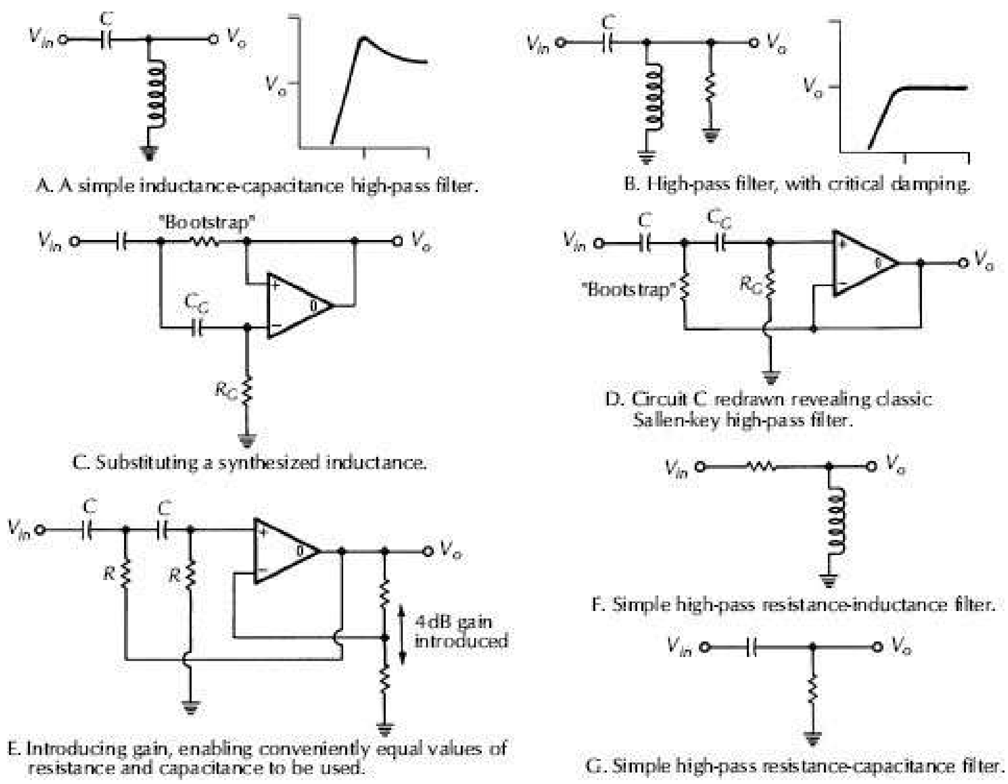


Figure 29-63. High-pass filter development.

A slight redraw of Fig. 29-63C gives Fig. 29-63D, a more conventional portrayal of the classic Sallen-Key high-pass filter arrangement. As the Sallen-Key filter evolves, it turns out that an equal value filter (where the two capacitors are equal and the two resistors are equal) results in a less than adequate response shape. An expedient method of tailoring and smartening up response to become Butterworth-like (working on the assumption that a couple more resistors are cheaper than a special two-value ganged potentiometer) is to alter the damping by introducing gain into the gyrator buffer amplifier (providing also a healthier mode of operation for the amplifier—followers are bad news), see Fig. 29-63E. A side effect of this technique of damping adjustment (which, incidentally, is independent of filter frequency) is that an input-output in-band gain is introduced. The 4 dB gain introduced necessary to render the filter frequency response maximally flat could be included in overall system gain, or alternatively a compensating attenuator could be instituted ahead of it. This could as well be arranged to be a fixed-frequency, band-end, single-order, high-pass filter to accelerate the roll-off slope out of band; a further alternative is to make the filter input out of two capacitors such that the input signal is attenuated by the needed amount yet the combined capacitance value is the correct value for the filter—this can be a bit of a nightmare to drive adequately, though. For many applications the free 4dB or so usually isn't a problem—it can simply be assimilated as part of the system-level architecture.

When the 4dB thing is a nuisance, or where an inverting filter stage is either convenient or necessary, the multifeedback configuration works well; indeed, lacking the problems of a near-follower as in the case of the Sallen-Key it uses the op-amp well. At high values of *Q* or extremes of frequency, some component values can get far from the ordinary midimpedance values seen elsewhere in the EQs and filters described here, and one should be aware of possible noise or op-amp current-drive limitation issues as a consequence. Unlike the Sallen-Key described, it is not readily possible to continuously vary the turnover frequency, and it uses three capacitors as frequency-determining components rather than two. Otherwise, for fixed frequency filters this is a very friendly topology.

29.11.23 Second or Third or More Order?

Without delving too deeply into psychoacoustics, the ear notices easily third or more order filters being introduced for much the same reasons as a high-Q bandpass filter is obvious. There are severe modifications to the transient response of the signal path and ringing-type time-related components are introduced into the signal spectrum.

An application where this effect is not overly objectionable is where the filters are defining bandwidth at high and low audible extremes. Within the audible band though, the ear is quite merciless toward such artifacts.

The transient response modification and timedomain effects are not the end of the story; the relationships between instrument fundamentals and their harmonics in the turnover area of the filter are likely to be interpreted as unnatural, especially should the fundamental be attenuated with respect to the harmonics.

Second-order filters, assuming moderate filter Qs associated with Bessel or Butterworth characteristics, score well in both respects.

There is less transient response disturbance and less tonal characteristic modification. There are few who would dispute that they sound more natural and musical than tighter filters. A small wrinkle is to leave a small controlled amount of underdamped bump in the filter frequency response. This has two consequences: one is the slightly more rapid out-of-band roll-off, but the other, a subjective effect, is that the extra program energy introduced by the hump serves to help offset the loss of energy below the turnover frequency. The perceived effect on introducing the filter is more of a slight change in sound rather than a direct drop in low-frequency response and strikes a better subjective compromise than techno-striving for the ultimately flat, perfectly measuring filter.

29.11.24 Equalization Control

Achieving bare response shapes of whatever nature—high-pass, low-pass, bell-shaped bandpass, or notch—does not really constitute a usable EQ system. The shape, even if variable in frequency and bandwidth, is either there or not, in or out, no subtleties or shades; some means of achieving control over the strength of effect is vital to the cause. By far the most common (but certainly not the only) control requirement and one easily understood by operators is lift and cut, where the frequency areas relevant to the various filters are required to be boosted or attenuated by any variable amount within given limits. Determining these limits alone is good for an argument or two, dependent on such disparate considerations as system head room, operator maturity, and, obviously, application. An EQ created specifically for wild effects is not a stable device. An adjustment of 20dB is not unknown (and not, unfortunately, unheard); a 6dB adjustment, in contrast, is often far more than enough particularly for spoken voice. A general median accepted by most manufacturers is to provide between ±12 and ±15dB level adjustment on channel-type EQs.

29.11.25 The Baxandall

Hi-fi-type tone controls needed similar basic operational high-frequency and low-frequency boost-and-cut facilities, and a design for this dating from the 1950s by Peter Baxandall has since been an industry standard in assorted and updated forms. A development of the Baxandall idea is represented in Fig. 29-64 based around today's more familiar op-amp technology rather than discrete transistors or tubes. Fig. 29-64A shows a virtual-earth-type inverting amplifier with the gain (being equal to the ratio of the feedback resistor R_F to the series resistor R_S) continuously variable from near-infinite loss (min) to near-infinite gain (max) with unity in the middle. If a fixed-gain-determining leg is introduced and the variable leg is made frequency conscious, as shown in Fig. 29-64B (in this instance by crude single-order high-pass filters—the series capacitors), the gain swing only occurs within the passband of those filters. The through gain for the rest of the spectrum is determined by the two fixed resistors. If this fixed chain is replaced by a second frequency-conscious network that does not significantly overlap the original one in bandwidth, the two chains independently modify their frequency areas, Fig. 29-64C. The fixed chain is only necessary where the gain is otherwise unpredictably defined by a frequency-conscious network.

The belt-and-braces low-pass arrangement (for low-frequency boost and cut) of Fig. 29-64C can be rationalized into the more elegant circuit of Fig. 29-64D. This circuit more closely resembles the definitive Baxandall circuit. Rather than isolating the low-frequency boost-and-cut chain with increasing inductive reactance, the control is buffered away with relatively small resistances and bypassed to high frequencies by capacitance. The control takes progressively greater effect at lower frequencies as the rising capacitative reactance reduces the effective bypass. A further refinement is a pair of stopper resistors, small in value, that define the maximum boost and cut of the entire network.

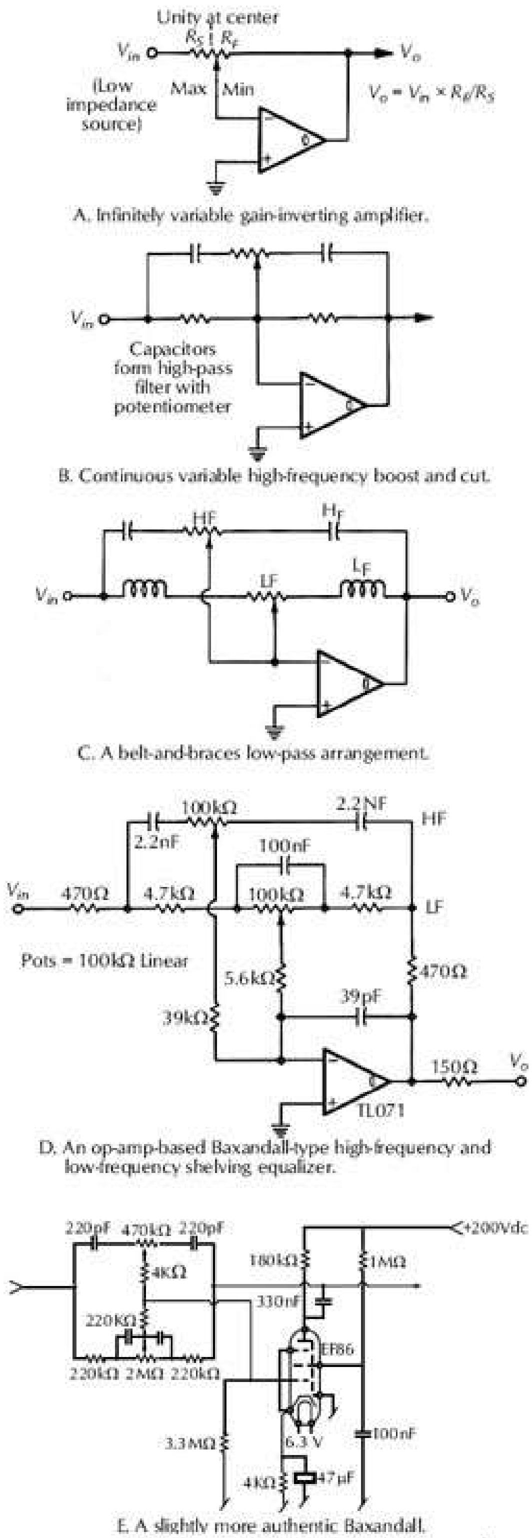


Figure 29-64. Development of Baxandall-style equalizer.

Naturally, a more complex EQ can be configured around the same arrangement. A midfrequency bell curve is easily introduced by any of the means in Fig. 29-65, giving a good hint on how to avoid using a real tuned circuit using inductors.

A variable signal either positive or negative in phase to the source V_{in} can be picked off from a pot straight across the existing high-frequency and low-frequency chains, taken to an active filter arrangement to derive the needed amplitude response shape. The signal is then returned to the loop at either the virtual-earth point (to which the high-frequency and low-frequency chains are tied) or to the noninverting reference input, Fig. 29-65D, depending on whether the absolute phase of the filter is positive or negative. Industry favorites seem to be this approach using either a Wein Bridge bandpass or a state-variable integrator-loop type.

Any number of such active chains may be introduced, provided two great hangups don't intrude excessively:

- Hangup 1 is the interaction between frequency groups. Hanging on two control chains that operate at the same frequency either adjustably or through overlap can at best be deceiving or at worst self-defeating. In the Baxandall (as with most other arrangements), if maximum gain (say, 15dB) is attained at a given frequency by one control, a second similarly tuned chain, cranked for maximum, will not give the expected additional 15 dB gain. The overall loop is already operating close to the maximum gain defined by the stopper resistors. A notable measured result is for the maximum boost-and-cut capability of a sweep-mid bell curve to be restricted at the extent of its range where it overlaps across the shelving high-frequency and low-frequency curves.

  A rough rule born from hard experience of squeezing the most EQ from the least electronics is to not allow overlap incursion beyond the point where either curve has ±6dB EQ effect individually. Overlapping is best achieved from the comfort of another EQ stage, although that too invokes other compromises. Paralleling sections (addition) around one stage is efficient but interactive; sectional isolation demands the more expensive multiplicative technique of sequential stages. Another consideration is that many operators who tune by ear like and happily accept additive sections, whilst those more analytically inclined prefer the multiplicative.

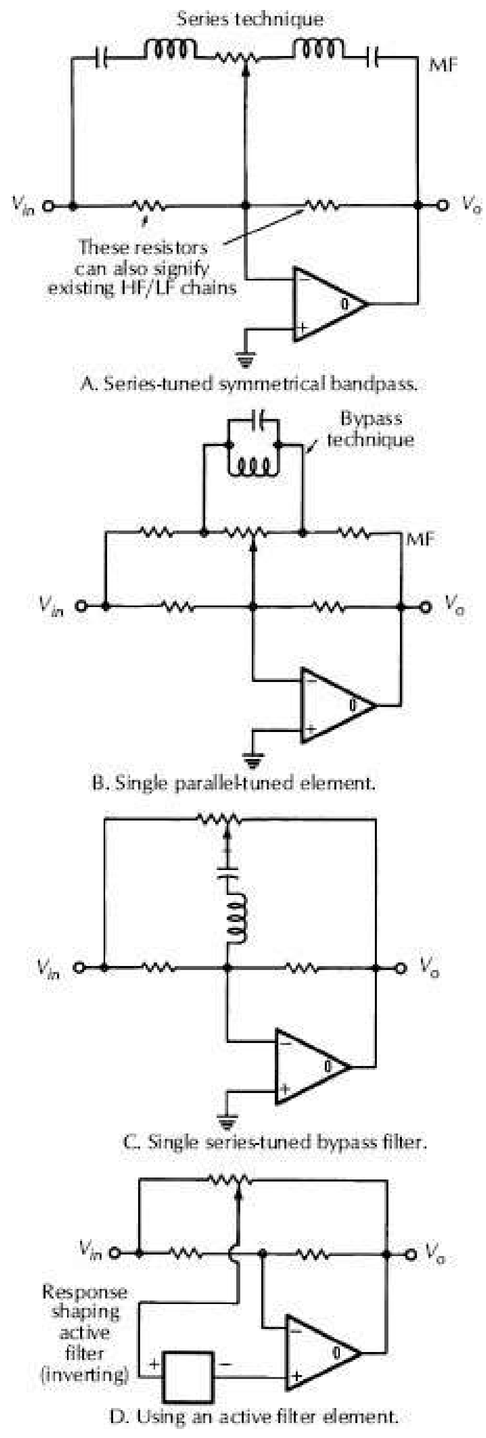


Figure 29-65. Resonant frequency selective elements in the Baxandall equalizer.

- Hangup 2 is noise. The basic Baxandall, using purely passive frequency-determining components, is a fairly quiet arrangement.

With controls at flat, it is theoretically only 6dB noisier than the unity-gain noise of the amplifier plus additional thermal noise due to network resistances—all in the −100dBu region. The noise character varies with the controls, as would be expected of an amplifier whose gain is directly manipulated at the frequencies in question—high-frequency boost, more high-frequency noise, and so on, but these are low in level regardless.

As soon as active filtering is involved, more noise is unavoidably introduced, often highly colored and consequently much more noticeable. What is worse is that it's present all the time irrespective of control positions. Even with its appropriate control at neutral center, it is quite usual to hear a midsweep swoosh in the noise changing with filter frequency. This is, along with the strange spectral character of the noise emerging from some filters, notably, the integrator-loop variety, a result of unoptimized impedances and dubious stability almost inherent to their design.

29.11.26 Swinging Output Control

The source impedance versus feedback impedance ratiometric approach of the Baxandall is not the only way of achieving symmetrical boost-and-cut, as stunning as its simplicity and elegance may be. A method of enclosing the controls within the feedback leg of a noninverting amplifier is developed in Fig. 29-66. This has the advantage of leaving the noninverting input of the op-amp free, obviating the need for a preceding low-impedance source or buffer amplifier. Roundabout to this swing is the necessity of a buffer amplifier or quite high destination load impedance since the output is variable in impedance and included within the feedback loop of the op-amp. Serious control law modification, potential phase margin erosion with consequent instability, and certain head room loss are among the penalties for careless termination.

Unity gain in Fig. 29-66A is achieved when the attenuation in the feedback chain equals the output attenuation; the feedback attenuator causes the op-amp to have as much voltage gain as the output attenuator losses. Replacing the two bottom legs of the attenuators with a swinging potentiometer, Fig. 29-66B, provides a boost-and-cut facility; when the pot is swung toward min, the feedback leg is effectively lengthened to ground, and the amplifier gain is reduced somewhat. Meanwhile, the output attenuator is shortened considerably, reducing the output accordingly. At max the reverse occurs. The feedback leg is shortened, increasing the loop gain of the op-amp while the output attenuator is lengthened, losing less of the available output. A small stopper resistor defines the overall gain swing about unity, which would otherwise range from zero to earsplitting, respectively.

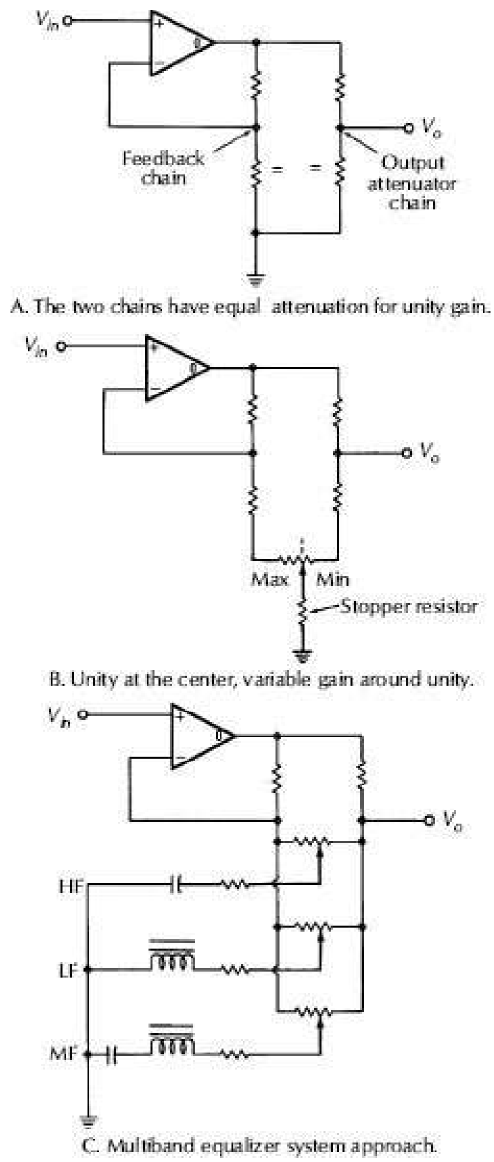


Figure 29-66. Swinging output equalizer.

Introducing reactances and complex impedances into the potentiometer ground leg (or legs as in Fig. 29-66C) results again in boost-and-cut control over the frequency bands in which the reactances are lowest, i.e., high frequency for capacitors, low frequency for inductors (real or fake), and so on. This arrangement, which is in a few professional systems and in some Japanese hi-fi, has only one major drawback other than the previously mentioned output-loading considerations. In order to achieve reasonable control dB-per-rotation linearity, the two attenuators (feedback and output) need to be of about 3 dB loss each with the control at center. This implies that the obtainable output voltage is 3 dB below the output swing capability of the op-amp, landing a headroom deficit of that amount in the equalizer stage—probably where it is most needed.

29.11.27 Swinging Input Control

Avoiding the head room headache but utilizing a rather similar technique, the swinging-inputs gain block of Fig. 29-67 is very promising. Here, the feedback attenuator remains unchanged, but the output attenuator is shifted around to the noninverting input of the op-amp. At minimum, the input attenuation is quite vicious while the feedback leg is long, making the op-amp deliver only a small amount of gain. When the attenuation characteristics are reversed for maximum, the op-amp works at a high loop gain, while the input is only slightly attenuated. Unity is achieved at control center where the input attenuation equals the make-up gain of the amplifier.

There is a fascinating tradeoff between noise mechanisms in this circuit arrangement. Assuming a maximum of three controls (for fairly basic high-frequency, low-frequency, and midsweep curves) before interaction becomes a major hassle, the amplifier can have between 10dB and 20dB of fairly frequency-conscious background gain, (i.e., with all controls flat) rendering it at first sight significantly noisier than a Baxandall. However, the impedances around the amplifier are around a decade lower. This considerably reduces thermal noise generation due to resistive elements and op-amp internal mechanisms.

In addition, the noise generated by the active frequency-determining filters is, with the controls neutral, injected equally into the inverting and noninverting inputs of the op-amp. Differential amplifiers being what they are, common-mode signals (such as this equally injected filter noise) get canceled out and do not appear at the output.

Interaction can still intrude, and care is required to prevent excessive frequency band overlap. Center-tapped pots (the tap grounded) eliminate many interactive effects but at the cost of increased invariable background gain (noise) and peculiar, almost intractable, boost-and-cut gain variation linearity versus control rotation.

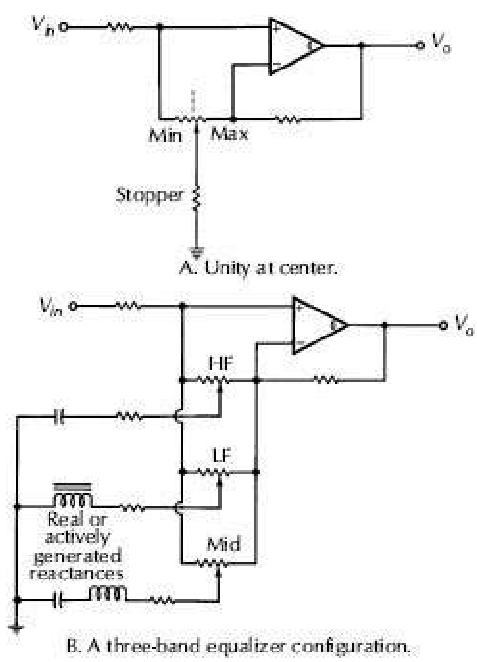


Figure 29-67. Swinging output equalizer.

29.11.28 A Practical EQ

A three-section parametric EQ with additional versatile shelving-type high- and low-frequency controls is detailed in Fig. 29-68. It is designed to be easily shortened to high-frequency, low-frequency, plus a single midband parametric section, for applications that don't demand the full complement of facilities. Each individual section is switchable in or out to allow preset controls. Simple in-and-out comparisons with tie-down resistors maintain the dc conditions of the unused filters to minimize switch clicks. Even a brief look at the circuit reveals a major benefit. The signal path through the EQ is merely via three op-amps, IC_2 is an input differential amplifier, and IC_3 does duty as the output line amp. In the shortened version this path is reduced to only two op-amps, IC_1 and IC_3, which serve also as a swinging-input EQ gain block. IC_2 and its associated circuitry are unused in this simplified version.

The unusual components around the differential input stage provide unity differential in unbalanced out levels while providing an identical impedance (with respect to ground) on each of the two input legs. Naturally, the more precise the component values, the better the common-mode rejection is likely to be.

29.11.29 The First EQ Stage

IC_2 in Fig. 29-68 is the first swinging-input stage. It has two nonfrequency overlapping filters hanging off it, one section covering 25 Hz to 500Hz, the other covering 1Hz to 20kHz. Each filter network creates a complex impedance form against frequency that looks like a series LC-tuned circuit to ground. This fake-tuned circuit (formed from two constant-amplitude phase-shift networks in a loop, named the CAPS-variable filter) reach parameters ordinary filters cannot reach.

The center frequency is continuously and smoothly variable over its range using reverse-log potentiometers; *Q* remaining consistent over the entire swing. The *Q* itself is continuously variable between 0.75 and 5 (very broad to fairly sharp, representing bandwidths of 1.5 to 0.2 octaves, respectively). Positive feedback inside the loop, which defines the *Q*, is balanced against negative feedback, which controls minimum filter impedance and, correspondingly, amplitude. Interestingly enough, this circuit relies on the input impedance of the swinging-input stage as part of the negative feedback attenuator. Fortunately, this impedance is reasonably constant irrespective of boost-and-cut control positioning.

In the absence of complementary square-law/reverse square-law dual-gang potentiometers ideally required for the purpose, readily available log/antilog dual-gang pots, retarded a bit to a reasonable approximation, control the positive/negative feedback balance. As a result of this compromise, the filter crest amplitude (maximum effect) varies within ±1dB as the Q control is swept; in comparison to the dramatic sonic difference from such a *Q* variation, this tends to insignificance. The result of all this, at the output of IC_2, is a pair of resonant-type curves of continuously variable place, height, depth, and width.

29.11.30 Second EQ/Line Amp

A reasonably hefty pair of transistors is hung on the end of IC_3 to provide a respectable line-drive capability, in addition to the use of the amplifier as a swinging-input EQ section. There is enough open-loop gain in the combination of the op-amp and transistors (over a much greater bandwidth than mere audio) to cope with 15 dB of EQ boost and output-stage nonlinearities.

Differing from the last EQ stage, this one only has a single midfrequency bell-curve creator, operating over a range of 300Hz to 3kHz, together with high- and low-frequency range impedance generators.

29.11.31 Low-Frequency Control

Gyrating inductance to create a conventional low-frequency shelving response (variable in turnover frequency by a 220kΩ antilog pot) is achieved around IC11. A fairly large (2.2 µF) series capacitor forming a resonance is switchable in and out. The value of the capacitor is carefully calculated to work with the circuit impedances to provide an extreme low-frequency response that falls back to unity gain below the resultant resonant frequency without compromising the higher frequency edge of the curve. The *Q* of this arrangement reduces proportionally to increasing frequency. Typical resultant response curves, Fig. 29-69, show just what all this means, demonstrating an extraordinarily useful bottom-end control.

29.11.32 High-Frequency Control

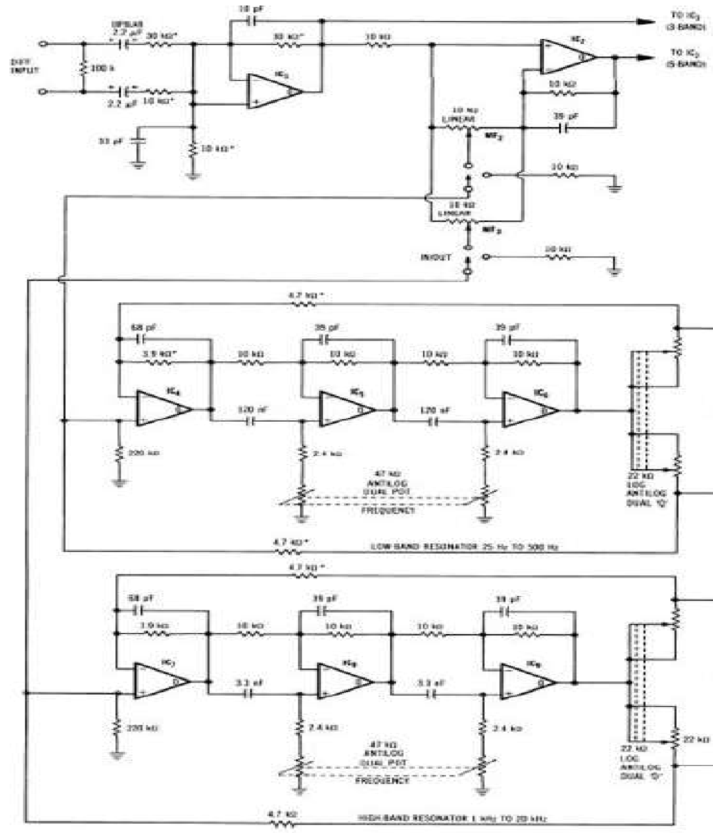
Unusual is one way to describe the high-frequency impedance generator and its EQ effect. It is essentially a supercapacitor, or capacitative capacitor. In other words, it's a circuit that, when in conjunction with a resistor, causes a second-order response as would normally be expected of an inductor and capacitor combination—a slope of 12dB/octave as opposed to a single-order effect of 6dB/octave. Fig. 29-70 shows what it does as an EQ element.

The response is hinged about 1kHz. The control varies the frequency (between 5kHz and 20kHz) at which the gain reaches maximum (or minimum if the boost-and-cut control is cut). The slope between 1 kHz and the chosen maximum frequency is virtually a straight line representing a nearly constant dB/octave characteristic, with a nearly flat-top shelving characteristic.

In electronic terms, this is achieved by progressively degenerating the supercapacitor until it's no longer super, i.e., it eventually ends up looking like a simple, single capacitor.

29.12 Dynamics

Dynamics processing is as common in today's signal paths as equalization. Many commercial mixing consoles carry dynamics as standard on a per-channel basis. Formerly, the occasional necessity of hanging in external dynamics processing was the major justification for channel insert (or patch) points. These patch points were typically either before or after (pre- or post-) equalizer in the signal chain as shown in Fig. 29-71.


DIFF INPUT
100 k
2.2 μF
30 kΩ*
10 pF
10 kΩ
33 pF
TO IC_3 (3-BAND)
TO IC_3 (5-BAND)
10 kΩ LINEAR
MF_2
39 pF
MF_1
IN/OUT
4.7 kΩ*
68 pF
39 pF
3.9 kΩ*
120 nF
220 kΩ
2.4 kΩ
47 kΩ ANTILOG DUAL POT
FREQUENCY
33 kΩ
22 kΩ LOG ANTILOG DUAL 'Q'
LOW-BAND RESONATOR 25 Hz TO 500 Hz
3.3 nF
HIGH-BAND RESONATOR 1 kHz TO 20 kHz

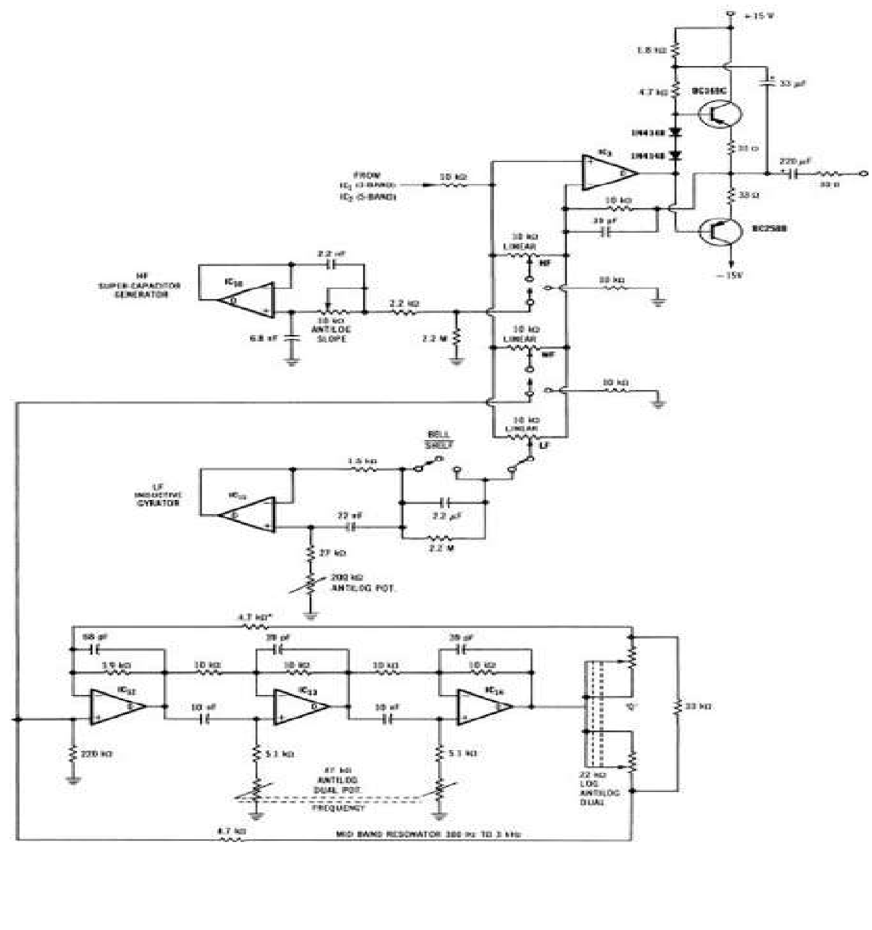

+15 V
1.8 kΩ
4.7 kΩ
33 μF
1N4148
1N4148
33 Ω
220 μF
BC258B
−15V
FROM
IC_1 (3-BAND)
IC_2 (5-BAND)
10 kΩ
10 kΩ LINEAR
HF
HF SUPER-CAPACITOR GENERATOR
2.2 nF
10 kΩ ANTILOG SLOPE
6.8 nF
2.2 kΩ
2.2 M
MF
BELL
SHELF
LF
1.5 kΩ
LF INDUCTIVE GYRATOR
22 nF
2.2 μF
2.2 M
27 kΩ
200 kΩ ANTILOG POT.
4.7 kΩ*
68 pF
39 pF
3.9 kΩ
10 kΩ
10 nF
220 kΩ
5.1 kΩ
47 kΩ ANTILOG DUAL POT
FREQUENCY
'Q'
33 kΩ
22 kΩ LOG ANTILOG DUAL
4.7 kΩ
MID-BAND RESONATOR 300 Hz TO 3 kHz

Figure 29-68. Five-band equalizer circuit diagram.

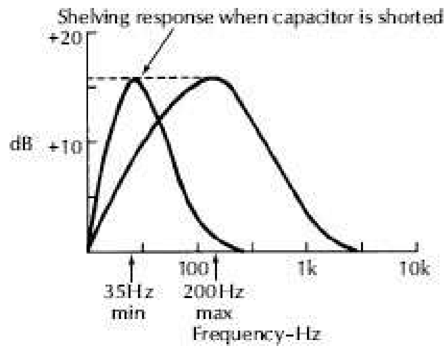


Figure 29-69. Frequency response of the low-frequency section of Fig. 29-68 (control at maximum gain).

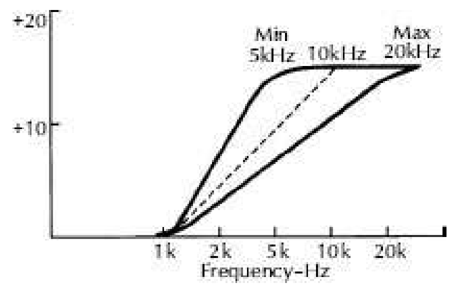


Figure 29-70. Characteristics of the high-frequency section of Fig. 29-68 (boost-cut control at maximum boost).

Preequalization actually amounts to postinput stage; the purpose of an insert point here is to control unruly input signals that may otherwise clip later in the chain—in particular within the EQ section where serious amounts of gain at some frequencies may be instituted. Characteristically this pick-off point is after any high-pass filtering, so that any low frequency rumble that may cause false triggering within the dynamics may be removed. On the other hand, the post-EQ insert point allows control over the entire channel immediately prior to the channel fader and is commonly used for automatic gain controlling.

Dynamics is automatic control of signal level by an amount determined by the characteristics of the signal itself. In a linear 1:1 circuit, what goes in comes out untouched and unhindered. For example, if we have a circuit that automatically senses the input signal and uses that measurement to control the output signal, if the input signal rises in level by 6dB, the output signal is controlled to rise only 3dB. The output signal has been compressed by a ratio of 2:1 with respect to the input signal.

There are four basic types of dynamic signal processing:

1. Limiting.
2. Gating.
3. Compression.
4. Expansion.

It is arguable that limiting is a special case of compression and that gating is similarly a special case of expansion, effectively reducing the number of basic groups to two. Although in practice the means of achieving these pairs of effects are indeed similar, true definitive compression is a long way from limiting, as gating is from expansion. The discussion about time constants etc. within the following Limiting section is directly applicable to each of the other sorts of dynamics.

Fig. 29-72 shows a now customary style of input-output signal level plot of a compound (more than one dynamics type active) dynamics section, with typical slopes for each of limiting, compression, and expansion. (Measured plots of actual dynamics sections can be seen also in the discussion of digital dynamics later.) This style of dynamics display is common on programmable consoles; here it can be a handy form of visualization or representation as the various types are discussed. Linear, i.e., no processing, is represented by the dotted line, showing equal output for input. Unusually (!) a section of linear remains in this compound curve, displaced upward by 15 dB; this is a normal occurrence when automatic gain reduction such as compression and/or limiting is used. In order to make up for the gain reduction above the threshold, some gain is applied to compensate and bring the output signal back up to usable levels. This is often called makeup, or buildout gain.

29.12.1 Limiting

This is the conceptually easiest and most commonly applied form of dynamics processing. Nearly any audio heard outside of a recording studio has been passed through a limiter at some point in the chain; TV audio, radio, even (and especially) CDs. This very pervasiveness has had the rather odd unforeseen effect that culturally we have gone beyond acceptance and become dependent on the sound of excessive limiting and heavy compression; even to educated ears a new CD can sound wrong or wimpy in comparison to what had been heard on the radio, where typically murderous additional processing is applied. Music retailers actually do get complaints like that. The race to be louder than loud has unfortunately spilled over into the record production arena back from radio where it had previously raged alone sparked by the AM radio loudness wars of the 1960s and 1970s; the deleterious effect will have future musical historians scratching their heads.

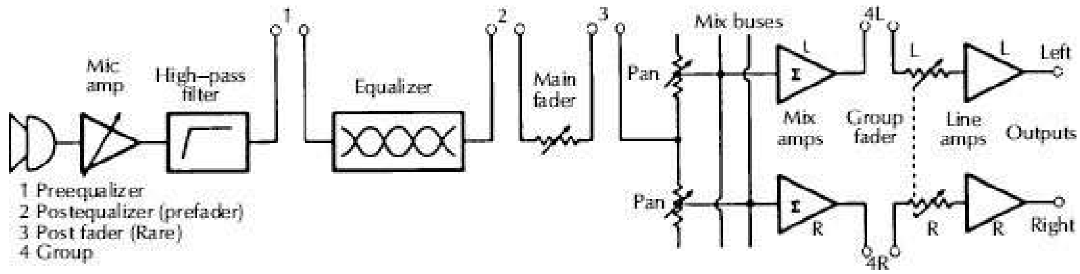


Figure 29-71. Typical insert points for dynamic processing.

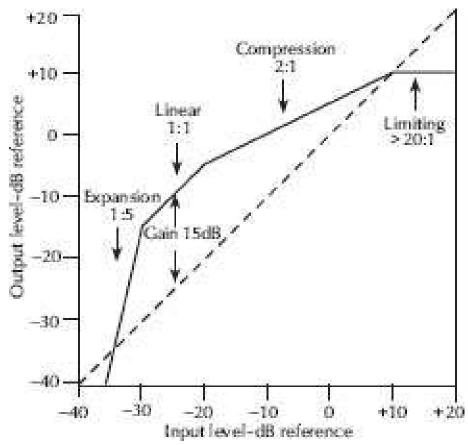


Figure 29-72. Dynamics Input/Output Plot.

Every recording and transmission medium has definite head room limitations, a maximum level beyond which the signal just plain overloads, distorts, or becomes a serious liability. AM radio overmodulation not only sounds horrid but causes interfering splatter up and down the dial, FM transmitter overdeviation causes adjacent channel interference and runs the risk of distortion in receiver discriminators; disk cutters can produce grooves that run into each other (long after they become unplayable by any normal pickup), PA loudspeakers fry, tape saturates, distorts, and screams, and best of all anything digital just plain runs out of bits and cracks up. The answer? A device that senses when enough is nearly enough and automatically reduces the source level such that a proscribed output level is not transgressed. This is a limiter.

Fig. 29-73A shows the all-time basic limiter—a pair of back-to-back diodes. These clamp the input signal from the source resistor to within their nonconductive range; beyond 700mV of either polarity one or other of the diodes conducts, sawing off any excessive signal. Brutal, but effective. The downside is gross distortion—serious waveform modification is going on, profligate audible distortion products are generated. Fig. 29-73B shows the same idea with germanium diodes. These tend to have a lower turn-on voltage (200–300mV) but a gentler knee, with the effect shown. This sounds considerably less harsh—fewer high-order distortion products are being created. In situations where ultimate signal quality is not necessary, but increased signal density (translated: loud) is required, these clipping circuits work like a charm; communication circuits often use this technique to saw the top 10dB or so peaks off speech and thus gain a nearly corresponding degree of increased apparent loudness. Broadcast airchain processors do the same but far more elaborately, the trick being to filter away and contain and control the resulting distortion products such that they become less agonizing, while retaining the high signal density clipping affords. Such is meat for a whole other saga.

The ideal limiter circuit is one that knows its input is going over the top and can reduce its gain such that the signal is left relatively undistorted but as high in level as it can be within the given constraints. Fig. 29-73C is a block diagram of such a device. The side chain circuit is a tripwire in this instance; if the amplifier output exceeds a stipulated level (just below what the destination can handle) the side chain develops a control signal that tells the input attenuator to drop the input signal sufficiently. The whole circuit operates in check—the bigger the input signal, the bigger the potential overload, the bigger the control signal, the more the attenuation. Below the tripwire—the threshold—the whole circuit behaves the same as an ordinary straight amplifier. Fig. 29-73D is about as simple as a decently performing limiter can get and was first noted in the original mid-sixties Philips cassette recorder.

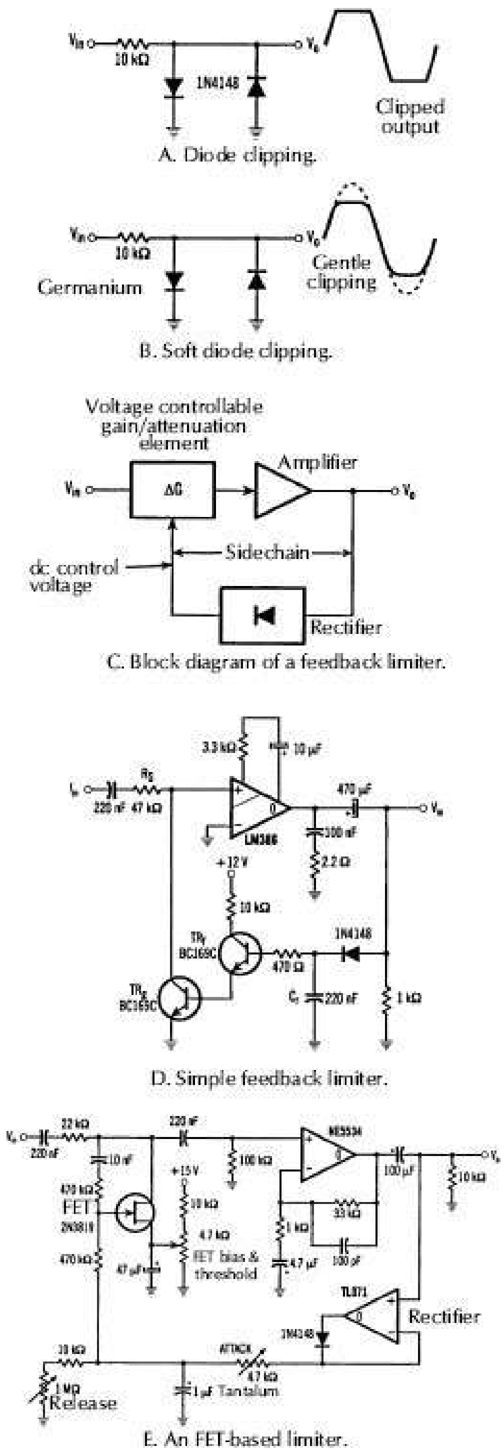


Figure 29-73. Simple limiters.

The LM386 is a small power op-amp commonly used to drive headphones or small loudspeakers but works well just as an ordinary amplifier, in this case at some 30dB gain. It is used here for its power output stage, which is hefty enough it can ignore the diode rectifier and side chain loading effects. This diode conducts when the positive-going output signal exceeds about 700mV and charges a reservoir capacitor, C_r. This is buffered by an emitter follower (TR_f) feeding the gain controlling transistor TR_g. When the voltage on the base of TR_f is sufficient to force conduction through the two base-emitter junctions, TR_g turns on, causing an increasingly low-impedance path to ground at the input to the amplifier. It forms a potentiometer, with the source resistor R_s attenuating the input signal to the level at which the rectifier and two transistors are just conducting. In this circuit that amounts to a positive-going output signal of about 2V (the added voltage drops of the rectifier and the two transistor base-emitter junctions). Simplicity has its drawbacks and in this instance they are noise and distortion. Although the distortion is in a different league from a diode clipper, transistors do not make ideal VCAs (Voltage Controlled Amplifier/Attenuator) and are somewhat nonlinear in this application. If, however, the signal across them is kept low (in this instance, −30dBu, the lower the better) it can be quite acceptable for a lot of applications. Keeping the signal low necessitates following gain to bring the signal back up again. That means amplifier noise.

There is a wide variety of possible voltage-control elements and Fig. 29-73E shows a smoother version of the transistor limiter using a JFET. The principle is much the same, only the side chain and VCA circuitry has developed somewhat. FETs have such spread characteristics that a preset is necessary to set their bias points. In normal operation the FET needs to be biased just nonconducting; that is, nonattenuating. This necessary adjustment also provides a means of varying the output level at which the limiter starts to bite, and incidentally some control over the ratio of the gain reduction. (The greater the bias, the more control voltage signal is needed to be generated before the FET turns on and starts attenuating, where it does so in a tightly controlled manner. A low bias results in a lower, or indeed no, threshold and a far gentler gain-control ratio. Carefully trading these two—a high threshold for a hard ratio or low and “smushy” threshold and gentle FET turn-on—against overall gain formed the basis of FET-based compressors such as the famed Audio and Design F760 and the UREI 1176LN.)

FETs have very high gate impedances, precluding the need for a follower. The control voltage is summed at the device gate with a sample of the input signal. Automodulation is an effect of FETs where the source-drain resistance (the resistance we’re depending on as part of the input attenuator) varies with signal voltage across it. This is attacked in two ways: first by keeping the signal across the FET low, as in the transistor limiter, and second by supplying the gate with some anti-wobble signal that does a fairly good job of forcing the source-drain path to wobble against and largely cancel the automodulation effect.

29.12.1.1 Side-Chain Time Constants

Between the rectifier and the FET in Fig. 29-73E is a simple resistor-capacitor network that determines how the side chain works and its effect on the automatic gain reduction. This is in contrast to the transistor circuit where the reservoir capacitor discharges through the transistors at one end and is charged rapidly through the diode from the other. Here we can adjust the rate at which the capacitor charges and at which it discharges. The implications of these on how the circuit behaves and sounds are crucial.

But why have time constants at all? If the idea is to provide protection for overloads, why bother with how they're handled? Well that's all diode clippers do, Figs. 29-73A and B. They have zero attack time, which means there is no delay or run-up to them when dealing with an overload. Similarly, they have a zero release time, meaning that once the overload is dealt with it's instantly business as usual. The trouble is, they sound horrible, as would either the transistor or the FET limiter with infinitely short parameters.

29.12.1.2 Attack Time

Fig. 29-74 shows the first few cycles of a train of sine waves that are in excess of the limiter threshold and the effect as the limiter tries to reduce the output to the prescribed level. Fig. 29-74A shows a zero attack time and not unexpectedly looks very sawn off. Lengthening the attack time somewhat, Fig. 29-74B, leaves a recognizable but mutilated crest, while longer still is even less bent, Fig. 29-74C. Unfortunately, the character of distortion products generated by this effect are very audible; they are loud (since it is a loud signal that is subject to control) and of high order and unlikely to be masked by the fundamental signal. Even at this stage it is clear that a longer attack time takes less toll of the input signal integrity; the less a waveform is modified, the better it will sound. Expanding the time scale to many cycles shows how lengthening attack time looks. A long attack time, Fig. 29-74F, gradually reduces the limiter gain until the signal is completely under control while imposing less immediate distortion on it.

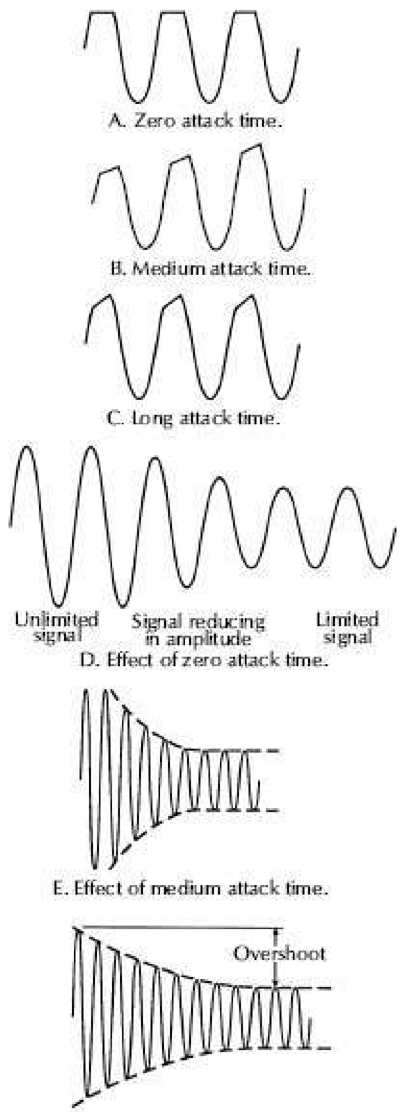


Figure 29-74. Attack time effect on waveshape.

The tradeoff is apparent. An attack time long enough to not mangle the program material also permits excess output level for as long as the circuit takes to bring the gain down sufficiently. Balancing this overshoot against leading-edge distortion due to short attack times is a subjective compromise.

Naturally the lower the frequency, the longer the time period between cycle crests and the greater distorting effect of attack times. An adequate attack-time for high frequencies can easily be still way too short for low frequencies, while an adequate attack time for bass is unnecessarily long for highs. That’s life.

It is normal in a high-quality studio dynamics section to use a full-wave rectifier for the side chain (as opposed to the half-wave shown in these two examples). This gives twice as many opportunities per cycle to sense and adjust the gain (one on the positive-going peak, one on the negative) in addition to allowing for the fact that few real-world signals are symmetrical; either the positive or negative peaks are more, sometimes greatly, pronounced.

Attack times are measured and quoted either in microseconds or milliseconds (i.e., the time constant of the reservoir capacitor versus charging resistor, which also corresponds approximately to the time a transient takes to be controlled) or alternatively in dB/ms which is the rate at which the attenuation changes.

29.12.1.3 Release Time

The purpose of a release time constant is manifold, but in the case of a peak limiter its value is primarily to minimize distortion, much as the attack time. If the incoming sine wave train is above the threshold and the limiter is trying to contain it, and if the release time is short, the gain will tend to recover between each individual crest of the sine wave. In effect, this is the reverse of the attack-time distortion. Although there is the brutality of a fresh set of attack-time-related distortion on each crest, it must not be forgotten that as the attenuation releases there is a less traumatic but nevertheless real change in shape of the rest of the waveform as its amplitude changes within a cycle.

Release times are normally maintained much longer than attack times. With the exception of true transients, which spike up once and then go away for an indeterminate period of time, if not forever, most sounds tend to stay around for a while—at least for a few cycles. It can be reasonably assumed that, once a signal has hit threshold, more of it will follow; given that, there is little point in letting the attenuation drop back just to be reasserted milliseconds later. The release time is a crude memory of the size of the signal the section is having to deal with at a given moment and by keeping the amount of attenuation relatively stable gives less work to and less damage to wreak for the attacking charge ramp-up. A longer release time constant gives the attack circuitry less to do except at the onset of material over the threshold.

There is always the danger with long release times (if chosen to minimize distortion) in that should a large transient come along, the limiter will do its job and promptly reduce the gain to prevent excess output level. Fine, but the long release time keeps the attenuation invoked sitting around for a long time, compressing following program material and a large amount of following information can be lost until the gain claws its way back up to normal level, Fig. 29-75.

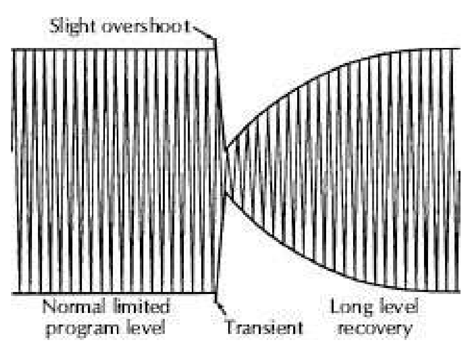


Figure 29-75. Effect of long release times on the output signal when subject to a large transient.

The subjective compromise between long release times (for distortion's sake) and rapid recovery depends largely on the program material. More so with release than with attack; too short a release time can really tear up bass frequencies. Distortions due to attack and release time constants—transient, intermodulation, and harmonic—owe themselves to the fact that gain is changing and rapidly. They are independent of the kind of device doing the attenuation, whether it is a humble transistor or an expensive VCA and they are just as obnoxious. By and large these dynamically induced distortions subjectively far outweigh the steady-state distortion characteristics of the devices. These only become important if the circuit is to sit in a signal path with little or no dynamics processing taking place. Once things start moving one is as good or bad as the other and the subjective quality of a unit is determined by how well the various timings tailor around the program material or how well it does so automatically.

Release settings are generally quoted in milliseconds or seconds, and sometimes as a decay rate such as dB/ms or dB/s.

In short, a dynamics processor lives or dies by its side chain.

29.12.1.4 Compound Release Time Constants

The hole-punching problem of a big transient hitting a limiter with a long release time constant can be attacked in a couple of ways. If subtlety is none too great a criterion—for instance, in an AM transmitter limiter where There Shall Be No Overmodulation—it is always possible to put a variation of the back-to-back diode limiter on the output of a timed one, Fig. 29-76A and deal with the occasional instantaneous distortion bursts. Set to clip immediately above the normal operating output level of the feedback limiter, it not only cast-iron stops excessive output signal swing but also prevents the transient from entering the side chain and digging too big a hole in the following audio.

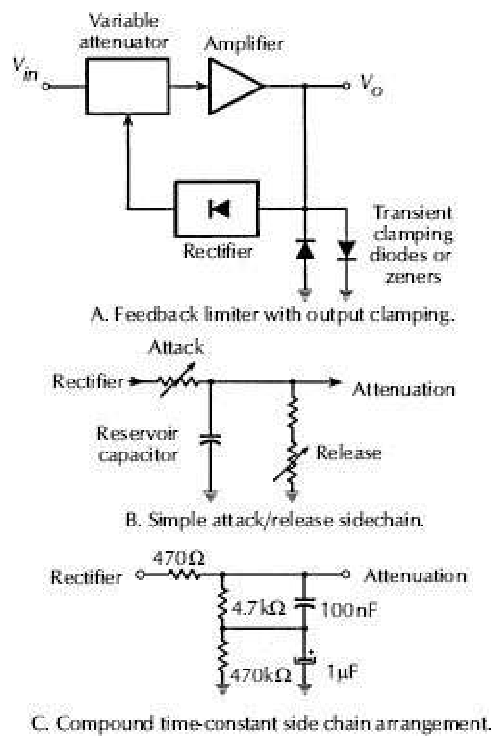


Figure 29-76. A feedback limiter with a compound time constant.

The circuit in Fig. 29-76C is used extensively. Instead of a single reservoir capacitor in the side chain with one attack defining resistor and one release defining resistor, Fig. 29-76B, a compound circuit can be arranged, Fig. 29-76C. A small value resistor and capacitor form additional, shorter attack and release time constants working in conjunction with a slower set. The extended attack and release times follow the general loudness envelope of the program material while the shorter ones, riding on the top take care of any short-term discrepancies and transients, generally to the tune of the top 5dB or so of processing. If a general-purpose, hands-off, no-tweaks limiter is needed, this arrangement with carefully chosen values can work very nicely and is the basis of some commercial outboard limiters in the auto mode.

29.12.1.5 Multiband Processing

A complex but very effective side step to the inapt time constant with frequency problem is shown diagrammatically in Fig. 29-77. Here the input signal is split a number of ways by frequency (in this case three, into bass, mids, and highs). Each band passes into a limiter with side chain time constants optimized for that band of frequencies; this allows the highs to be optimally treated with short time constants without compromising the other sections, and so on. The bands are recombined and passed through an envelope limiter, which only needs to catch a limited dynamic range and so has little effect on the path and the overall sound. Two bands is enough to remove the pumping effect of the usually energy-intensive bass-modulating higher frequencies; three bands allow for better time-constants for the all-important midrange to be established without compromising the high and low frequencies; more bands, say five or six, allow considerable program density (translation: loudness) to be built up while retaining musicality.

This is a very common technique in radio broadcasting airchain processing, and allows for better or worse far deeper processing than possible with broadband units and totally avoids side chain pumping effects, where typically heavy bass modulates the mids and highs content. It is usual for there to be a number of multiband stages, preceded by broadband AGC and succeeded by broadband limiting/clipping. The first multiband section (five bands is common) being compression and perhaps multiband AGC, feeds a second section of multiband limiting (31 bands of limiting is not unknown!). Needless to say, such devices can be quite an entertainment to set up; indeed, a whole subindustry of processor witchcraft has evolved in radio.

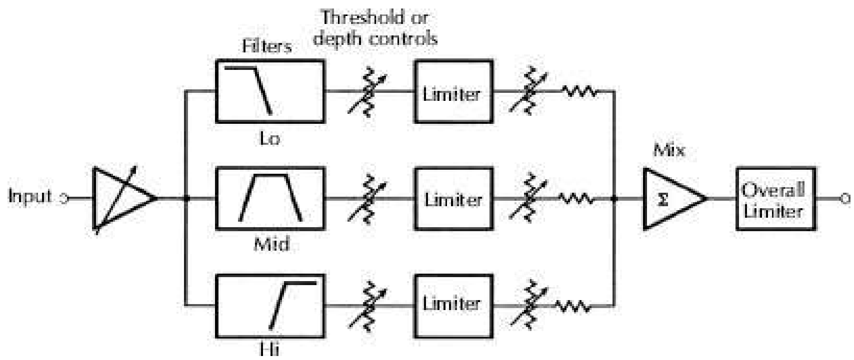


Figure 29-77. Multiband signal processing.

Adjusted well, these units can sound startlingly good (and loud). In corollary, they are far easier to make sound truly dire. Unfortunately, the multiband technique, either using discrete units such as air-chain processors or virtually as software plug-ins to audio workstations and digital consoles, has found its way into music production. Being low on the learning curve the results have rarely been beneficial.

29.12.1.6 Active Release Time Constants

Passively discharging the side-chain reservoir capacitor with a resistor is not necessarily the best way of going about things. Looking at Fig. 29-78A shows that the initial discharge rate is considerably faster than that farther along in time. With a gain control element (i.e., a Voltage Controlled Amplifier/Attenuator—VCA) having a linear control voltage to dB attenuation characteristic, the gain reduction on release would die away very quickly initially and steadily bottom out. This is bad news since a longer than necessary release time constant would need to be applied to preserve adequate low-frequency distortion. If the reservoir capacitor is discharged linearly as in Fig. 29-78B by a constant-current source instead of by a straight resistor (example in Fig. 29-78C), a tidy linear dB attenuation release versus time characteristic ensues; less release time need be wound in for similar LF distortion.

This can be taken a step further. Some gain-control elements with logarithmic (transistors) or square-law (FET) control voltage characteristics, for example, can be made to work with a passive release system to give a pretty good approximation of linear dB/time release. Adding constant-current discharge to one of these circuits gives a slow initial discharge (i.e., good low frequency distortion) with a more rapid tailoff, Fig. 29-78C, which removes unnecessary gain reduction quicker than any other arrangement yet. On program material this works very well, also serving to reduce pumping and suck outs from transients.

29.12.1.7 Hold or Hang Time

Given active discharge with a constant current source, it is always possible to turn the discharge path off. This has the effect of freezing the attenuation at the instant the discharge is removed; if there is no discharge path, the reservoir can't discharge, the control voltage remains static, as, consequently, does the attenuation. Recent refinements to dynamics include this feature; if the side chain is attacking in response to an increasing signal over the threshold, the constant-current discharge is turned off automatically. It remains turned off for a preset amount of time after the attacking has ceased and when the circuit would ordinarily be releasing. Instead, the attenuation remains static for that preset time after which the discharge is reinstituted and normal release decay occurs, Fig. 29-78D. The advantages are straightforward; there is no release time-constant-related distortion in the period of time the attenuation is frozen and the subsequent release can be independently set to return gain as quickly or as slowly as desired. Tailoring the attack, hold-and-release times around a given program source can render processing virtually transparent in many cases. Hold or hang time is quoted usually as a direct time in milliseconds or seconds.

A second application for freezing the gain reduction is if the applied signal level to the compressor falls silent, or below a certain threshold. If frozen until valid signal reappears, the compressor doesn't have to re-establish a control signal from cold with all the time and dynamic modulation involved—it can just pick up where it left off, so to speak. In spoken word for example, the compressor gain does not ride up (often bringing noise with it) in pauses, and breath noise is mitigated. This feature is variously called Freeze and confusingly Gating (as it is very different to the Gating shortly to be described below in section 29.12.2).

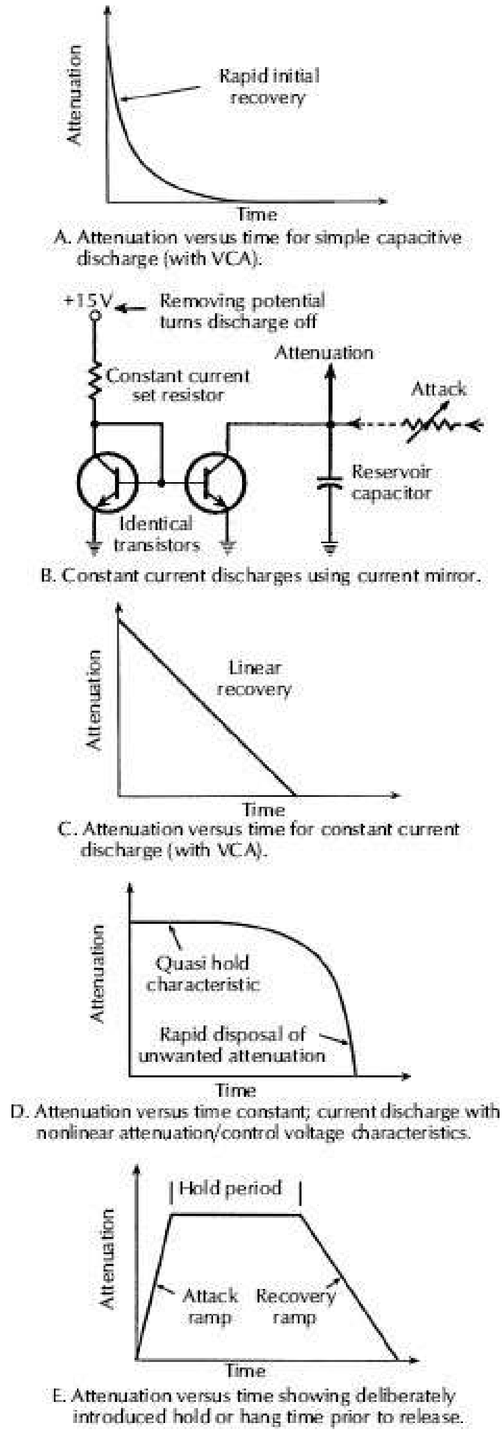


Figure 29-78. Active release side-chains.

29.12.1.8 Limiting and Compression for Effect

The principal creative purpose for limiting or compression (as opposed to the precautionary and damage-control functions outlined earlier) is to make things loud. Suitable parameters can also imbue low-frequency chunkiness or weight on the sound. Given a certain maximum headroom level in a transmission or recording medium, it is often desirable to increase the program density or reduce the dynamic range. A case in point (again) is broadcasting. Their legitimate purpose for compression is to render audible portions of program that may otherwise be too quiet and buried in ambient noise, interference, or static at the receiver end. Automobiles have a notoriously small dynamic window between receiver output capability and cabin noise. A compressor (or usually air chain processor) is set to give sufficiently high output for a quiet program and to automatically reduce the gain when it gets louder.

An interesting subjective side effect of this lies in psychoacoustics; if the ear hears something it knows to be quiet ordinarily at a certain volume, then a sound that it knows to be louder still seems louder even though a limiter may be compressing the two signals to the same level. A classic example is with reverberation tails—these are one of many means by which we subconsciously gauge relative loudness; if the original sound spawning the reverb tail is compressed to be closer to the level of the tail, the whole overall sound seems louder.

Normal program material consists of quite high transients and peaks above the mean average level. If these peaks are removed (brutally by a clipper or more subtly by a timed processor with short time constants), the average transmitted level can be increased correspondingly. The shorter the time constants, the more apparent loudness can be squeezed out. Usually, though, this is at the expense of quality.

Here belies the principal reason broadcasters like dynamics processing—the louder they seem on the air, the more listeners are attracted to the station. To this end excruciating amounts of gain reduction are common on-air in radio. It's a well-known effect that something that sounds louder—even marginally—is perceived as sounding better, at least in the short term.

It is also a strong reason compression is prevalent in individual recording channels within a console; each sound can be made not only more controlled in level, which helps balancing, but also denser and more solid sounding. The downside is that it's so easy to squash vitality out of a sound, trading liveliness and depth for something more up front but ultimately less interesting.

29.12.2 Gating

Gating is to a degree the inverse of limiting. It is the removal of an output signal unless it is of a sufficient strength; in other words, if the input signal is above a threshold level it is permitted to pass, but if it falls below the threshold it is attenuated.

Its purpose is usually to remove or reduce in level a signal when it is no longer usefully contributing to a mix, remove noise in between wanted sections of program and to generally act as an automatic mute. A true gate totally removes the undesired signal but in practice—for noise reduction in particular—a lesser amount of attenuation is invoked; this is set by a control and indicator called depth or maybe just attenuation. Gentle amounts of depth make the operation of a gate far less obvious together with the benefit that there is less intermodulation distortion if the gain is asked to change through less of a range.

The gate open or wake-up time is generally adjustable and determines how quickly the gate opens in response to a signal tripping the threshold. It is usually set very fast, though, such that none of the leading edge of the signal is missed. The hold time (if there is one available, usually) determines how long the gate remains open after the signal drops below the threshold and the release or decay time sets how quickly the attenuation returns. That these are a direct parallel to their behavior in the limiter is no accident; nearly all dynamics processing sections carry these controls. It only needs to be remembered with a gate that the open (attack) time has to do with how quickly attenuation is removed rather than applied as is the case with a limiter.

The ranges of time-constant values are typically similar to those for a limiter, but the threshold range can extend from 0dBu (or even as high as +20dBu in some cases) down to about −40dBu or below. The higher thresholds are mostly for key triggering while the low extremes are for noise reduction. Automuting settings are somewhat critical, needing to be above the general background level yet below the desired signal's typical level; there is always some tuning to be done, but figures of −10dBu to −20dBu are typical. Depth can be adjustable between 0dB and 40dB attenuation (some manufacturers optimistically state infinity).

Practical uses include automatic microphone muting (backup singers), spill removal (e.g., a snare drum microphone is usually gated so that when the snare isn't actually being hit, the microphone isn't picking up the rest of the kit), and noise reduction (just enough gating applied to recorder track returns to subdue tape hiss or air conditioning rumble, say). In all cases the parameters are set up to be as unobtrusive as possible. These vary from lightning fast attack and decay on a snare drum to fairly leisurely ramps in noise reduction.

In addition to the hold, or hang, time which prevents the gate from chattering on a marginal signal, an additional tool to prevent falsing is hysteresis between the signal level necessary to open the gate (open threshold) and that below which the gate considers the signal to have gone away (close threshold); this hysteresis (a few dB) is generally concealed from the operator.

29.12.2.1 Gating Feed-Forward Side Chain

Naturally, a gate cannot possibly operate with its side chain taken from the amplifier output as is the case with the feedback limiters described earlier—it would never open. It has to sense prior to the attenuator in the signal chain. This arrangement is called feed-forward sidechain sensing and is the prevalent method of generating control voltages in today's dynamics processors. Fig. 29-79 shows a typical gate circuit using this method; the input signal as well as going to the attenuator hits a variable gain amplifier, which determines the threshold. The more gain in the amplifier, the sooner the detector threshold is reached. Following the threshold detector—which is in this case a comparator type yes/no level sensor—are the various time constants. Depth is controlled by placing a limit on the amount of attenuation possible.

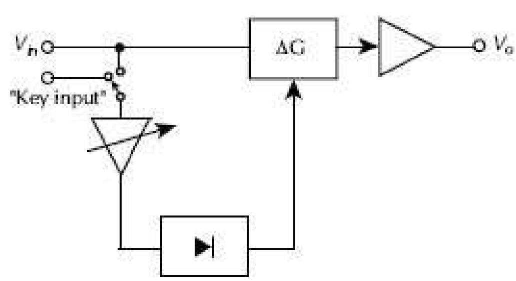


Figure 29-79. Gating feed forward sidechain processor.

29.12.2.2 Subtractive Feedback Gate or Expander

An alternative gating method cunningly uses a limiter, albeit with a very low threshold, to subtract from or cancel the straight signal, Fig. 29-80. The signal gains through both the straight path and the limiter (below its threshold) are arranged to be the same but out of phase; they cancel out. Above the limiter's threshold the limiter output remains fixed but the straight signal is left unhindered, so the two no longer cancel, leaving the straight signal predominating. Time constants of the effective gate are determined by those of the limiter, threshold by the gain of the amplifier within the limiter loop, and depth by contriving a mismatch between unlimited level and straight path level to produce less than total cancellation and some residual.

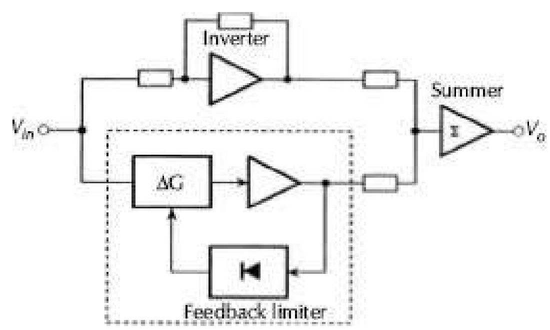


Figure 29-80. A subtractive gate, canceling a limiter from inverted input.

29.12.2.3 Keying

Keying is the triggering of a gate from an external source that is not from the signal that is actually passing through the gate. Perhaps the best and most commonly heard examples are keyed snare and kick drum sounds—all the rage in the dark ages of disco. In circumstances where a new drum sound was needed or alternatively the existing drum sound was not fit for human consumption (very common on live stages), the existing drum sound is used to key a gate that is carrying in its signal path something which can be convincingly shaped into a better drum noise. Favorites are white, or EQ'ed white, noise to emulate a snare sound; similarly, some tone around 20-60Hz (sometimes even ac line hum) when shaped by the attack, hold, and decay times of the gate can make for a good kick drum!

29.12.3 Compression

As briefly outlined at the beginning of this section, compression is where the output signal from the processor does not increase as much as the input signal is increasing. If an input signal jumps in level by 10dB, a compressor with a ratio of 4:1 would only allow the output to rise 2.5dB. Correspondingly, a drop in input level of 16dB into the same compressor would result in 4dB output level change. A compressor reduces the dynamic range of an input signal by the amount of its ratio.

A true compressor acts on all signals, regardless of actual signal level, in the same manner. No matter if the input signal is way down at –60dBu or up at +20dBu, a change in input signal level of a given amount will cause a similar, reduced, change in output signal. Practically speaking, there is no such thing as a true compressor; things that come close and work down to very low signal levels are used in noise reduction systems for telephone lines, tape recorders, and wireless microphones, where they are used with a complementary expander (see later) to reinstitute the original dynamic range.

Most compressors have a threshold below which they leave a signal unscathed (a 1:1 ratio) and above which they proceed to compress the dynamic range, much as a limiter does. The family resemblance becomes all the more striking as compressors with high ratios are considered. A 10:1 compressor above its threshold reduces a 10dB input level jump to just 1 dB. Infinity-to-1 reduces anything above the threshold to the same output level. Looks like a limiter, smells like a limiter. Generally, compressors are used at far gentler ratios (between 1.5:1 and 4:1) to bring up lower level program material in a less take-it-or-leave-it manner than a limiter while leaving some sense—albeit reduced—of light and shade, louder and quieter. They are also used to subtly make sounds chunkier—a degree of compression tends to accentuate lower frequency elements of a sound, which are those generally most predominant and are so controlling of the gain reduction.

Major differences between limiting and compression are in the nature of the side chains, the level detectors, and in particular typically applied time constants. Limiters almost invariably have peak detectors, such that the peak of a waveform is detected and, time constants allowing, protected from overload by the limiter; this is fine for the protection mandate of limiters. Compressors, on the other hand, tend to have much more relaxed attack and release times, such that they are less intense sounding than the typically frenetic limiter settings. Similarly, since the concern with compressors is less peak level than loudness to the ear, which tends to gauge by overall signal energy or power rather than peak values, the detectors are typically average or power sensing. The slower time constants go a long way toward this by essentially ignoring peaks and responding more to an average over the time imposed by the attack and release times. Deliberate averaging detection (as opposed to the more or less accidental) sounds much more even and unobtrusive than peak detection; taking it a step further, power detection by means of a root mean square (rms) level detector is better yet. In reality, though, there's little to choose between average and rms detection, since although they give somewhat different answers under test waveform circumstances, dancing around in the heat of audio battle they are quite difficult to tell apart. Occasionally either one or the other will be fooled by a difficult piece of program material.

The threshold tends to extend farther down for a range typically of –30 to +10dBu, and the ratio adjusts from 1:1 (straight) usually to infinity:1 or close (limit). Some commercial units extend the ratio beyond infinity to negative values; that is, if a signal progressively exceeds the threshold it gets progressively further attenuated! Although on first glance it seems a bit pointless, it does allow fairly nice sounding level control for a compound signal; it permits looser (longer) attack times than would be possible on an ordinary limiter, with the resultant overshoot merely propelling the signal farther downward away from possible overload. It is also good for some pretty silly effects on individual instruments.

Most compressors (and limiters) are "hard knee", meaning the transition at the threshold between when they are acting and not is quite definite. "Soft-knee" compressors instead gradually morph over the first few dB of gain reduction from inactive to fully compressing; in many applications this is sonically preferable, especially where compression is being used for control rather than for effect.

29.12.4 Expansion

An expander increases the dynamic range of its output signal in relation to the input signal. Its ratio determines how much: a 1:3 expander renders a 4dB level shift in input signal, which results in a 12dB difference at the output.

As mentioned under compression, true full-range expansion is a rarity and is generally only found as a complement to a compressor in a double-ended noise reduction system. In these circumstances they are nearly always of 1:2 ratio with an axis point (the level at which the input signal is the same as the output signal level) of around 0dBu.

Practical expanders come with a threshold setting, above which they leave the signal alone and below which gain reduction sets in. Sounds a bit like a gate? A gate can be emulated by an expansion with a ratio of 1:infinity; any signal below the threshold gets attenuated away completely—the one exception is that expanders don't always have a depth setting. The purposes of expansion are very similar to those of a gate, only generally it can sometimes do a better, less noticeable, job. A relatively gentle expanding slope (say 1:2 or 1:3) can provide the same degree of noise reduction as a gate with less abrupt changes in gain; since the signal is audible still (but quieter) and doesn't have to be resurrected with a start to normal level, fairly gentle (slower than a gate) attack times do not have as noticeable a softening effect on the required leading edge.

Expansion side-chain time constants are similar to those for a gate, as is the threshold range. Ratio, as with compressors, is usually 1:1 to 1:-infinity, although often "classic" implementations have a fixed ratio of something close to 1:2. Expansion is used as subtle gating in much the same way as compression is a gentler substitution for hard limiting.

29.12.5 Feed-Forward VCA-Style Dynamics

The feed-forward class of dynamics owes itself to the development of VCAs and similar log/antilog processing; it is exemplified by the classic dBx160 series. As far as consoles go, the mere existence of a VCA in the channel for fader automation begs for this style of processing to be incorporated. (VCAs are further discussed under Consoles and Computers, later). Fig. 29-81 shows such a processor in block diagrammatic form.

Key to VCA dynamics is the inherent exponential (logarithmic) control, which relies on reasonably simply implemented basic transistor behavior (base voltage versus current). Gain (or gain reduction) of a VCA is as good as linear dB/v, which can lend to a deterministic design approach (meaning one can pretty well predict what the circuit will do within narrow limits, without a servo loop to help). Simple log/antilogging lends itself to another typical feature of VCA dynamics sections.

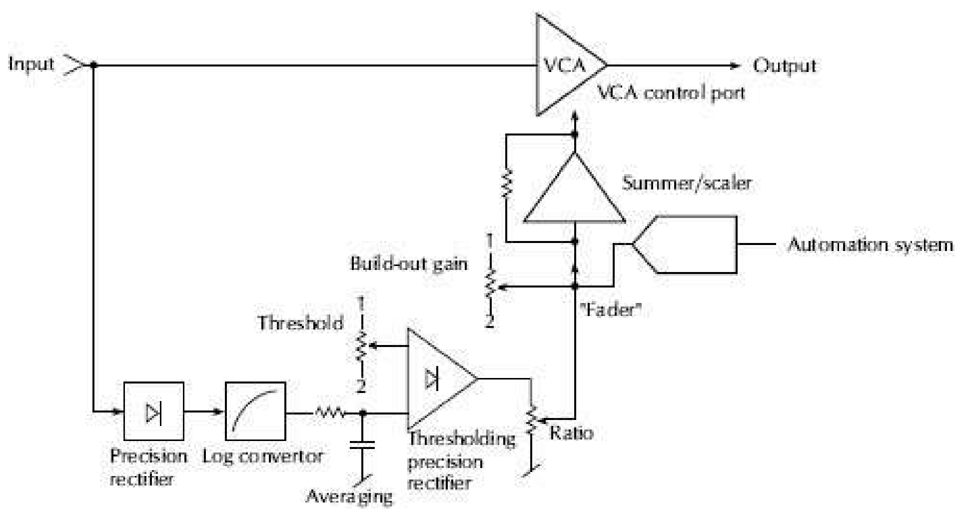


Figure 29-81. Block diagram–feed-forward VCA Dynamics.

29.12.5.1 Rms Detection

Hitherto, detection of signal levels in dynamics had been either peak or average. These were actually achieved by broadly similar circuitry with the difference dictated by the attack time applied after signal rectification; short attack times allowed the reservoir capacitor to charge immediately to the highest signal level applied, while longer attack times tended to smooth out the peaks, settling on an average value of the applied rectified waveform. And a sort of mushy continuum existed between the two.

Rms (root mean square) detection has the intent of providing a measure of the energy in an applied waveform, the actual power. The reasoning is that a power measurement is considered more equivalent to loudness. It is far from an exact relationship, but there is strong correlation. Rms is achieved by first squaring the applied signal, i.e., multiplying it by itself, not turning it into a square wave, finding an average value of that squared source, and then determining the square root of that average (unsquaring it). Seems like a lot of bother to go to, doesn’t it? Well, no one would have bothered if there was a reasonably straightforward method. This comes from an application of a log anticell.

A precision-rectified (meaning accurate down to very low levels) input signal is logged, and then its output is doubled (added to itself); doubling a log value squares the number it represents. This signal is then integrated with a time constant long enough to allow reasonable averaging of the lowest frequency under consideration; this incidentally defines the minimum attack time of the processor. This log value average is then halved (division of a log value by two is the same as finding the square root), delivers a log-world rms-detected output. (In this circumstance a subsequent antilog conversion is unnecessary. Actually, the square rooting is ignored at this point, too, since it can be achieved in a later scaling exercise.) The good news is that all that can be done with a handful of transistor junctions. Release time can be extended with a following buffered capacitor, but often the imbued time constant of the rms detection serves as symmetrical attack and release. This somewhat leisurely time response (necessary to permit good rms detection at low frequencies, devoid of distortion-creating ripple) in and of itself ensures that the behavior of such a dynamics section can't get too wild and interesting, but by corollary such processors do afford probably the least intrusive method of automatic volume control, which is a highly prized attribute on occasion.

29.12.5.2 Thresholding

The rms-detected control signal is then masked in a threshold determining circuit; typically this is a precision rectifier with its reference point determined by a threshold control voltage—the purpose of this is to ignore all variation of the detected voltage until it exceeds (in the case of a compressor) the threshold point, beyond which its output follows the rms detector output. Any control signal escaping the thresholder still has a dB/V characteristic, being still logged, following the input signal. Another (linear-think) way of looking at this is that a division takes place; the detected control signal is divided by the threshold, but with any result less than 1 masked out at 1, only greater than unity results being passed.

If the thresholder is designed to pass only changes below the threshold, then the low-level effects of expansion and gating are possible, signals above threshold being ignored. Separate thresholders and following conditioners are necessary for each desired function of the dynamics section.

29.12.5.3 Ratio

If this thresholded control voltage were applied directly to the (level-adjusted) control port on the VCA, something odd would happen: nothing. More precisely, above the threshold the control signal would rise in accord with a rising applied audio signal to the precise extent that the gain reduction resulting from it would be exactly the same as the increase in signal. The VCA output would remain at a fixed level for any applied signal level above the threshold. In other words, it is a compressor with an infinity:1 ratio, meaning that above the threshold, any amount of signal level variation will have no effect on the output. In yet other words, it would be a limiter (albeit with slow dynamic response).

Introducing a variable attenuator in the feed to the VCA control port from the thresholder affords altering the amount of dynamic gain reduction; less control signal variation, less gain reduction. A nice feature of this very simple approach in log world is that this attenuation (equivalent to solving for variable roots in linear) results in precise applied signal level to dB gain reduction ratios; for a given setting, if the input signal were to rise 6dB, the output would rise only 3dB; this ratio, 2:1, would obtain linearly for any applied signals above the threshold.

Astute readers may wonder what would happen if the control signal, rather than being attenuated, was amplified instead. Yes, the degree of attenuation would become bigger than the corresponding changes in applied signal, and the VCA's output would actually increasingly reduce beyond the threshold. This effect was brought to stardom by the Eventide Omnipressor.

29.12.5.4 Side-Chain Sources

Figs. 29-81 and 29-82, the block diagram and schematic respectively, of a feed-forward style dynamics element, both only consider the side chain as being taken from the same place as the input to the VCA gain controller. This need not be the case at all. In fact, with a standard channel's signal processing it would be quite a limitation, forcing the dynamics section to be solely post everything, prefader. The side chain input may be separated and instead taken from pretty much anywhere upstream; the overall effect is (with only some odd side effects) the same as physically moving the whole dynamics section to that location. For the most part, the audio doesn't care that the control voltage from the sidechain has no relation to that being passed through the VCA. The main areas where things might sound awry are if there is a significant (and audible) deliberate time delay between sense and activation or extremely short time constants are invoked. The first disconnect is obvious, and the second is almost irrelevant since it wouldn't be sounding very nice anyway.

Taking this a step forward again, a prime virtue of a VCA-based system is that many different sources can operate simultaneously on the one relatively expensive VCA gain element. If there are multiple side chains (say, one for keying/gating/expansion, another for compression) these again need not sense from the same pick-off points but places more suited. The gate would likely sense postinput filters, while the compressor would likely be farther downstream, one side or the other of the EQ. This situation would work, but would result in behavior unlike having a discrete gate up front and a discrete compressor downstream. In the literal case, i.e., with two separate dynamics elements—the incoming signal would be gated before it hit the compressor. In this virtualized case, though, the actual audio signal hitting the compressor side chain would not have previously been gated and hence cause the compressor to act differently than if it had. Subtle, maybe, but a definite difference.

29.12.5.5 Makeup Gain

The side chain's thresholded and ratioed control signal is summed in with a voltage representing the amount of buildout gain (necessary to compensate for the signal level reduced by the effect of compression/limiting) and is also, in the case of a typical console channel, summed in with the voltage from the automation system representing the fader position. This summation is scaled to suit the actual (highly sensitive) VCA control port, to which it is fed.

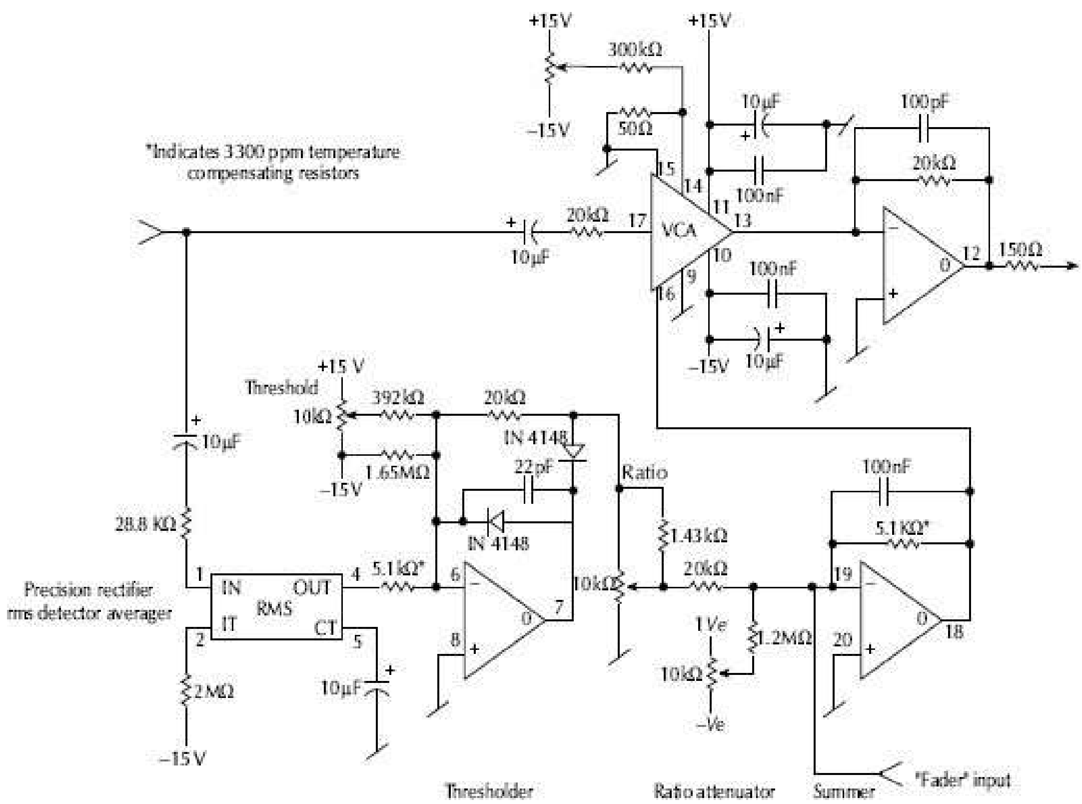


Figure 29-82. A feed-forward VCA compressor using THAT 4301.

29.12.6 A Practical Feed-Forward Design

A highly integrated part, the THAT Corporation's 4301, has both an rms detector and a VCA built in, in addition to some op-amps for glue. A simple, very low parts count compressor with all the active elements contained within this part is shown in Fig. 29-85. As can be seen, it relates strongly to the block diagram of Fig. 29-84.

29.13 Mixing

29.13.1 Virtual-Earth Mixers

The circuit diagram of Fig. 29-83 in its simplicity belies the hidden design that is in the relationship of the circuitry to its mechanical and electrical environment.

This is where the care and feeding of op-amps (section 29.7) and grounding paths (section 29.8) really pay dividends. Mix-amp stages, with large numbers of permanently assigned sources such as in the main mix buses, are as crucial to the overall well-being of a console as any front-end stage could be. In a typical situation, as a unity-gain virtual-earth mixing stage with 33 sources (channels plus access), the amplifier is being asked for about 30dB of broadband gain, as much as any other stage in the chain including both the microphone preamp and/or secondary input stage.

29.13.2 Noise Sources

All the following about mix devices assumes that system grounding is impeccable. Jolly good. That said: That mix-amp gain is sometimes referred to as noise gain is not accidental. Unless care is taken to balance fader-back channel noise contributions against this self-generated mix-amp noise, the latter could well predominate and arbitrarily determine the noise floor for the entire console. Similarly, channel noise contribution should equal or outstrip mix-amp noise, but not excessively so: ideally they should equally contribute, to the extent that channel-off noise contribution should not necessarily impact the overall bus noise, while bus noise should not significantly impact channel-on noise. Self-noise generation in the mix-amp is predominantly the amplified thermal noise of the paralleled source and feedback resistances, device input current noise, and surface generation and recombination noise. The last two can be minimized by device choice. Thermal noise is physics and is here to stay. Common sense on first glance says to make the mix resistors as low in value as possible but this has the downside that too low a value would cause quite large signal (hence, ground) currents to be thundering about. On a less technical and more economic level, it necessitates somewhat beefier and more serious buffer amplifiers on each source to feed the buses.

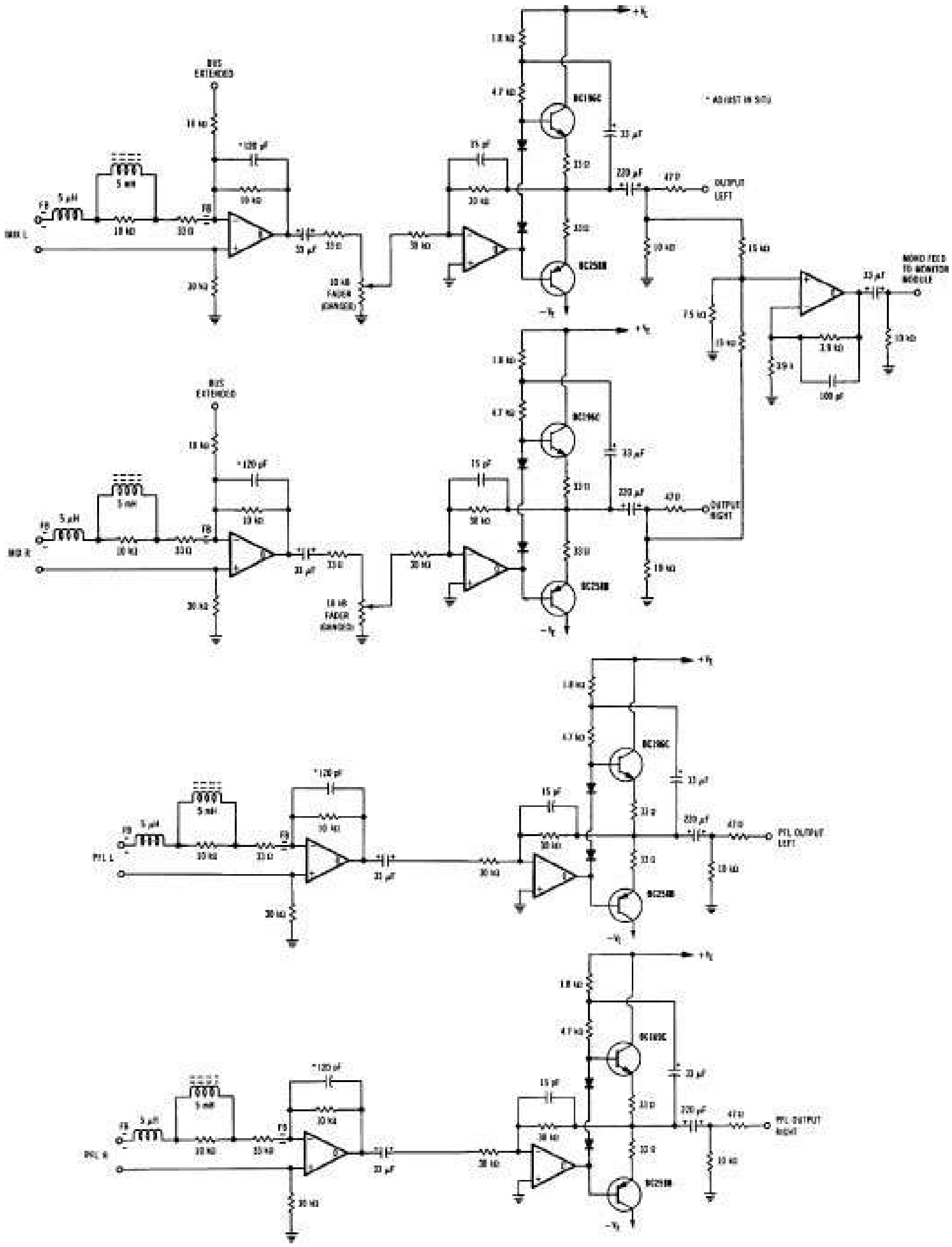


Figure 29-83. Master group electronics.

Ordinarily though, the mix resistors are of such a value that, in the context of a complete mixer, the combined effectively paralleled resistance is well below the optimum source impedance of nearly any mix-amp device available, so the primary noise modes are those above-mentioned device vices. This certainly isn't too difficult with

FET front-end devices, with their high OSI (optimum source impedance). These devices have a couple of other major benefits in this application though by virtue of their FET inputs. Input current (hence, input current noise) is extremely low, and being FETs they don't have the many low-frequency junction and surface noises inherent to bipolar devices. It seems a paradoxic absurdity to use an ultrahigh input impedance device for zero impedance mixing, but in many ways and under some circumstances they're better suited than bipolars. On the other hand, the intrinsically superior noise performance of a 5534-class device can pay dividends in this application. Like so many cases in console design each individual application needs staring at for its own optimum solution. This is all really only a problem for those who have the luxury of designing small mixers or where it is more or less guaranteed that only a small number of sources will be allowed to hit the bus simultaneously and hence where the parallel impedance of the sources remains fairly high. In most midsize and large consoles without these constraints, mix device noise will likely predominate. Device choice will be down to its self-noise (of course), and output current capability if the summing resistor value is low, and ability to cope with a big hairy capacitative bus sitting on its input current node. Integrated mic-amps have been successfully used as differential passive mix-bus amplifiers, which with their very low OSIs stand a chance of getting closer to that low bus impedance and low bus noise nirvana. However, as alluded to earlier, the channel-off noise contribution from all those bus-driving amplifiers in all those channels is more likely to then predominate. It is a balancing act.

If bus noise performance truly is a major concern (as it could possibly be on a tracking console) removal—as in physical disconnection—of all unused sources from the bus at all times is the best way to get the noise gain down and that bus impedance back up to where mix-amp noise can be optimized to it. No way to run a railroad or a mixdown console, though.

Things can get a bit startling if the resistance/OSI relationship is awry, particularly on relatively small consoles. Above the OSI as much as below its OSI, device noise becomes an increasingly important noise contribution. Many years ago in a mixer design with bipolar device mix-amps and quite high mix resistors, the measured bus noise was actually quieter on a 20-channel version than on the 10-channel original. It wasn't until much later that what was actually happening finally dawned. Increasing the number of source resistors reduced the bus impedance, previously well above the OSI of the amplifier with only 10 sources, to closer to the OSI, where input noise voltage was contributing less.

Theoretical source impedance and device contribution tell less than half the story in a practical design. They may be quantifiable in the isolation of a test bench, but thrown into a system they can all seem a bit meaningless. It's all largely a matter of grounding and out-of-band considerations.

29.13.3 Radio-Frequency Inductors

Inductors are used between the bus and the amplifier input in Figs. 29-83 and 29-84. A simplistic view is that they are there to stop any radio frequency on the mix bus from finding its way into the electronics, but this is only part of their purpose. The ferrite beads and small chokes (about 5 μH) are there to increase the input impedance and hopefully help decouple the bus from the amplifier at very high frequencies. The larger inductance creates a rising reactance to counteract the falling reactance of the bus capacitance. If left completely unchecked, this capacitance would cause the mix-amp extreme high-frequency loop gain to turn it into an RF oscillator. Feedback phase leading around the amplifier stops the gain from rising, but if it were not for some series loss (accidental or deliberate) in the input leg, it would be insufficient to hold the phase margin of the amplifiers within their limits of stability, especially at bandwidth extremes where device propagation delay becomes significant in the loop. A small series resistance can provide this loss while also defining the maximum gain to which the circuit can rise. A parallel inductor-resistor combination improves on this in a few important respects.

The inductor is calculated to present low in-band (<20kHz) reactance, allowing the mix-amp to operate on the bus in its intended virtual-earth (zero-impedance) configuration. The reactance rises gently at the audio high-frequency end, imparting little frequency response anomaly but a definitely beneficial partial phase straightening against the inevitable effect of heavy bus capacitance.

At even higher frequencies, the inductive reactance continues to rise until the combined network impedance is limited by the resistor, which is of high enough value to define amplifier out-of-band gain to a reasonably low value. It is low enough, however, to stop the inevitable inductor-bus capacitance resonance from getting completely out of hand. Making a stable inductance-capacitance oscillator is one way of preventing spurious instability, but is not exactly the desired end here.

While FET inputs are far less prone than bipolar inputs to the intermodulation and direct demodulation effects that cause RF interference to appear out of nowhere, this fairly healthy brace of filtering may be helpful to those living near a source of high-powered very high frequencies, such as a group of television transmitters.

29.13.4 Virtues of Grounding

Grounding paths for virtual-earth mixing, especially in long mixers, are always the final arbiter on how far down the system noise floor will go and how susceptible the mix stage is to extraneous fields and earth currents. In this age of digits, ground paths are especially crucial. Remember from Fig. 29-62 how the ground noise on the noninverting input of an op-amp mix stage gets amplified up by the noise gain of the stage? This implies that a ground noise of −100 dBu will end up at about −70dBu for a 32-source mixer, which is hardly adequate.

A simple, but so often ignored, rule with virtual-earth stages is to make sure that the ground reference has got the same dirt on it as the signal and vice versa. Yes, ground follows signal. If both ground and signal have the same noise in the same phase, there is a chance that the noise will get ignored as common mode and not amplified in the mix-amp. So, for each mix bus, there should be a parallel ground bus being fed by the last relevant ground reference from each channel. Avoiding a major bus-length ground loop (otherwise known as a single-turn transformer!) means that all the heavyweight signal current in the channel proper, e.g., fader/mute/mode switchers, has a direct wire to central ground while the mix-amp has a respectable output referenced ground to work against, clean of channel signal currents but representative of the reference of the buffer amplifiers. The mix-amp does not take a direct system central ground of its own.

29.13.5 Passive Mixing

There are, of course, alternatives to single-bus virtual-earth mixing. Passive resistor mixing, Fig. 29-84, is quite viable for fixed-assignation systems that are not going to be chopped, changed, or switched in and out. A major advantage is that bus capacitance is merely something to be taken into account in terms of frequency response and phase rather than directly imperiling the stability of the mix-amp. For passive mixing, the mix-amp is just a buffer amplifier to make up the loss in the resistor tree; RF filtering becomes simple with known filter source and load impedances together with the ability to refer against ground. A primary weakness is that the bus is unbalanced and is of some impedance at audio (albeit fairly low due to paralleled sources). As such it lays itself wide open to induced noise and capacitatively coupled crosstalk. Despite this, it is a method used with considerable success for many years in quite a few production mixers.

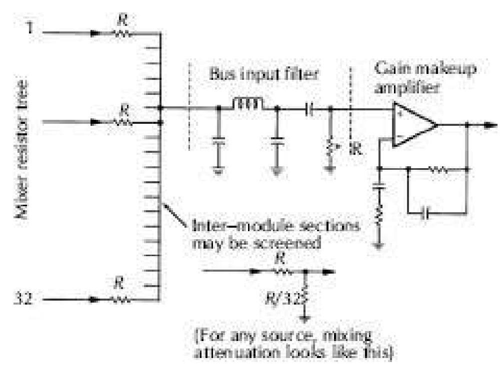


Figure 29-84. Passive mixing arrangement.

In all cases but especially in small mixers, say with fewer than eight sources, there is a theoretical noise advantage to passive mixing over virtual earth. As an extreme example, simple summing of two sources of passive mixing calls for 6dB of gain to make up for the loss in the summing network. A virtual-earth mixer needs around 10dB. Beyond eight sources this advantage tends to the insignificant.

29.13.6 Devolved Mixing

Distributed or devolved mixing, Fig. 29-85, uses local mix-amps to sum relatively small blocks of channels; the outputs of these local amplifiers is then taken to a common summing point. This quite neatly obviates having to deal with a long bus but does create a practical problem of locating the distributed summers.

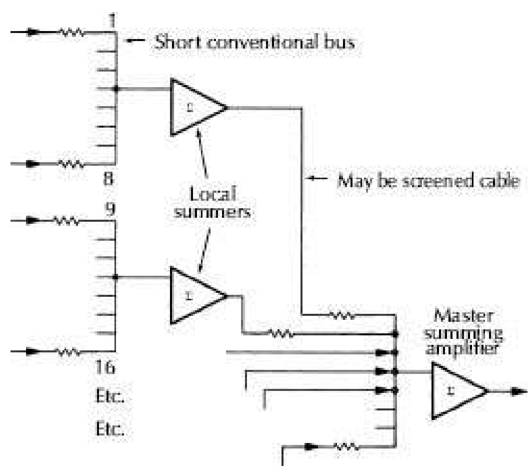


Figure 29-85. Distributed or devolved mixing.

Both passive and devolved systems have the advantage that large amounts of the bus can be run in shielded cable. The extra capacitance here does not have the awful consequences it does with long virtual-earth summing amplifiers.

For consistency—if this approach is taken—all buses should be run devolved. This means the submix facilities for the PFL buses, effect sends, foldbacks, the main stereo/monitor mixer, and analog subgroups (if used). Also, provisions must be made to arrange the master mixer for each of those at the grouping end.

29.13.7 Balanced Mixing

The earliest form of signal mixing consisted of directly paralleling the sources, which were generally medium-impedance (nominal 600 Ω) and balanced. This form of passive balanced mixing persisted until semiconductor electronics and its readily achieved zero impedance transpired. The balancing was done entirely by transformers; again, things that have fallen at least partially by the wayside. As a technique it was simple (for the technology at the time) and maintained all the advantages balanced systems have in general—principally a welcome robustness and immunity to interferences, induced noise, or crosstalk.

Balanced or differential mixing has become practical again with falling component costs and the development of simple electronic differential and floating balanced input and output circuits (see sections 29.9.6.1 to 29.9.6.4). Fig. 29-86 shows how differential sources of the trivial kind (straight and inverted) can be mixed onto a balanced virtual-earth mixing bus, created and sensed by a superbal input stage.

Although requiring a comparatively large number of parts, the performance of such an arrangement in the context of a large multitrack console is truly staggering, especially noise, head room, electromagnetic field rejection, and crosstalk. The noise improves in two respects:

1. No longer is the mix-amp amplifying the noise on its reference

ground. It is referenced to itself, effectively.

2. Square law noise summation—twice the signal (coherent) means 6 dB gain, two lots of incoherent noise 3dB gain, bingo, 3dB noise advantage.

Headroom, by virtue of two signal paths carrying the same information differentially, is 6dB higher. (Naturally the noise and head room are interrelated; whichever is more pressing in a given circumstance necessarily takes precedence in the level architecture.) The RF field and crosstalk rejection improvements are dramatic, but they really ought to be expected from the naturally self-canceling nature of balanced systems.

All the problems of keeping virtual-earth mixers tidy and stable apply twofold here; of course, bus buffering is strongly recommended, mostly to allow the bandwidth definition around the superbal to be effective.

Passive balanced mix-amps can be arranged around integrated mic-amp devices such as the THAT 1510; being single-ended output doesn't lend them to dynamically generating differential virtual—zero impedance mix buses, but does allow the choice of mix resistor values versus mix width to optimize the parts mix-noise contribution. It is nowadays difficult to consider any serious large console design that doesn't use balanced mix buses.

29.13.8 Pan Pots

As outlined earlier pan pots are a means of positioning a monophonic image somewhere within a stereophonic image plane. About the simplest pan pot is shown in Fig. 29-87A where a pair of linear potentiometer tracks are complementarily wired; one goes up, the other goes down. All well and good and even the sums work out nicely; if the L and R outputs are subsequently remonoed, the summed signal remains at the same amplitude regardless of pan pot position—center, either end, or anywhere in between. Subjectively, though, the image seems too loud at the peripheries (i.e., extreme left and right) and subdued in the middle.

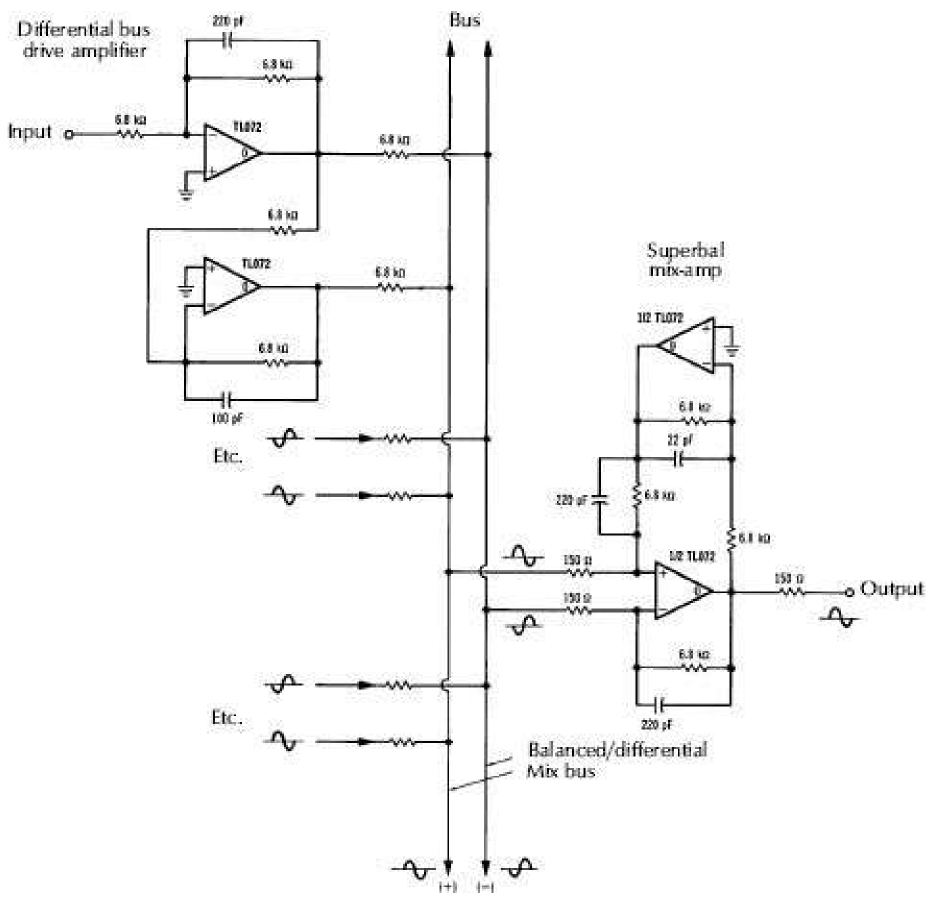


Figure 29-86. Differential balanced mixing.

Replacing the linear pots with a ganged log/anti-log pot (the log section wired upside down) performs much the same function but with a different law, Fig. 29-87B. If a signal is panned steadily right, the left hand output is steadily attenuated, leaving the right output fairly steady in level (in practice it shifts about 1dB). The center position sees both L and R only attenuated slightly (<1dB) with respect to the starting mono signal. Not surprisingly this has the opposite subjective effect to linear pots: the image seems louder in the middle than at the peripheries. Despite that, often this law is more appropriate, particularly where the pan pot is used as part of multitrack odd/even panning or in use as a correctional offset control. In these cases there is virtue in leaving at least one side fairly unscathed.

Somewhere between these two extremes (if extremes is the right expression for a 6dB difference) should lie a happy medium at which the signal keeps an even subjective level panning across the image plane and also tracks well, i.e., has good correlation between control position and image position. Easy? This has been the subject of raging controversy and opinion for decades. Should it be 2, 3, 32, 4, or 42dB down at center?

As is often the case, those closely involved with the theorizing somewhat lost touch with how a pan pot is ordinarily used. A pan control usually remains rusted at an initially set position for hours, days, weeks, months—however long the mix takes. If a pan pot is used dynamically for effect during a mix, its very drama drowns any question of whether it was “a wee bit quiet in the middle”.

A single pot used as in Fig. 29-87C allows a choice of central down points by adjusting the relative values of the source resistors and pot value, but at the cost of slightly iffy tracking (most pan effect tends to happen at the extremes of the control travel) and ultimate panning. When panned hard one way it is nearly impossible—due to wiper-track resistance—for the diminished side to achieve complete attenuation. If 40-odd dB is good enough then this may be the one. Shown are values for a 3 dB down panpot.

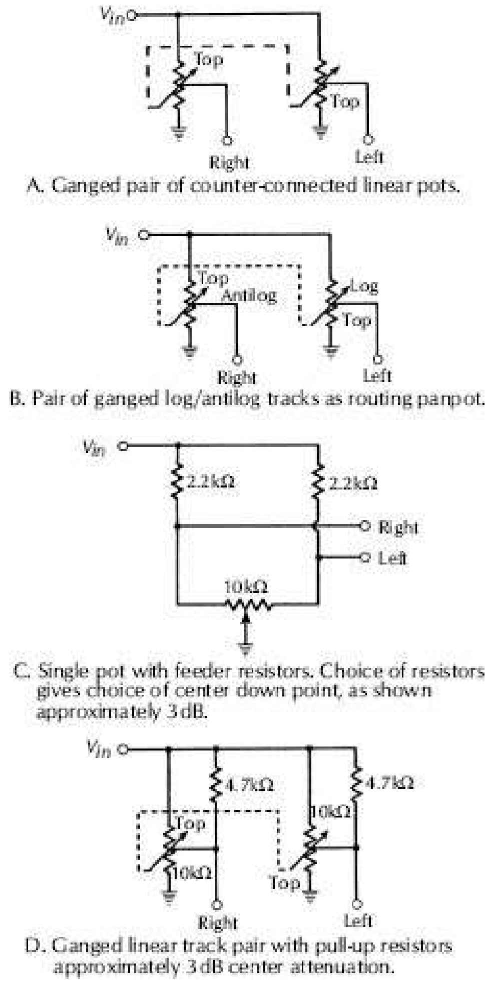


Figure 29-87. Panpots.

During the ′70s the BBC had the understandable problems of multitudinous operators, countless consoles of varying antiquity, and a considerable number of console suppliers. Their aim was consistency and to this end evolved a dazzlingly simple arrangement shown in Fig. 29-87D recommended for inclusion in new supplied equipment; it works.

29.13.9 Surround Panning

Surround has many variants, but for the purposes of discussion here 5.1 will be considered; other formats have basically similar requirements that may be taken as a subset or extrapolated from 5.1.

Well, quad is back (and it hasn't forgiven). At least in terms of four of the 5.1 signal paths, left front, right front, left rear and right rear. Panning for these is achieved in much the same manner as it was in those chillingly far-off days, a joystick that controls relative proportions of the source signal to the four output, or a pair of pots that do substantially the same; one pans left/right, the other front/rear. The "1" of the remaining 1.1 is center; it is to where central dialogue or vocal is panned; this is usually achieved with a blend or similarly named control, which cross-fades the source signal between the center channel and the quad pan pot. In this manner, a source can be directed to any one path, all, or a combination for desired effect.

The last 0.1 is a bass sublow channel. The .1 means that it is (but not always) band limited. It usually has its own level control independent of the full-bandwidth panning.

The surround panned outputs form a channel with six dedicated surround mix buses, which are treated as a married set within the console much as the main stereo mix-bus is/was.

29.14 Monitoring

Monitoring is probably the single most important section of a console. Without it the engineer cannot listen to the results of his labor. At its simplest, monitoring consists of a power amplifier and loudspeakers hung across the main output(s) of the console, with the auxiliary functions either unused or preset. In public address (PA) work the PA actually is the monitoring; the only other function necessary is prefade listen (PFL) and then really only during panic mode. At an alternate extreme the monitoring demands for multitrack recording extend to an entire secondary submixer replete with panning, pre/post foldback effect feeds, and stand-alone soloing, together with listen access to all console send and return ports. The in-line console principle makes efficient use of electronics to combine often coincident signal and monitoring path requirements for normal multitracking techniques. If the architecture is well thought out, it is operationally rare to need to listen to anything other than the main stereo bus output; this output serves as both the multitrack monitoring bus and the stereo mixdown bus.

Three distinct types of monitoring activities evolve in multitrack work:

1. Mainline—The stereo bus encompasses the multitrack recorder sources/returns and stereo mixdown. This can be read as surround bus if appropriate.
2. Transient—This allows short-term check listening of individual channels for reassurance or adjustment, using PFL or solo functions.
3. Auxiliary—This provides access to the assorted foldback/effect feeds, effect returns, mastering machine, and subsidiary machine returns.

From an operating point of view, the foregoing activities seem to form natural divisions. From a technical stance, it's a different matter entirely. The solo (in-place monitoring) function is very closely related to the stereo bus. In fact, it uses exactly the same signal path throughout—and can be seen simply as a modified use of it. PFL, though, despite a similar operation (only prefade as opposed to post pan listening), actually requires an entirely separate bus and mixing system. Its output is switched to override the main path into the monitors. (It may seem strange to go through all this for a spot-check function that tells less than the stereo in-place solo, until it is remembered that a solo disrupts the mix while a PFL is nondestructive.) Conversely, an operator usually has a psychological hook about the main stereo bus monitoring being the gospel unblemished signal path and that all the auxiliary functions are somehow less polished and somehow tainted. In reality, the monitoring chain normally selects directly between all its sources, merely treating the stereo mix as one of the many. No special treatment is desired or given.

29.14.1 Solo, Solo-Free, and Prefade Listen

An assumption is made that the solo function is such that if a console channel is soloed, all other sources contributing to the main stereo bus are muted, leaving the desired channel in isolation at its set level and panned position. An exception and extension to this are for other channels (principally those returning effects to which our soloed channel may be contributing) to remain unmuted in the stereo mix during solo operation; this is done by using the solo-free button on those channels still needed. Solo-free detaches the channel from the console wide muting/solo activation logic.

Soloing individual channels wet, i.e., with all its attendant effects, is a common need; at a stage in a production where things are dripping in reverb and sundry funny noises, soloing in context only makes sense—by that time it is well known and redundant what something sounds like dry. A channel's sound has become an amalgam of the source and applied effects, not just that of the source.

The upshot of this is that solo monitoring is inherent to the stereo mix path. If that path isn't selected for monitoring, then neither is the solo. So, although a solo overrides the main stereo mix (unless disabled altogether by a master function, solo safe), it cannot override anything else, unlike the PFL.

Although PFL could just be brought up as another monitored source, it is made to emulate solo in single-button touch operation, with the added advantageous capability of overriding everything—whatever is selected to monitoring. Hit a PFL button anywhere on the console and, if desired, it will be what you hear in the monitors. Alternatively it can be arranged to just come up on headphones or a "near field loudspeaker" so as not to disturb the main monitors.

29.14.2 Monitoring Controls

Now we've worked out how to get what signal and at what priority into the monitoring chain. What other torture do we put it through?

1. Level control, which is used to adjust the volume. Usually a big knob or a fader of its own. The most used control on any console —just ask any console manufacturer's service department.
2. Mute is used to turn the row off occasionally.
3. Dim is used so that you can hear what people say.
4. Mono is still used in radio and TV.
5. Phase reverse is used to make sure you haven't already done it inadvertently. (This function together with the mono button makes for one of the quickest ways in history of lining up analog tape machine azimuth.)
6. Split is unashamedly borrowed from broadcast monitoring

technology. This routes a mono sum of the main stereo mix bus continually to the left side of the monitor chain and a mono sum of whatever source is selected (including PFL override) to the right side, providing simultaneous monitoring of two different sources—one of which would almost certainly be console output anyway. (Split's origins lie in network radio, where announcers on the air have to talk up to program junctions and smoothly hand over to another studio or network feed, news, or whatever at a cue. In order to do this, they have to be able to hear both themselves and the network they are opting into to hear the lead-up and handover cue.) Other than its primary design use, the split function is used considerably under other normal programming, affording random source monitoring without losing track of what the main console output is doing. It's also used extensively in program prerecording and production, enabling, with practice, real-time multisource edits (jump edits). Split can fulfill the requirement for single-loudspeaker mono monitoring, by simply selecting the right side to a dead source.

7. Desktop loudspeakers, or idiot speakers, are used to do transistor-radio and cheap hi-fi impersonations, also affording a respite of sorts from the sometimes wearing grandiosity of normal monitor loudspeakers.
8. Near-field loudspeakers (relatively small speakers, usually perched on the console's meter bridge) are used as a twofold reality check during mix: they are close enough to the engineer for the room acoustics to be unimportant; they are closer in size/quality to what the majority of listeners will be using. Often, they are used as the prime monitoring with big monitors and idiots being used briefly to make sure nothing's gone amiss

or become overblown.

29.14.3 Related Crosstalk

In a program sense, two forms of crosstalk are relevant. The first, related crosstalk, is a signal bleeding over into another signal path that is carrying a musically and temporally related signal, e.g., between the left and right of a stereo pair or between adjacent tracks of a multitrack recorder. It happens quite often and is fortunately not often subjectively obvious or embarrassing; usually they're playing the same song!

Crosstalk within multitrack recording systems is usually little short of horrifying. As a result of the large physical size of the console, ground paths are unavoidably long and ground currents generate (and cross-inject into other paths) crosstalk voltages across the resultant ground impedances. Capacitance between interconnecting cabling, looms, modules, buses, indeed everything, results in a reasonably suspect electrical overall crosstalk performance. Naturally, the better the design and construction, the better a console tends to be in this respect. One typically gets what one pays for in this regard.

This was overshadowed and mitigated by analog multitrack tape machine crosstalk between tracks—a safe order of magnitude worse than even a horrid console could ever be. These tape machines not only had the same electrical problems as consoles but also had many magnetic heads in very close proximity, all dealing with a tape medium not notable for magnetic isolation anyway. It was all tolerable and usable simply because all the crosstalk was related and blended in unnoticeably.

29.14.4 Unrelated Crosstalk

Unrelated crosstalk is the clashing and cross-bleeding of signals that have nothing whatsoever to do with each other and are a mutual embarrassment.

In console monitoring a hostile signal, i.e. a delayed replay B check of a master, can be screaming about in uncomfortable proximity to the main stereo mix paths. Broadcasters face this same problem all the time. All their sources are hostile unless brought up on air.

This is unrelated crosstalk, where the bleeding signal is totally dissimilar and irrelevant to the interfered signal. Basically, if any unrelated crosstalk is audible above system background noise, it will be noticed.

An insidious sort of unrelated crosstalk came in the forms of assorted chirps, buzzes, and sizzles stemming from the relentless march of digits into console design and operations. The Society of Motion Picture and Television Engineers (SMPTE) time codes and automation codes were bad enough, but trying to get computer clock droning and computer display squeaks out of the mixing buses and audio paths is not one of life's most enjoyable tasks.

Designing it out in the first place is the only way to deal with computer noise:

1. Make sure all the logic grounds and analog grounds interrelation makes sense or are tailed back separately and never meet.
2. Scrutinize printed circuit layouts to make sure there are no digital signals adjacent to or on the direct opposite side of the board to any analog signal.

3. Intersperse lots of ground traces, and spill entire ground plane layers in the board.
4. Screen high-current high-speed digital signals.
5. Try to allow only static digital control lines onto analog boards—this means decoding digital buses elsewhere other than on audio boards.
6. Ground-plane everywhere there is board space, fill supply and ground layers fully.
7. Choose logic families—or at least interface devices—that are low current and devoid of large power-rail gulps. CMOS is just fine.
8. Decouple everything for all signals—decouple digital for AF and analog for RF.
9. Work on your karma.

29.14.5 Quantifying Crosstalk

"If you can hear it or measure it, it's failed." Such is the empirical crosstalk test. A more formal test was originally the test for interchannel crosstalk, i.e., between any channels in a console; it's also used for any dissimilar path crosstalk measurements. In short, it asks for better than 60dB of isolation of 6 kHz between the paths, measured with a standard peak program meter (PPM) with a CCIR 468 weighting filter in line. Since this CCIR curve has 12dB of gain at its crest (at 6kHz), the specification is actually calling for better than 72dB of isolation at 6kHz, which is neither easy nor often realistic. Such a figure is occasionally not far above system noise floors. Remember, it's a peak measurement; an rms measurement would be some 7–10dB lower. Nobody said it was going to be easy. Crosstalk's a tough problem.

29.14.6 Meters

Some indication to the operator of the signal levels running through the console and, most importantly, the levels that are being sent to other places is necessary. In Fig. 29-88 a pair of level meter feeds are taken from the top of the dim switches; thus, they follow monitoring. A further pair permanently hung across the main stereo mix output is optional. It's customary to provide metering facilities on each channel; in this design the feed is taken following the monitor path source/return switching. This allows level indication of what is going to a tape track during recording and an "all is well" playback display. Gazing at a row of meters hanging off a multitrack playback, it's surprisingly easy to tell what each is indicating. This is an important cue to a mixing engineer.

There are plenty of proprietary meters of the popular standards and types, plus quite a few strange ones, too. It's all a matter of personal preference and the information hopefully gleaned from the assorted needles, lights and blinkies dancing before the eyes.

Without jumping into the argument of average versus peak-reading instruments, it is relevant to state that the choice will directly affect the operational levels, the level architecture, the recorder nominal level lineups, and the various tweaks, such as the input stage limiter threshold in this design. In digital world, reference level is now 0dBfs, or clipping, rather a change from old-school references typically 20dB below that.

It is strongly recommended to refer to Chapter 30 *VI Meters and Devices*.

29.15 Multitrack Console Considerations

Elsewhere in this chapter, versions of practically every kind of electronic subsystem that finds its way into today's analog mixing consoles has been described, explored, and analyzed. Here is an outline of how they are variously stitched together into a practical recording console.

A system can be defined as a means of reducing the versatility of its component parts. Ideally, there should be no system, but practicality dictates that there must be one. The thought is mortifying: hundreds of elements, the microphone amplifiers, differential input amplifiers, line amplifiers, equalizers, filters, and routing matrices roaming loose and needing to be coupled together for each individual operational requirement.

We need a saving grace, and fortunately there is one. Engineering and balancing habits are pretty well entrenched, giving rise to a few well-defined, commonly used elemental combinations. Rationalizing these combinations and arranging easy selection of them as necessary is a good compromise. We've not so much lost versatility as gained a family of operating modes.

29.15.1 Inline Function Modes

A simplified representation of the four basic channel operating modes found in an inline recording console is given in Fig. 29-89A for recording, Fig. 29-89B for mixdown/direct to stereo, and Fig. 29-89C for overdubbing. The Xs show the switching points. Briefly, main multitrack operating modes and their implementation in this system are outlined here.

29.15.2 Recording Mode

In the recording mode, the object is to get a live source, e.g., microphone, through the signal modification chain (i.e., limiting, equalization) and on to a track or tracks of the multitrack machine. Level control on this path is by the main fader (or VCA fader if automation is applicable). Before and after monitoring of the tape track dedicated to the channel is routed onto the main stereo monitoring/mix-bus via the secondary level control.

29.15.3 Mixdown Mode

The machine return is brought through the modification chain and mixed onto the main stereo monitoring/mix-bus via the main/VCA fader. The machine monitoring chain is disabled.

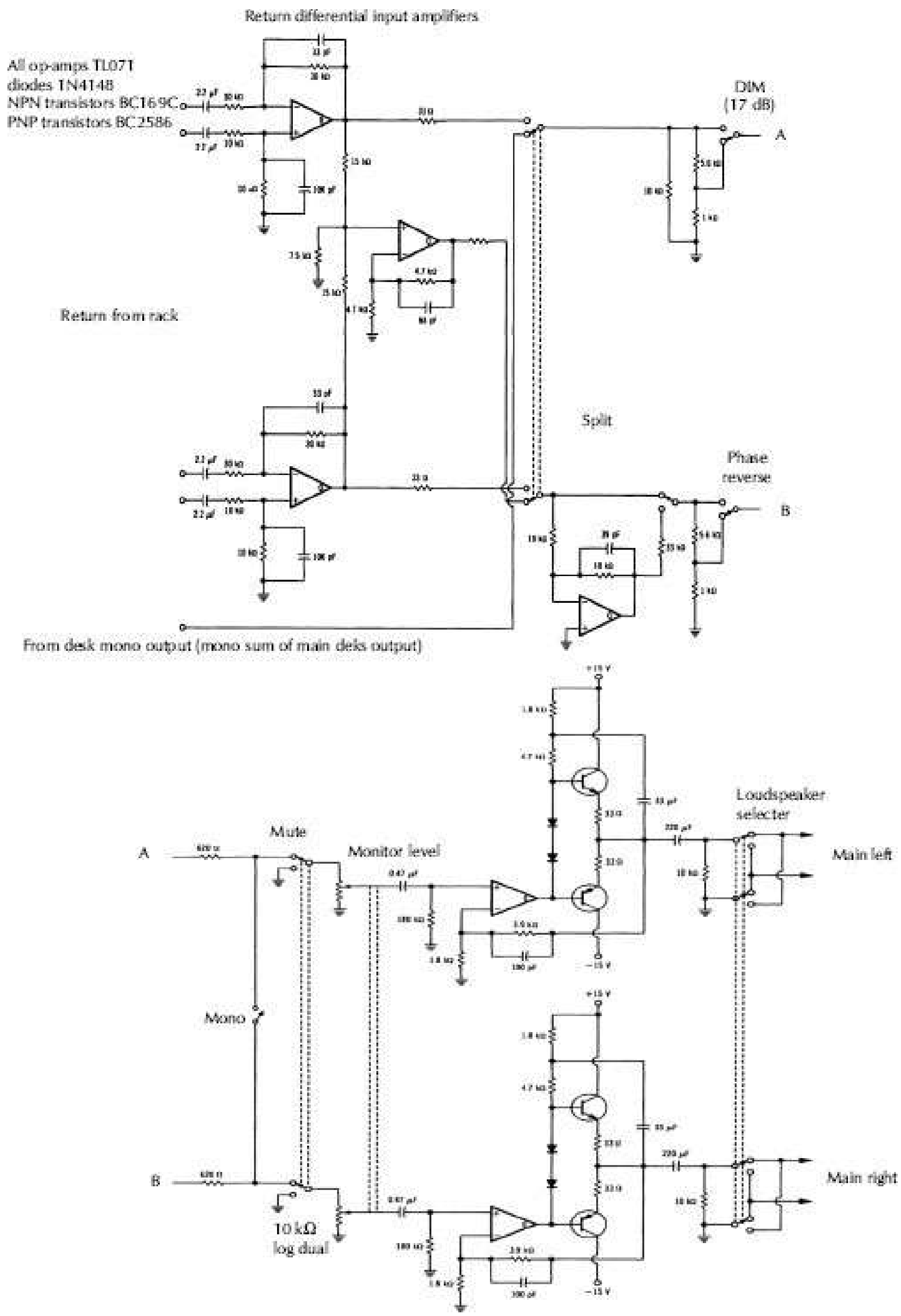


Figure 29-88. Monitor audio path (desk end).

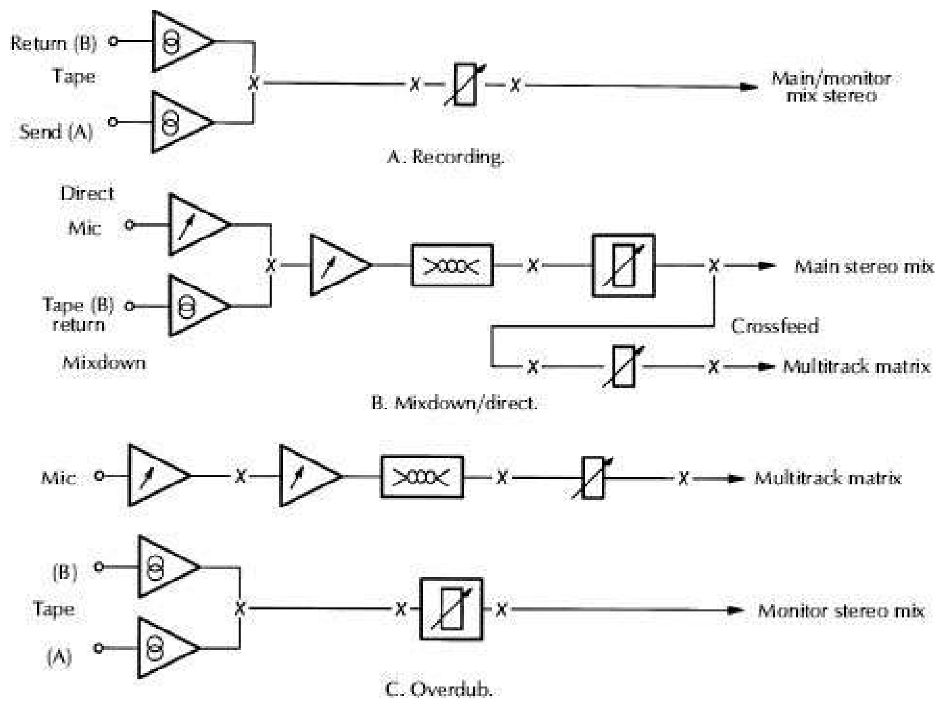


Figure 29-89. Channel function modes.

Since a major justification for keeping the multitrack routing open during mixdown is to provide additional effects feeds, this is best served if the secondary level control is fed post main fader and post mute/solo switching. To enable this, a crossfeed electronic routing is included in Fig. 29-89B. However, independent control is restored when required if a fader reverse is called.

Another mode, direct to stereo, is a derivative of mixdown. It enables live sources to be mixed straight on to the main stereo bus, obviating the need to use multitrack routing.

29.15.4 Overdub Mode

A halfway house between record and mixdown, the overdub mode is intended for use when most of the console is in mixdown but individual channels are being laid or touched up. The signal flow is the same as in the record mode, only with the main/VCA and secondary level controls interchanged. The main/VCA fader in this mode, therefore, controls the monitor feed into the main stereo mix-bus, which ties in with the operation of this fader on all the other channels that are in mixdown.

A handy interlock exists in this mode to facilitate single button drop in. When the channel system function is selected to overdub and the monitoring path is set to A check (machine input), a relay closing pair is made that may be plumbed into the remote control access of the machine. Provided the track is armed ready, hitting A check automatically drops the machine into record simultaneously.

29.15.5 System Level Architecture

Nonunity-level architectures in analog consoles are regrettably necessary under some conditions—detailed here are ways (quite typical fixes within most console designs) that are directly applicable to this described console.

Given standard +4 dBm referred VU meters, under normal operational circumstances, head room in any console is perilously skinny. Various ways of dealing with potentially inadequate head room are in use, Fig. 29-90. A favorite is to run the entire console system at a depressed level, usually −4dB, the necessary 4dB makeup at the end being done passively by an output transformer ratio stepup. This is a poor choice for two reasons. The transformer stepup arrangement is overly critical to termination impedance, and the frequency response could suffer with a heavily reactive load such as a long line.

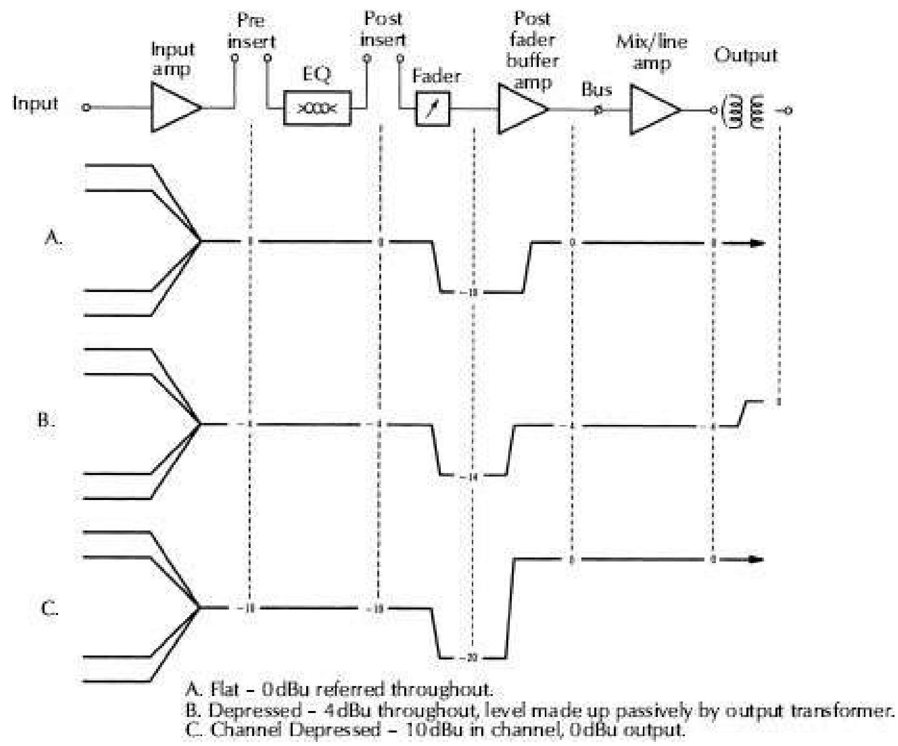


Figure 29-90. System level architecture.

A more modern solution on a similar theme is to adopt a depressed level of −6dBu and make up the level at the output in a quasi-balanced electronic output stage—in this way head room is not compromised at any point along the way.

Headroom is mostly a problem in input channels, before the channel gain-controlling element, the fader. Both ragged unpredictable input sources and equalizer gain gobble up the nonmargin. Hopefully, beyond that point the levels and, hence, the mix are easily and well regulated by the faders. Dropping the channel operating level by 6dB or 10dB helps matters tremendously, and the gain is made up either in the mix-amps or the postfader buffer amps (the latter being normal). This does compromise bus noise (quiescent console output noise), but since the main justification for doing it is the high level of signals present, the pluses outweigh the minuses. This depressed channel system is worthwhile in any circumstance, regardless of metering type, where there is likely to be a great unknown lurking on the end of an input line.

Some of the disadvantages are that all the channel insert points operate at the depressed (10dBu) level, which may or may not give problems in some less than versatile outboard devices. The more immediate concern is that other internal channel circuits will need adjusting.

Ultimately, it is up to the designer to make the product—this mixer—as transparent and free for the operators as possible. Messing around with system headrooms does somewhat fall into the category of kludge, but as a means to the end of facilitating the music-making process less painful, who can argue? A console is after all a creative tool, not a museum of technical operating standards.

29.16 Digital Consoles, DAWs, and Computers

Varying levels of logical control, remote control, automation, and data storage and recall were common on analog consoles long before digital consoles came along, on which of course there is no option. The "Why?" of such digital control of analog was a much harder question to answer at the time, but decent solutions to most of the requirements were found.

Automated fader systems led to the quest for storing and recalling all console functions, since the instantaneous storage, recall, and automation of console settings had immediate application in several spheres of activity. But by the time this was all achievable with a comparable level of sonic performance as from a nonautomated console, digital mixers were becoming a reality.

Since they were of necessity completely programmable, and the control hardware and software was by its very nature already there and in place to do automation to some degree or other, it was a done deal.

What emerged out of the growth of digital control was the fascinating question of control-surface ergonomics; far from being shackled to the hardware beneath them, the surfaces could actually be designed to best suit their purpose.

A further set of considerations comes from the wisdom, necessity, and/or desirability of siting the guts of the console (the signal-processing bit) remote from the control surface. Apart from the need for extensive communication between the two (usually attacked by networking-think) the effect on the design of the console architecture is actually surprisingly minor, and the impacts such as they are will be dealt with piecemeal as required. No, the control surface became the real battleground.

29.16.1 Fader Automation

The first victims of automation were the faders. Once heavy multitrack (16/24 track) had become commonplace, a severely limiting factor of human physiology—only ten fingers—proved something of an obstacle in a mixdown situation demanding considerably in excess of that number. The hitherto classic solution—reduction mixes of subgroups of tracks to a more manageable quantity—forced another tape generation; not optimal considering one of multitrack's touted advantages was freedom from bouncing.

To be able to remember, and subsequently modify if need be, fader movements during a mix seemed like a good idea. There were, and still are, two fundamental approaches to this requirement:

1. Remember the physical position of the fader and on recall arrange for it to move physically to its required position.

   This first technique was introduced initially by one major manufacturer (Neve's NECAM system) and with the availability of reasonably economical motorized faders is now widespread, even on low-cost systems. Most others fall broadly into the second camp. Moving fader systems are dearly loved by their users because of their unequivocal indication at all times—by the actual physical fader positions—of what the system is actually doing. It has one other major benefit—the involuntary hysterical laughter it spontaneously generates from anyone who for the first time sees a swath of motor-driven faders dancing about on their own.

   With the ready availability of such moving faders at cost points suited to nearly every level of application, moving fader automation systems have become the de facto standard for both automated analog and digital mixers.

2. Drive a voltage-controlled amplifier (VCA) from the fader and on recall reapply the appropriate control voltage to the VCA—the fader itself is not then controlling the VCA, Fig. 29-91.

   VCA systems remain viable in analog consoles, though, since they offer advantages at that crucial fader point that moving faders cannot alone fulfill. Although VCA automation systems were once implemented in a purely analog fashion, the fader position values being stored by a PWM or voltage-to-frequency conversion methodology on an analog tape track, these techniques mercifully gave way to digital manipulation and storage as soon as it was practicable.

   A nulling indicator, as described later, is usually employed to match actual VCA gains to that notionally indicated by the fader.

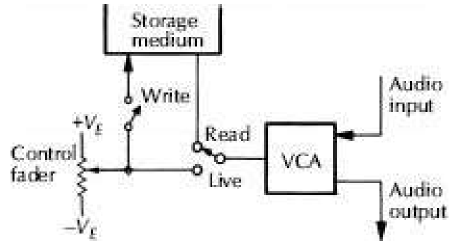


Figure 29-91. Simplistic VCA-type fader automation.

29.16.2 VCAs

Several functions in mixing consoles cry out for a perfect and consistent controllable gain block. In addition to automated fader systems, dynamics control and other analog-controlled gain stages could all benefit by something that looks like Fig. 29-92. It is a black box to which audio is applied, from which audio is extracted, and a control port that determines how much audio is passed. Ideally the law of the control signal should be predictable and consistent. No biasing, no tweaks, no singing, no dancing. Should be easy, right?

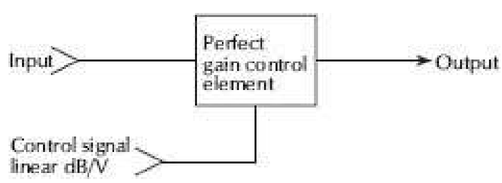


Figure 29-92. Ideal gain control.

As seen elsewhere, raw active electronic devices can be used as gain-variable stages with varying degrees of success, compromises, and weirdnesses; their limitations are various but notably include limited audio signal handling capability, high distortion, and often nonlinear (or nonsensible) control-voltage laws. In feedback-style automatic gain-reduction circuits such as compressors and limiters, law nonlinearities tend to disappear within the servo-loop feedback and have either negligible or, better yet, an interesting effect on the behavior of the circuit in response to stimulation. Effectively biased —to avoid their turned-off regions—FETs can have a square law response, transistors an exponential (logarithmic) response of collector current with respect to applied base voltage. Logarithmic?–dBs are logarithmic!

29.16.2.1 Transistor Junctions

The departure point for the journey to VCA-dom is Fig. 29-93A, which is for our immediate purpose actually pretty useless but elsewhere is known as a cascode amplifier. The upper transistor's emitter serves as a non-voltage varying load to the lower transistor, allowing it to achieve large bandwidth current gain free of Miller effect; the upper transistor (as essentially a common-base amplifier) has no current gain but serves to buffer the load in its collector from the lower transistor, which is busy doing all the work. Varying the base voltage of the upper transistor has little effect on anything other than altering the maximum voltage swing capability on the load, certainly not gain, which is all very much different from the long-tailed pair of Fig. 29-93B. Note that the upper stage is and can be used differentially, as is the output, but it works single-ended too. Here the current through the load is modified by signals applied to either or both upper or lower stage transistor bases. The overall current through the arrangement is set by the lower transistor, which is shared by the upper two; assuming both of the upper two transistor's bases are held at the same voltage, the currents will be shared equally; if one is raised with respect to the other though, its share of the current will rise, having stolen it from the other and vice versa (the total current stays the same). So, wobbling the lower transistor's base will change the overall current, an upper's base that in both upper transistors, complementarily, the combined effect is multiplicative gain variation. (Conveniently, one of the signals [usually the audio] can be applied to and recovered from the pair of upper transistors differentially, although it is not unusual for them to be driven one-sided, the opposing base grounded.) The one remaining drawback is that the operating points of all the devices are moving around in accord with the control voltage applied to the base of the lower transistor and so the control voltage unavoidably appears as part and parcel of the derived output signal.

29.16.2.2 Gilbert Cell

Fig. 29-93C shows what is the essential heart of a good VCA—two long-tailed pairs back-to-back. Actually it's a bit more like three; a long-tailed pair with a long-tailed pair in each output leg. So universal is this basic configuration that it has become the Hoover of VCAs—it is what springs to mind when VCA is mentioned; variations and extensions to this theme are used extensively. Called variously the Gilbert cell or, by RF guys, a double-balanced modulator its main attributes are the innate cancellation in the output of both the applied signal and of the control voltage (CV); all that appears at the output is the product of the applied signal and CV. Product implies it is the result of multiplication, which it indeed is. The circuit is the basis of a good if hardly perfect analog multiplier. Better yet, since it uses good old transistors with their exponential base-voltage to collector-current response, the control law is for the large part linear with respect to decibels of gain and attenuation, reasonably predictably and repeatably. Which is why all the bother and complication is worthwhile.

29.16.2.3 Log-Antilog Cell

A different approach, resulting in a different internal topology is what could be called the log-anti-log approach and is schematically described for simplicity (although the actual integrated implementation is nothing like this) in Fig. 29-94.

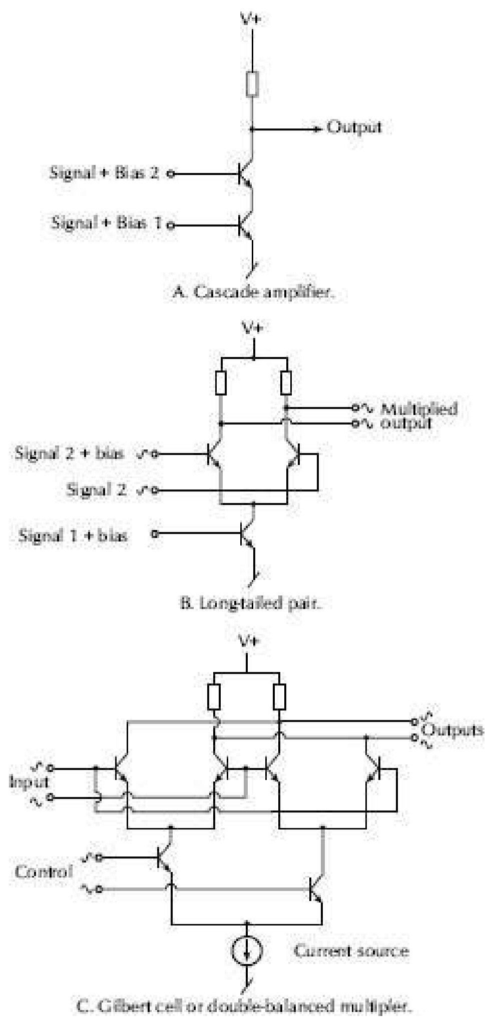


Figure 29-93. VCA design.

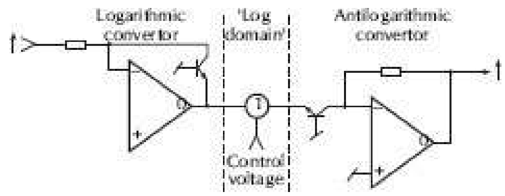


Figure 29-94. Log-Antilog VCA principle.

It again relies on the exponential relationship between the base voltage and collector current of the transistor. The first stage is a log convertor, converting the (positive-going in this example) input signal into a (negative-going) logarithmically representative voltage; summed in with this is the control voltage, which, since we're in log domain, is linear voltage per dB; the composite is then antilogged back into the real world. The effect of adding or subtracting the control voltage is to increase or decrease the linear end-to-end gain.

29.16.2.4 Commercial VCAs

Commercial IC VCAs typically use one or other of these approaches; VCAs are almost always acquired and used in IC form. One built out of discretes will work, of course, but the inherently much closer matching of active semiconductors on the same substrate reduces much matching of parts and tweaking out of various offsets, and the manufacturers have gone to the bother of thermally compensating and biasing everything up such that it "plays nice" in the real world. Nevertheless, for optimum operation of this arrangement or circuits based on it, preset adjustments are the norm, even for integrated versions. Stabilization of operation against temperature is a further complication, if perhaps less so for consoles that will spend their lives in air-conditioned environments.

Beyond the basic and remarkably well-performing circuit element, there are, of course, other issues that come along with real live electronics. A prime consideration is noise; the best operating points for the transistors vary depending on the parameter that needs to be optimal (see the discussion on microphone amplifiers for collector-current versus noise); what may be right for noise almost certainly isn't right for adequate large-signal handling. Attempts have been made to provide for both by altering the bias point of the transistors dynamically in accord with applied signals, such that they're closer to right for both the low-level region (noise) and high-level operation. Another approach has been to parallel up many of the IC VCAs in order to improve the combined devices' noise-voltage to noise-current ratios to improve noise and better suit operation at ordinary audio signal impedances and levels (this goes hand-in-hand with paralleling or using multiparallel input transistors to optimize OSI in transformerless microphone amplifiers).

Input buffering and conditioning of the control signals makes them easy to use; already quite linear, the linearity can be extended over a greater control range and can be arranged to be changed at so many dB per volt of control signal (say, 20dB/V) as to be convenient with the A/D and D/A converters in an automation system and the voltage swing off dc-driven faders. Typically, though, the control port sensitivity on integrated VCAs can be much higher than this—a few mV per dB—and needs to be treated with respect.

29.16.2.5 Control Voltage Noise

This discussion highlights a crucial design issue—that of ensuring a very quiet control signal. This might seem an odd concern until one realizes that at typical control sensitivities, mere mV of undesired ripple or noise on that control line will modulate the through audio noticeably. Note modulate. Since the balanced modulator that is the VCA will not permit control voltage itself into the output, a real audio signal has to be passing through the VCA for this modulation to take place. This, more than anything, is the underlying cause of VCAs' largely undeserved reputation for sounding dirty. Like all these kinds of aspersions, there is a germ of a reality behind them. And this one is CV noise and crosstalk. Any circuitry involved in the CV should be handled with the care one would apply to the "real" signal path; there is a very real temptation to be more casual (and cheaper) with control stuff that should be avoided.

A less-than-obvious concern springs from the fact that the audio path and the control path are not only cross-disciplinary, but are architecturally dissimilar. Audio paths are (assuming mixer channels) following the signal flow, as are their grounds, while the control voltages for a number of, if not all, channels are being handled en masse and distributed star fashion. If ever there was an inadvertent recipe for a ground-induced noise problem, this is it. If the CV is referenced to a ground that is moving in any fashion at all in relation to the audio ground at the VCA, then that difference is effectively added to the CV as far as the VCA is concerned creating noise modulation.

29.16.2.6 VCA Channel Application

Fig. 29-95 shows a typical implementation of a high-end integrated VCA.

The THAT Corporation ("son of dBx") VCA-type 2180 is a current-in, current-out device for audio, hence, a standard current-to-voltage convertor using a good bipolar op-amp following. Note also a seeming overkill op-amp on the control-voltage summer. The control port is where things get interesting. The control feed to the VCA can be a summed combination of several different sources (a quick point here—since VCA control voltages are logarithmic, adding voltages results in multiplication in the VCA, or in other words, the dBs represented by the voltages add or subtract):

1. The channel fader, only it isn't, really. It's actually the output of a D/A convertor that is either reflecting the fader position as sensed by an A/D convertor, or replaying a prior fader position from the automation system. But, for now, we'll call it the fader.
2. Gain-reduction control from the channel dynamics. It is common to use the high-quality fader VCA as the gain-control element for on-channel dynamics. It presupposes the dynamics have deterministic feed-forward detectors and conditioners. Obviously, a feedback-style compressor could not use this VCA.
3. VCA subgroups. A common feature on sound-reinforcement consoles, these are controlled by a central set of a number of VCA group master faders (eight is common). These generate a control voltage, each of which is bused up and down the length of the console; each channel has the option of selecting one (or more, if one likes danger in one's life) of these voltages to be summed into its own VCA. This is a very convenient manner of grouping related channels under a control without having to create real audio subgroups.
4. VCA master. Again, a centralized fader only operating as an overall master over all channels contributing to the main mix bus. Although seeming to be redundant being that there is almost certainly a real audio fader on the mix bus output, a VCA

Master has the advantage that all the levels of sources contributing to a mix can be adjusted, rather than the output of the mix stage. Helps avoid headroom problems in the mix stage.

It should be stressed that the 5 V supply for the 0 to 5 V control signals should be oppressively regulated and fabulously quiet, squeaky clean. Borrowing some off the nearest micro and hanging 100nF across it doesn’t count—sorry.

In console designs with sophisticated computer control, all but the local channel dynamics control signal are manipulated and summed digitally and this composite result is fed to the channel VCA via a D/A convertor; this dramatically simplifies the multiplicity of summed analog control voltages per channel.

29.16.3 Digitally Controlled Amplifiers

VCAs are not the solution for all variable control in analog circuitry. In order to be driven from a digital control system a D/A convertor output needs to be used to derive an analog control voltage for each VCA. This can get very expensive very quickly. A gain controllable stage that can be more directly connected to the controlling microcontroller is desirable.

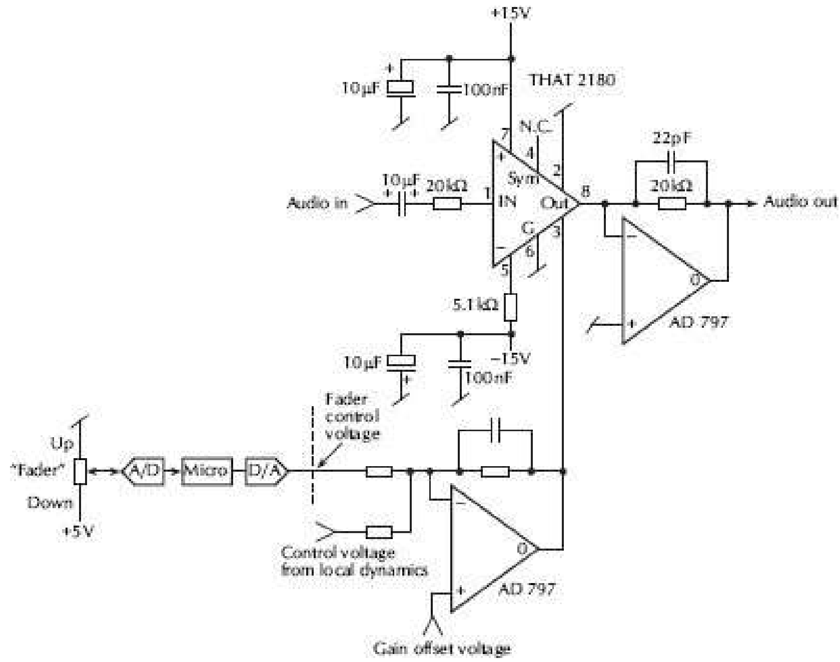


Figure 29-95. Simplified channel-style VCA using a commercial IC.

29.16.3.1 Multiplying DACs

A more direct approach, meaning it can be driven directly off a digital control system, is to use a multiplying digital to analog convertor (MDAC) (don't you just love that "multiplying" bit?)—In particular, a referenced-input four quadrant multiplier, which implies it can produce an output both positive and negative in potential (or current in this particular case) and which is proportional to a voltage applied to its reference terminal, Fig. 29-96. Now, bear in mind these devices were never intended to be used in this way, but luckily they do so well.

The audio signal is applied to the reference pin; a digital number, in this case 12 bits wide serially fed into the device, is applied to the 12 bit R-2R-style ladder DAC, Fig. 29-109; the audio signal is attenuated in proportion to the applied digital number with respect to the 12 bit maximum (4096 steps). The output current is sensed and converted back to a voltage by the following virtual-earth input amplifier, using the friendly internal feedback resistor around the op-amp. The interface is dead simple, linearity is pretty good, the signal handling is excellent, and the noise isn't bad—dancing in the streets!—except every time the gain is changed (a new digital word transferred into it) it makes a little tick noise, which is very audible on high-level signals, low-frequency signals, and especially the combination. In fact, as the gain is moved (a la fader), classic zipper noise is very evident. The only good news about all this is that for a large part, program material's spectral content masks this noise. However, when the device is used as a frequency or Q-determining element in an equalizer, the effect becomes comical; depending on one's sense of humor. There are two approaches necessary to nail this noise, since it is actually due to two separate causes.

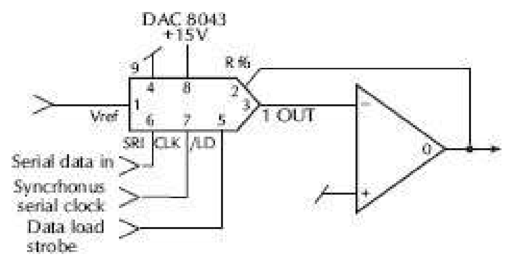


Figure 29-96. MDAC as rudimentary gain control.

29.16.3.2 Charge Injection

This is a near unavoidable effect in CMOS and other electronic switches, where a tiny amount of differentiated charge impinges itself onto the signal path from transitions of the control port. In a multiplying DAC, any number of bits may be changing as the gain is varied, and so the total charge injection varies correspondingly. It is, however, almost completely independent of the applied audio signal.

Cancellation works well, with reservations. One approach is simply to use a second MDAC with its own inverter that sums into the virtual-earth point of the main MDAC path, with its reference pin undriven. However, with only slightly more complexity the arrangement as in Fig. 29-97 emerges, which is our old friend the Superbal differential summing arrangement fed by both DACs being driven differentially. Not only does this provide a gain-control stage with full-rail differential signal-handling capability, but also the charge-injection noise is substantially canceled. To get the best noise cancellation, however, the DACs really need to be matched (a DCR test through the ladder is a reasonable guide for matching) or pairs on the same substrate employed.

29.16.3.3 Zero-Crossing

The second impulse noise cause is attempting to switch a high-value signal; any truncated or very rapidly level-shifted high-level signal is going to go “click!” (Run tone through a switch and turn it on and off a few times. The switch click—nicknamed tone-click—will vary in intensity seemingly at random; that’s because the switching is occurring at random points through the sine wave’s cycle. Those at or near the crests of the sine wave will click loudest.) The simple solution is, don’t. If one arranges only to change gain while the applied signal is crossing through zero or is at a low level, this manifestation will all but disappear. Zero-crossing detection is standard on gain-variant devices such as digitally-controlled microphone preamplifiers used in today’s consoles, external/stand-alone I/O (Input/Output) interfaces, and audio processors. The techniques, now integrated into converters and such, are very similar to that described here for internal console design.

“Tone Click” reduction is a control issue rather than an audio path issue; Fig. 29-97 illustrates differential MDACs with a peripheral circuit that achieves near zero-crossing. The MDACs in Figs. 29-96 and 29-97 are double-buffered. In other words, it is possible to load a new gain value into them without disturbing the current operational gain and then transfer the new value over when desired by means of the/LD (Load) control pin. The arrangement shown allows the controlling micro to do a “hit and run” on the circuit, depositing the new gain data and telling the circuit to take it at the next zero-crossing; the micro doesn’t have to hang around waiting for a zero-cross to occur. It can be zooming around setting up other MDACs in the meantime or attending to other microlike things. The circuit is addressed, the new gain data serially clocked into the MDACs’ first buffer and then the micro nudges high the ARM line. (The ARM line needs only an instantaneous +5V pulse; positive feedback around the first comparator keeps it set. Likewise, it can if desired be nudged down to dis-ARM—nice, but useless.) Comparators wait for the applied audio signal to fall into a low near-zero signal window, at which point an instantaneous strobe pulse for the/LDs is generated, which latches the new gain data into the MDAC ladder and simultaneously cancels the ARMing.

This arrangement would not be the avenue best traveled for dynamics, being that it takes a comparatively long time to load in data and wait for zero-crossings, limiting apparent responsiveness —VCAs are a far better course for dynamics—but for anything else it works a treat. With reasonably matched 12 bit MDACs this gain-control circuit is virtually transparent and even works well in high-Q filters and EQs. It’s still not inexpensive though, and the dawning realization of exactly how many of these circuits (or DAC/VCAs) would be necessary to fulfill complete automation of a decent size mixing console, and just how much they'd all cost, has quite a stunning effect.

29.16.4 Discrete Logic and Programmable Gate Arrays

The ubiquity of these parts now has led to a hardware design approach that is at once bold yet somewhat alien to those who still remember tape-and-dot layouts; everything on a board, say switches, resolvers, converters, etc., are taken directly to pins on a gate array; the interface to the host microcontroller is brought to the gate array; then how it's all interconnected, strategized, timed, polled, strobed, and indeed processed, becomes a pure (software) programming exercise for the gate array. Errors and changes similarly become just software changes, too, not board re-spins. They have revolutionized console control methodology.

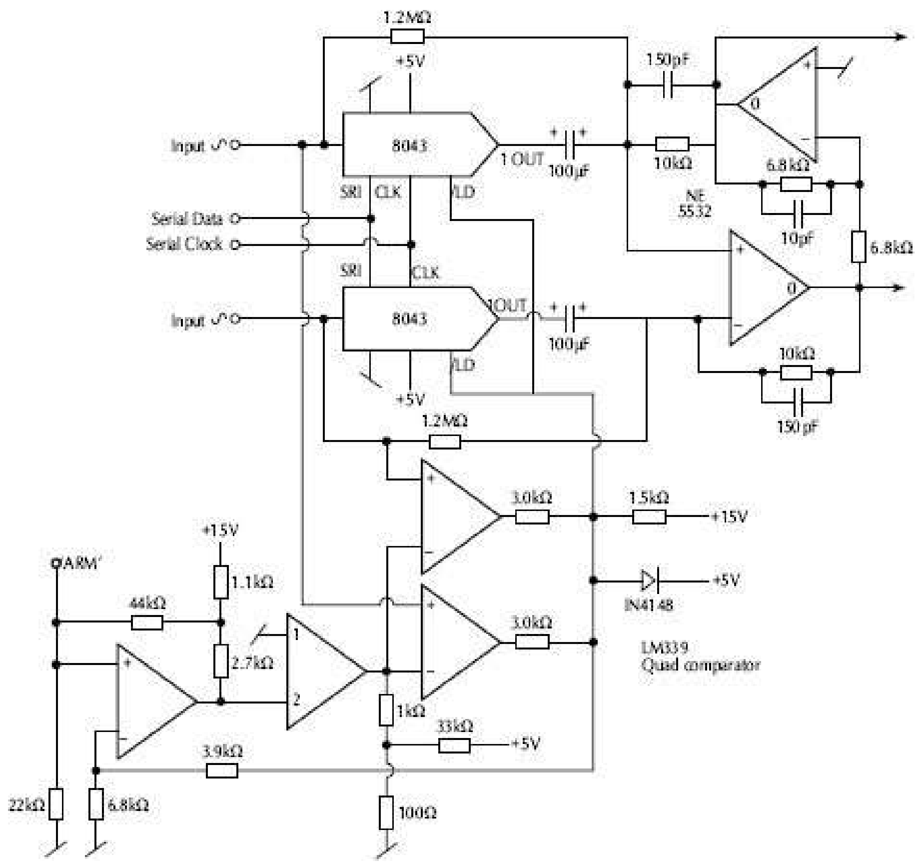


Figure 29-97. Differential MDAC gain control, with zero-crossing enable.

29.16.5 Recall and Reset

Remembering the position of controls in a conventional console was the great innovative burst of the late 1970s. All that was required was a rider pot on the back of variable controls (although this can be a bit difficult with dual-concentric pots) and an extra pair of contacts on switches.

29.16.6 Data Acquisition

Switch closures were polled and sensed in batches of 8 or 16 (dependent on the bus width of the micro in use for the automation) while each individual pot position was resolved to the accuracy afforded by usually an 8 bit analog-to-digital (A/D) converter—256 possible positions. Although high for resolution and practical resetability of most pots, it is actually harder work reducing the capability than leaving it be! This may be true for pots, but with high-quality faders the resolution may feel too coarse and "lumpy"; 12 bit resolution was often necessary. Ground noise reduced the useful resolution to significantly less in most practical systems to a 10-bit equivalence. The micro addresses and runs around the console compiling a data field in memory representing the entire console's control status for that polling period.

Instead of a single micro struggling to do entire surface along with whatever else with which it may be being tasked, with the cost of such devices being pocket change, it is not unrealistic to throw a microcontroller or an FPGA at each channel simply to perform these tasks; a significant reduction in parts can be afforded. The console's main micro then just has to collate the reapings of each channel's micro.

29.16.7 Recall Display

Storage of this control data on a mass medium such as hard disk or network is a fairly simple computer file-management exercise, as is recalling it. What to do with the recalled information is now the question.

It is assumed for the moment that this particular requirement is informational recall only, not hardware reset (i.e., setting up the parameters of the channel to their stored values). Eyeball comparison and human tweaking is the resetting mechanism assumed here. The comparison is between a recalled value displayed on a meter, LED column, bar display, null indicator, or up on a GUI (Graphical User Interface) screen, and the immediate real value read from the control in question and displayed on an adjacent display. As the relevant control is tweaked, its indicated value will be higher or lower than the stored value; when the two are matched, then the control position is the same as it was when the snapshot was taken. Fig. 29-98 shows in simplistic form the basis of the matching process.

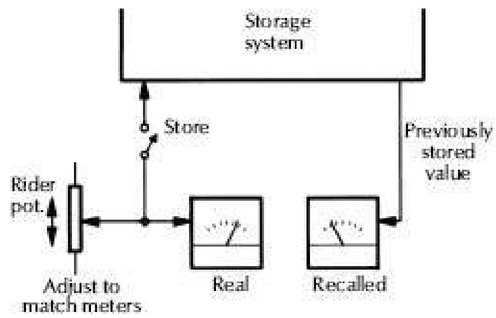


Figure 29-98. Simplistic manual reset system.

GUI displays are presently the easiest way of performing this matching. So much information is visible at once, which is a blessing in this circumstance.

Even with increasingly common totally programmable/recallable consoles, screen-based display of control statuses is a very useful function, if in addition to localized feedback to each individual control. It looks good, too.

29.16.8 Nulling

Null indicators are particularly easy to implement and use. They usually take the form of a pair of LEDs adjacent to the relevant control. If the real value is higher than the recalled value, the upper LED lights; if it is less, the lower one lights. If they both come on, the two values are matched. Even simpler nulling indicators take the form of a single LED that only comes on (or alternatively goes out) when the two values match. A nicer arrangement is a single-cell green/red LED giving an unequivocal “go” or “no go” indication. This device makes it particularly easy to spot anything out of order on a channel.

As laborious as these facilities may be operationally, a complete reset of console parameters can be achieved. It is considerably less laborious and inaccurate than writing everything down. Typically, only really salient settings are reset in this manner, say, the microphone channel settings for a vocalist overdubbing over multiple sessions.

Obviously, manual eyeball resetting is a far cry from the ideal; what might not be apparent here is that ignoring the resettability aspect this now gives the control surface the means to control other equipment, say a DAW’s corresponding features. Oooh…

29.16.9 Resetting Functions

The next logical step in developing computer assistance is for the machine not only to remember console settings but also to reestablish the console to its previous operational state on command. This means that if the multitrack routing on channel 27 was going to machine track 15 when the console status was stored, then regardless of what has happened or how the routing may have altered or configurations changed, upon recall channel 27 will go to track 15.

Every switched function that is intended to reset needs to be made electronically controllable; the techniques are detailed in earlier sections. This replacement, by and large, has already been implemented with other ends in mind, such as simplifying PC layouts, avoiding large physical switches, and, not the least, facilitating some of the tortuous signal rerouting required in a modern production console.

29.16.10 Motorized Pots and Faders

Motorized pots and faders look and feel like conventional ones, only a clutched motor drive controlled by a servo allows the mechanics to be reset to any point on their travel. A rider track, either a normal resistive track encoded by an A/D converter or a digital track direct input allows a microcontroller to keep track of the position. Comparing its present position with one previously stored drives the servo to equalize the two, i.e., return the control to its prior position.

These are extensively used particularly in automated or soft consoles, where one physical control can be responsible for many channels or functions.

29.16.11 Resolvers

Resolvers are continuously rotating (no end stop) controls that otherwise look like a conventional potentiometer. Indication for these is commonly arranged to be a circularly disposed set of LEDs around or within the resolver knob rather than linear, adjacent. Such arrangements of varying degrees of cleverness are a staple of control surfaces nowadays. A resolver, when rotated, sends out two streams of pulses, half overlapping as in Fig. 29-99; in other words, they are 90° out-of-phase or in quadrature. This is enough information to determine not only how fast it is rotating (by counting the number of pulses from one of the trains) but also in which direction. These two, rate and direction sense, are enough for a controlling processor to analyze and appropriately perform control.

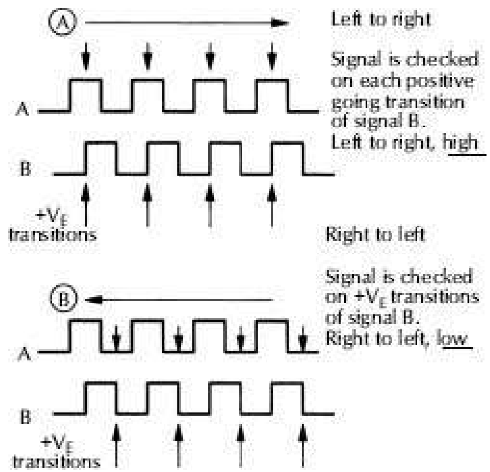


Figure 29-99. Quadrature streams from resolver indicating how speed and directions are sensed.

The simple circuit of Fig. 29-100 sorts it out; it's a 4013 D-type latch. The data port is fed by one train, while the edge-triggered clock input is fed by the other. If the clock is triggered by the rising edge of the A train and the B train is active, then the latch output goes high, indicating one direction of rotation (left to right in Fig. 29-99). In the other direction, the rising clock edge from A corresponds to B being inactive, so the latch output goes low.

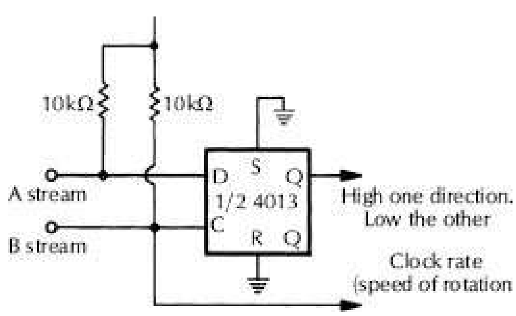


Figure 29-100. Resolver decoder (using a D-type flip-flop).

It is rather a simplistic circuit that assumes that the making contacts of the resolver are perfect and no false triggering will occur. With more swanky optical resolvers this may be true, but with mechanical ones a little debounce clean up prior to the D-latch gates may be advisable.

29.16.12 Control Surfaces

A large problem with recording and live consoles was precisely that—they got large. Console channels grew into long, thin strips for purely historical reasons, and the manufacturing technique of hanging all the signal-path electronics on acres of dense PC card has just tagged along with little evolution. Removing the audio electronics (analog or digital) from the control surface into a remotely controlled equipment rack seemed quite an obvious development, although initially a technically unwieldy one. Many types of analog circuits lend themselves to direct remoting. For example, VCAs for level control need, in essence, a single dc control line. Others, such as equalizers and microphone preamplifiers, don't. Noise and difficulties in extending nonzero-impedance configurations are both significant problems. As with everything else, these areas of difficulty look quite different given a dose of digits. As has been seen, with only minor compromises, digitally controlled remotable audio circuits of all sorts are realizable at some cost and complexity; it is entirely possible for the control surface to become now just that, no audio need go anywhere near it.

The question of whether the control surface and signal processing electronics should be divorced and live in environments possibly better suited to them is a more vital one with digital mixers than with analog; although possible, it was (is) actually very expensive to do remote, fully digitally controlled analog (DCA) circuitry that sounded decent and wasn't riddled with clicks, burps, and fizzies. It rapidly became more expensive than having a fully digital signal path, which has really sorted that argument out once and for all.

29.16.12.1 The Single Channel Concept

There is an immediately apparent redundancy with large traditional consoles—rows and rows of identical channel modules. The first intuitive step would be to reduce all those to just one set of channel controls that is selectable or assignable to any channel that needs tweaking. The first modification to this rather simplistic single-channel console concept is that the main channel level faders need to be kept continuously available in front of the operator; a button adjacent to each of the individual faders (the "ME!" button) calls the set of assignable channel controls to the channel to which that fader is related.

The second modification concerns the assigned controls. They have to be separately acting for indicating. On being called, the indicating part of the control adopts the settings pertinent to that channel; the control, whether it be knob, switch, or fader style, can then act on the selected channel with the indicators following their action on the remote circuitry. A wide and glittering variety of controls have evolved to suit this requirement, but they all basically have a row or concentric ring of indicating lights, fluorescent indicators, or LCD panels disposed around the digital resolver control knob. Alternatives to knobs, switches, and indicators such as interactive GUI screens can suffer from an ergonomic disconnect between the physical operation of the control and from where the relevant feedback is displayed, unless the control becomes a mouse-driven widget on the screen. Touch screens with haptic control are an excellent solution.

A third modification to the initial rationalization concerns the many auxiliary mixes found in a console, whether they be for effect feeds, foldback, multitrack monitoring etc. Although the controls for these are traditionally regarded as channel controls, intuitively they are thought of and operated on horizontally across the console; if someone's setting up a foldback mix they'd most likely be working along the row of controls for that mix bus (to which they'd also almost certainly be listening via monitoring) and have very little interest in any other channel controls at all. Making the operator select each channel at a time to do such a routine mix setup is a very retrograde move—it imposes an unwelcome multi step process that diverts concentration from the task at hand to the means of achieving it. Quite sensibly then, any same-function bus-oriented controls should become accessible together. This is precisely the rationale behind the channel faders all being accessible simultaneously. Ideally, a row of interactive knobs, one per channel across the console, the function of which follows the feature of interest (like that foldback mix) is appealing. Such have been variously called smart bus or virtual controls. If control surface space is a driving issue, a neat bit of further rationalization can come into play here; the consolewide set of controls implied in having parallel access to auxiliary buses (meaning that in addition to a fader for each channel there would be an auxiliary bus control also) can be avoided if really necessary by using the already existent faders. After all, if we're busy setting up an auxiliary mix, we won't be overly concerned about other mixes, including the main one. Even if something does need instant attention, reassigning the faders to main mix is only a button away.

So here is the essence of control surface rationalization. There would be a row of moving faders and possibly a smart knob, one for each channel, with an adjacent control select button (ME!) that renders a singular set of channel controls (whether glass or physical) operative on that particular channel. We would also have a row of buttons (with again possibly GUI supplementation) that selects on which mix-bus(es) the fader row (or additional smart bus) is acting.

Early practical experience with this showed, even with operators who came to grips with the single-channel concept readily, that there should be ideally more than one set of channel controls—it is a common requirement to play two or more channels against each other in a mix. Secondarily it was felt that having a set of controls always set up on the one critical channel in a mix (the money mic) and having one or more floating surfaces perhaps represented a better compromise. This represents the fourth major modification to the single-channel concept, although most rationalized designs still lean to just supplying the one set of channel controls.

The great beauty of making all controls transient (i.e., not totally dedicated to any one channel's function) is that all console functions are implicitly digitally stored, recallable, manipulatable, and automatable.

29.16.12.2 Commercial Console Control Surfaces

There was once a very sensible argument is that there was a very large user base familiar with the ergonomics and accomplished in the use of a traditional knob-per-function analog console, so why on earth force a learning curve on them? In less than a generation, almost all console control surfaces have become rationalized to some extent, and their users regard the old behemoths if not in awe, as quaint at best.

By way of interesting variants, Figs. 29-101 and 29-102 show a world-class analog production console, the SSL AWS, and the SSL Live DSP-based live sound console. Although both can operate independently, they both necessarily live in a DAW world: The analogue console is intended to work just as comfortably hand-in-hand with a DAW and with highly integrated control over it, whilst providing the capacity to 'track' (originate material) in analog, a preferred workflow option by some. The digital console can also readily interface with either a DAW, mostly using it as a show multitrack recorder, or with a stand-alone recorder.

Figure 29-101. SSL AWS – an analog music tracking console. Courtesy Solid State Logic.

Figure 29-102. SSL Live—a digital live-sound console. Courtesy Solid State Logic.

The SSL AWS remains close to a knob-per-function traditional console format. At the other end of the rationalization spectrum is the Innova-Son Compact, shown in Fig. 29-103. This was as close to the single-channel concept as it was possible to get; other than the faders all controls were centralized, following the ME! button's activation on the desired channel. There were some very clever features not obvious from the photograph: all the faders are moving faders; if a channel ME! button is pressed, all the group faders move to represent that channel's feeds to each of those groups. If a group ME! is pressed, all the channel faders move to represent each of the channels' contributions to that group. A very impressive surface and fun to drive.

Figure 29-103. Innova-Son Compact sound-reinforcement console. Courtesy Sennheiser USA.

It is not a surprising leap to see that a rationalized surface such as that can have the input (and output for that matter) channels paged; this means that a switch can instantly throw a whole second (or more) batch of controlled channels up onto the surface. Superficially a great idea, since a modest-sized surface can drive a much larger overall console, this seemingly facile addition is far harder to come to grips with operationally than rationalizing the channels themselves ever was. It is quite unnerving to have half a console disappear! It takes considerable effort to design and engineer a surface with enough clues as to the background channels' existence and well-being to make paging a comfortable operation.

Fig. 29-104 shows a highly considered control surface design somewhere between the two extremes of knob-per-function and completely rationalized. The medium-format Wheatstone D-32 television audio console has centralized EQ, dynamics, routing, surround panning, auxiliary and mix-minus feeds, which are brought into play by, guess what, a ME! button on each relevant channel. Additionally, though, it is to be noted that there remains a considerable amount of localized control on each channel; these controls are what an operator needs to get his or her hands on immediately (and which of course can differ between setup and on-air contexts), with no intervening selection step involved. (Remember, broadcast is a high-stakes no-second-take environment). Input metering is adjacent to each fader, and two sets of channel ID indication above each fader help assuage paging concerns; full console status and metering are spread across the numerous GUI displays in the penthouse.

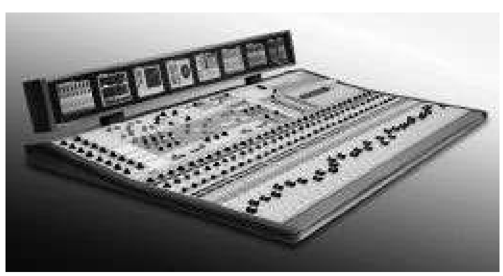


Figure 29-104. Wheatstone D-32 live console. Courtesy Wheatstone Corporation.

There are circumstances where the use of a room might be quite diverse over the course of a workday or likewise the technical adeptness of the users; in mind is that of a radio broadcast studio. One could have the problem of there being a perceived baffling sea of knobs for a disk jockey, yet insufficient control for a commercial producer. A convenient solution, falling out of the soft control surface concept, is shown in Fig. 29-105; the hardware surface is very basic—just what an on-air presenter needs—but the (removable if need be) screen can be ME'd and mouse driven to have a full set of EQ, dynamics, and effects per channel: happy advertising producer. It is even be possible to run the console entirely from the screen, with no regard or need for the hardware surface.

Figure 29-105. Wheatstone Evolution 6 Console. Courtesy Wheatstone Corp.

And so in a few short paragraphs, we've moved from a knob per function to no knobs at all. Already in the game of control surface design sans frontiers, manufacturers are rightly taking the measure of their clientele and producing surfaces much closer to their actual requirements than ever before, liberated by digital control. It is very encouraging. There is no universal perfect control surface solution; seemingly polar protagonists of the knob-per-function and doodle-on-a-screen approaches are equally exactly correct.

29.16.12.3 Control Surface Intelligence

Even if there is no signal processing going on in the same box, the control surface still has an awful lot going on inside it, Fig. 29-106. Typically there will be a large embedded controller, or even a PC-style microcomputer to administer things such as control surface host; it will likely be of the ×86 persuasion, or a capable embedded processor such as an ARM. Divorced from the tyranny of thin module strips, economies of scale come into play: a reasonable number of controls' data may be garnered, and reasonable numbers of indicators driven, without the indirection and bottle-necking encumbrance of a console long data busing scheme. Even a large console's worth of controls and indicators is easy pickings for a large processor treating them as medium-speed peripherals through industry standard buffers or FPGAs. Should a devolved scheme be necessary from a semimodular or macromodular approach to the surface each submodule may be looked after by a smaller embedded micro or even an FPGA, with communication from each back to the host by fast serial link. (On a smaller scale, all the requirements of an entire small broadcast control or recording surface can be handled by a single modest FPGA, radically simplifying interconnection, because there isn't any!)

Chances are, this host will also be feeding data to a subsidiary LCD screen driver processor (or two, or three), talking down Ethernet to the signal-processing host sending it fresh parameters (or even coefficient sets if they're being calculated at the surface end), receiving back from the processing host packets of metering information to be divvied out to the appropriate displays, and last but certainly not least attending to the level of automation (static snapshots or real time) in which the console is operating. To this end, it almost certainly has mass storage, such as loads of flash ram, and/or a hard drive.

29.16.12.4 Multiuser, Multifunction

User arguments can run something like, “But we might want to change several things at once, and Fred the producer likes to look after the monitor mix while I do the rest.” The control software would naturally allow simultaneous control actions on a pair or across a group of channels to be ganged, which is fairly trivial and not the point being addressed. The main engineer console can be regarded and would be regarded by the host computer in console systems as simply a terminal, albeit the main one. There is nothing to stop other terminals of greater, equal but probably lesser or deliberately limited facilities having access to the main body of electronics, sharing the network and its resources, in other words. In practice they would have access to and be able to manipulate a preprogrammed subset of the total capability (e.g., our producer friend’s monitor mix) concurrent to the main terminal or control surface. Another obvious secondary terminal would be a second or even third set of assignable channel controls for multi-op situations, although we can’t help wondering how often they would be redundant except in the all-hands on-deck film mix-down world. As a capability it would go a long way to soothing the frustration of engineers new to the concept who are wary of losing so many controls at once in sacrifice to the new false god rationalization!

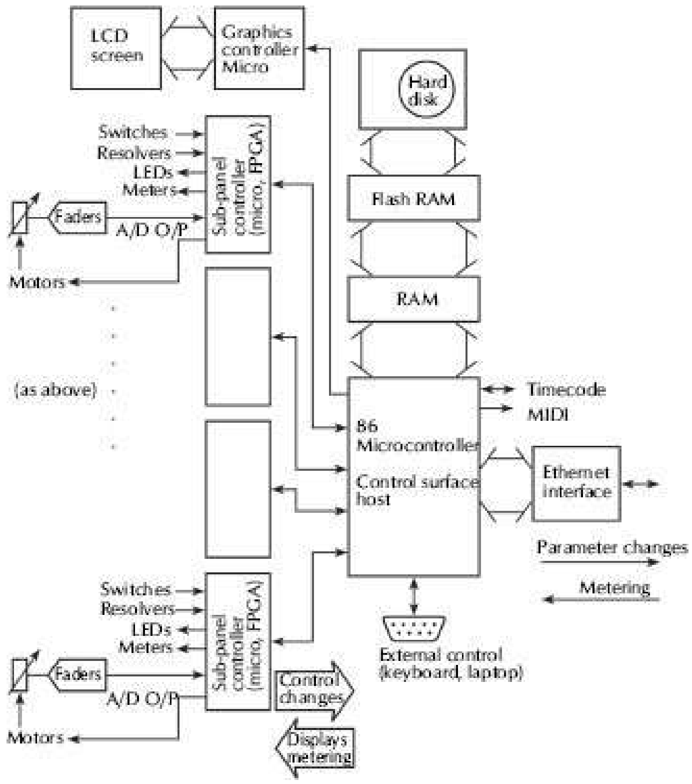


Figure 29-106. Control surface control architecture

Simultaneous access to the same set of information is what the term multiusers is all about. Multiple control surfaces pose no real issues—control systems and networking operate so quickly in relation to the rate of changes a mixing engineer can make that several concurrent operators sense no interaction at all.

In computer terms the system described bears more than a passing resemblance to a hardware-related database, remotely controlled by a terminal or terminals down a network. Again, in computer terms, it's a pretty small database, and at least on the control side a pretty lightly loaded network, too.

29.16.13 Goodbye Jackfields, Hello Routing

Considering that one of the easiest audio subsystems to organize using digital technology is signal switching, it's astonishing jackfields lasted as long as they did. Analog switching matrices are at such a level of development that they can be considered transparent to the system. Digital routers of course have no impact on signal quality whatsoever. Neither in any way, even when many are cascaded, create performance limitations. They are dense (many thousand source/destination crosspoints will fit in the same rack space as 144 jack holes) and decreasingly expensive—much less expensive per crosspoint than a comparable jack circuit. Control is soft and the operation can thus be anything from a humble computer terminal, PC application, to effectively complete seamless integration as part of the console's control surface. Of course, within assignable systems, the matrix is controlled by an interactive control surface by the operator, all routings and parameters being storable, recallable, and resettable as are the rest of the parameters of the console, in real time if desired. Try that with 50 patch cords!

Inputs and outputs of everything internal to the console (equalizers, dynamics sections, front-end amplifiers, line-output amplifiers, and so on) and everything external to the console (effects, machine input and outputs, and so on) all appear as sources or destinations on the matrix. The concept of insert point has disappeared; anything can go in anywhere. After decades of things getting more complex suddenly things have become simple again—there is no system and no prewired interconnections. A system to fit a given circumstance is built up from scratch using all the circuitry building blocks interconnected as required via the matrix. A repertoire of usual starting points—preassembled patches—is stored and recalled as needed.

In pure digital signal-processing systems, this is taken a step further, where processing elements can be arranged in order at will or dropped into place in the form of plug-ins. No arbitrary system: full circle.

29.16.14 Integrated Control of Outboard Gear

A great many bits and pieces of traditionally outboard signal-processing gear (known vernacularly as toys) are involved in the successful production of present-day program material. Already the term outboard is flimsy since via the system matrix, or via plug-ins, their signal paths are already firmly internalized. The old music-industry serial communications link MIDI, despite its limitations, still bears integration into any studio interactive system. The centralized control point for these is the interactive main control surface for the operator, and a MIDI-controller application is required unless the console as a whole (strangely but sometimes quite sensibly) merely becomes a MIDI slave to an external controller. MIDI controller as console?

Why, yes! It is no coincidence that major players in the DAW world, and hence with influence tendril-like into the rest of pro-audio, were initially strongly into the instrument, machine-control, musical synthesis, and arrangement world of MIDI control (e.g., Steinberg, with CuBase); it helps explain, too, why there is such a tight integration of MIDI music-making capability with audio processing in these DAWs, and their look and feel is unmistakably MI (musical instrument) in flavor, owing much less to traditional recording.

At last the impossible studio system of a mere few years ago, integrated, completely automated, and resettable in real time in conjunction with effects, storage machinery, and other systems such as video, is here.

29.17 Digital Consoles, DAWs, and Computers

29.17.1 Digital Audio Workstations (DAWs)

Control by GUI only, where all audio functions are controlled by mouse or haptic touch manipulation of on-screen widgets in the form of pseudo-knobs, buttons, and sliders, has been a natural progression, if for no other reason than it is cheap—there's no physical surface to build or buy! A GUI, though, presupposes that the actual signal processing is already in digitally controlled form, usually pure digital. Although a GUI can be part of an embedded system controlling a traditional digital console as described later, often it is part and parcel, along with the control code and signal processing code, wrapped up within a PC. This does not make it a non-console, all the parts and processes that make up a console are in place, just in the one place, including the recorder!

DAWs rapidly transcended being the dinky two-track editing tools they started out as and have become the de facto console experience in many spheres of audio. Characterized at heart by being (or having at least the appearance of being) software applications that run on the familiar PC or Mac; by absorbing the recording into the PC's hard drive, by providing access to just enough audio signal processing, by rationalizing control extensively so that it fits adequately on a screen, DAWs rule the nonlive and production audio arenas. While many DAWs totally run within and skirt the processing limitations of the host PC (which are becoming less limited as PCs become more capable daily) in some cases extensive additional DSP farms are employed to do the heavy lifting, leaving the PC mostly to do the user interface. In either case, the PC-based DAW is a perfectly valid multitrack production environment. Paradoxically, that which was the DAW's initial strength—the convenience, familiarity, and low cost of the PC environment and GUI—is now the major (there are others) drawback. Screen-based DAWs using point-and-click represent the ultimate in rationalized operation, and do not lend themselves well to other than single-operation-at-a-time usage.

This is the predominant reason simple DAWs are still eschewed in any live audio activity, with more traditional (less rationalized!) console surfaces and processing maintaining favor. That said, there are burgeoning after-market and own-brand control surfaces expressly to augment and improve operation of DAWs, and many traditional console manufacturers have embraced the underlying technologies and merged the two approaches quite seamlessly; these range from a small surface of little faders all the way up to major surfaces such as the SSLs above.

A crucial point is that although DAWs may seem to be a breed of their own, with the exception of very tight integration with the recording medium, they are indeed digital mixing consoles. That said, using DAW software purely for its recording capabilities is common in conjunction with traditional consoles, as is the use of stand-alone recorders which can again integrate seamlessly, using typically MADI or an Audio-over-IP variant to move the audio around. Which all goes to say that assuming the control surface suits the bill, one might never know, or actually care, what the underlying technologies may be!

29.17.2 DSP-based Consoles

It is an impossibility given the nature of this book and the space available to give a thorough treatise on digital mixers and their techniques. It gets pretty mathematical, pretty scary, pretty quickly. What is intended here is an outline of typical audio digital signal-processing considerations, methods, and limitations from an intuitive and practical standpoint and ultimately in the context of a practical digital console design.

An analog versus digital divide still exists simply because as with any pair of such disparate technologies, what is easy in one can be hard in the other and vice versa; digital can do some things that are practically impossible for analog—time-related machinations, for example, which are typically gruesome in analog. Tritely, it used to be said that real-deal EQ and dynamics were the province of analog, being that it has hitherto been easier and cheaper to achieve nice-sounding, complex, and flexible phase and frequency response shaping with a handful of analog components; this has become less of a black-and-white proposition though as the size, speed, and power of digital signal processors have increased and relevant expertise and ears were applied. Very fine digital EQ, dynamics, and effects are indeed possible. Suggestion otherwise “is ‘fightin’ words” and many would suggest that digital audio processing has now surpassed analog in all important respects.

Particularly in mixing, switching, and routing there has been a dramatic bipolar switch over to digital purely on ease and inexpensiveness of implementation as appropriate parts became readily available; Fig. 29-107, a photograph of a couple of LSI ICs and a handful of support parts, illuminates this blindingly; of course it could be argued that the same could simply be achieved in analog with a mere 144 op-amps and 2304 VCAs, but by whom or why is uncertain.

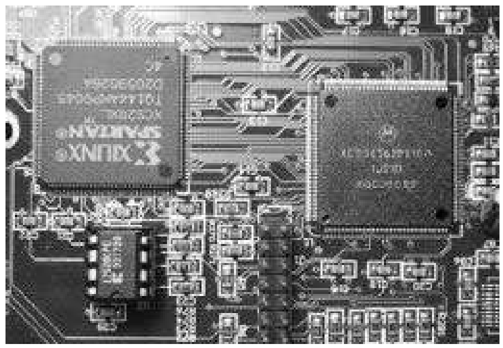


Figure 29-107. Mix stage of a digital console.

Digital recording and transmission are covered extensively in other sections of this book. That these are where digits first made their mark on pro audio is hardly surprising; once the speed of the associated processing speed and bandwidths were high enough, well-proven techniques from the communications and computer world were applied to the problems of storing and moving the fairly prodigious number of bits digital audio demands. After all, the major telephone companies worldwide have been using high-speed digital streams for decades. Early successes for audio include the conversion of the BBC's nationwide radio network program distribution system to digital in 1971. The turn of the 1980s saw the first few serious digital tape machines, heralded by the 3M/BBC design; then a little thing called the PC happened. Hard disk recording moved from the high-end esoteric to the bedroom studio and is now both ubiquitous and universal. And very good. The pro-audio digital revolution is almost complete. Resistance is futile.

Fig. 29-108 shows about the simplest example of a DSP (digital signal processing) system possible. The processor itself, in old days racks of discrete logic and latterly specifically tailored microprocessors, is sandwiched between means of coupling it to an analog world outside. We'll first look at the converters and then the DSP bit.

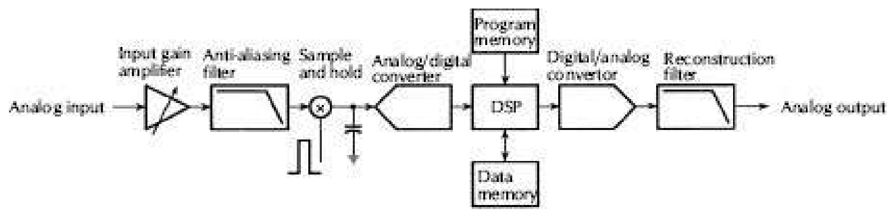


Figure 29-108. A basic digital signal processor in an analog world.

29.18 Converters

29.18.1 A/D Conversion

29.18.1.1 Resolution

A DSP processor needs a stream of digital words of sufficient resolution to adequately portray the actual input signal level at a given moment. This resolution is determined by the number of binary bits in each word; each bit corresponds to a doubling of the resolution, or roughly 6dB of dynamic range capability. Phone systems typically use 8 bits (approximately 48dB dynamic range linear, although effectively more when companded), the BBC's original distribution system was a 13 bit system (78dB), CDs 16 bit (96dB) and most production and recording systems a nominal 24 bit.

The A/D conversion process is fraught, particularly with high-resolution converters, and the actual dynamic range is often much less than theoretically possible from the number of bits. System noises, either from the analog paths or crosstalk from various digital signals, are the predominant limitation; gross errors used to come from nonmonotonicity. Strictly speaking, a converter should, if the input signal is increased by one unit of the resolution, reflect this by increasing the value of the output digital word by one bit. Often, particularly at transitions of the major bits, this goes awry and an untoward jump in output level occurs. (As an example, the transition from 01111111 to 10000000 in an 8 bit word is a likely point of nonmonotonicity. Although this only reflects one increment of resolution change, a lot of convertor bits are changing simultaneously; the more that have to change—especially the wider the word—the more chance for error. Trust is being laid in the converter's manufacturer that each successive bit carries exactly twice the weight of the previous one; with very wide converters the increments of resolution are tiny and the odds are increasingly slim. In a 16 bit converter the most significant bit has to be accurate to within at least a bit of resolution for the device to be monotonic; this corresponds to an accuracy of better than 0.0015%. Enough said. The more bits change in a transition, each of the individual bit's tolerances come into play and errors are far more likely.

The wholesale shift to sigma-delta—type A/D converters, which are inherently monotonic, has all but buried this problem now in most practical circumstances. Integrated IC sigma-delta converters are available very inexpensively with what used to be considered science-fiction performance. As mentioned, the primary limitation is noise, either induced digital mush or from the necessary analog parts in the mixed-signal format of these devices; such is not an ideal environment for low-level analog. The low supply rail voltages (3.3 V is considered big these days) mean that the additional dynamic range available from conventional high rails (such as typically ±15V or more) is simply not available.

Although it is actually quite difficult now to find a convertor that is rated at any less than 24 bits (and to be fair, their internal structure, in particular the word width of the FIRs, is 24 bit), the actual performance depends on how many of those bits represent useful data and how many are marketing bits.

29.18.1.2 Sampling Rate

In addition to the required resolution, speed of conversion plays a great part. In order to give an accurate portrayal over time of an input signals waveshape, there need to be enough conversions for the digitized signal to be reconstructed to an exact analog of the original signal. The lowest theoretical (Nyquist) sampling rate is twice the highest frequency intended to be processed. This implies at least two digital word conversions taking place for each cycle of (typically in audio) 20 kHz. In practice the sampling rate is made even higher, and figures of 44.1 kHz (domestic) and 48kHz in professional audio are the most common. 96kHz and higher converters are readily available, and many keep their production (as opposed to distribution) recordings at that rate, although the additional implicit bandwidth is unnecessary.

29.18.1.3 Converter Limitations and Requirements

Currently the figures of 24 bit linear conversion at a 48kHz rate are de facto standard values in pro audio. Although these parameters are capable of very respectable sonic performance, certainly comparable or in excess of the analog recording and transmission methods digital has supplanted, their practical implementations fall somewhat short of the performance of analog electronics. This is not a snipe at digital; it is clear to anyone involved in the art that any shortfalls are practically irrelevant.

How much resolution is actually required? A good quiet balanced-bus multitrack console's typical input-output path can be expected to have some 26dB head room above an operating level of 0dBu and a noise floor some 90dB below that for a 116dB dynamic range. A similar quality mixer summing a fair number of sources can still be reasonably expected to have a noise floor of −80dBu corresponding to 106dB dynamic range. These values imply digital word widths of some 20 bits and 18 bits, respectively. converters of these capabilities are readily available commercially, if implemented well. A further related question is what is the highest dynamic range signal source? A very good condenser microphone with a very quiet FET in it in a very quiet room is probably the best candidate; it might be able to cope with a 130dB SPL gunshot at the high end of the range while still hearing breathing noises in the room at the other. Why one might want, other than as a science experiment, to encompass this total range without gain adjustment begs to be answered. Practically, though, the dynamic range of nearly any meaningful source (that could end up in a mix) or a finished product (that people might want to spend money to listen to) is actually considerably less than that of available digital hardware implemented and used properly. However, one must not get too cavalier with this approach—clean masters (either multitrack or mixed) should retain a dynamic range well in excess of the intended distribution medium to allow for losses in reduction and processing of the masters for same. Just a little compression can put a big dent in dynamic range.

29.18.1.4 Anti-Aliasing Filters

Early converters used brutal brick-wall filters at the Nyquist frequency to prevent ultrasonic frequencies from being mirrored down (aliased) into the audio and prevent further ultrasonic signals from heterodyning with the sampling frequency.

For example, a 40 kHz signal passing into a 48kHz rate sampler will produce an 8 kHz byproduct that will definitely become audible upon signal reconstruction. There is no way these filters could ever be described as anything other than a bad thing. Their temporal response was appalling, their effect reaching far, far down into the desired audio passband. More than anything else, it was these filters that gave digital audio a bad name in its early days.

Sigma-delta converters come to the rescue. Oversampling, a technique of taking and reconstructing samples at a multiple of the sampling rate (4, 8, 16, or so), allowing the nasty filters to be both relaxed in brutality and moved correspondingly higher in frequency, dramatically improved this situation—the filters had far less in-band effect. Sigma-delta converters typically initially sample 64, 128, or even 256 times above the nominal sample rate with the consequence that the antialiasing can be reduced to as little as a gentle single or double-pole filter; the band limitation is done inside the converters by a phase-linear FIR filter, with considerably reduced sonic impact.

Nevertheless, there have been experiments that indicate that even such benign internal filters at 20 kHz are with some program material and under some conditions audible, in comparison to the same class of filter set twice as high in frequency. Since the only way to properly engineer such a filter at 40 kHz is to double the sample rate, it seems that the predominant improvement (and ever-so-slight at that) of a 96+kHz system is not the increased bandwidth available—arguments will continue to rage about our ability to hear/sense stuff up there, and even the desirability of its existence—but that doubling the rate is the only means of pushing the last vestiges of filter effects from audibility. Since this means doubling the amount of processing hardware in a system, it is not a light decision.

29.18.1.5 Types of A/D Converters

There are three types of converters with possible application to digital audio. Although without question the sigma-delta type rules the roost in pro audio, enough applications use flash and successive approximation for them to be considered here.

Flash Conversion. Flash conversion involves a long train of comparators, such that a given signal amplitude will trip a given number of comparators and fairly simple conversion logic can turn their outputs into a binary word. It is the fastest conversion method as far as logic propagation times; a change in input level is instantly reflected in output code. The down side is the sheer number of comparators needed for a sensible size word width, one for each possible level of resolution; also the offset inaccuracies of the comparators tend to dwarf the required resolution! This said, they are little used, except in some hybrid converters where a 4-bit flash convertor will provide the major resolution of a wider word, leaving the remainder to a more accurate type.

Successive-Approximation Encoder. The successive approximation encoder is a very common form of encoder, especially where high speed at high accuracy and with low latency (processing delay time) are required. There is but one comparator unlike in flash, easing accuracy. Conversion takes at least as many cycles as there are bits of resolution. Operation consists of comparing internal voltages, weighted in accordance with the bit value, against a frozen (by a sample and hold circuit) sample of the input signal. It needs to be frozen since the conversion is not instantaneous and the input signal level could change in the time it does take. The most significant bit's value is half the permissible input range, the second a quarter, the third an eighth, the fourth a sixteenth, and so on in binary weighting. They are applied in turn to the comparator, MSB first. If the sample is larger than the MSB, then the MSB is left asserted; if not, it is dropped. The next weight is applied; if the input sample is still larger than the combination of MSB and No. 2, then No. 2 is left asserted, and so on. Eventually all the bits are tried against the input sample with the bits remaining asserted, forming the 1s of the digital word, the remainder the 0s.

Both of the above converters generate an absolute digital value of the input signal at each sample period.

Sigma-Delta Conversion. Sigma-Delta, or also called Delta-Sigma, conversion starts off in essence by measuring relatively how far the input signal moves, up or down, rather than stating exactly where it is and at a very high sampling frequency. Conversion occurs much more often than the required output sampling rate (e.g., 48kHz) often 256 times higher. The conversion itself is much simpler though. Simplistically, at each conversion it only has to make the decision whether the input signal has moved up or down from where it was last sampled. Its output is a very fast stream of up and down signals; the sampling is fast enough that it can keep pace with the input signal's probable changes, sensing automatically whether large level shifts or tiny ones are taking place. Subsequent intelligence (filtering) keeps track of this torrent of single-bit state changes and renders it down a to conventional digital word width at a conventional sample rate for an absolute output value.

As a method it has many advantages, not the least of which being the enormous internal sampling rate; the input anti-aliasing filter can be relaxed considerably, both in order and cutoff frequency (often it just consists of a single- or double-order filter set much higher in frequency than with other encoders—and sometimes left out completely!). Filtering is left to the internal decimation process.

They are also monotonic, having none of the problems of the other types of comparator level or ladder accuracy. What they do have, and which can be a concern in some applications, is a comparatively very long latency (signal processing delay time) before a relevant sample pops out for digital digestion; at normal sample rates and depending on the length of the FIR decimation filters within, this latency can be around a millisecond or so. Sigma-delta A/Ds predominate in pro-audio.

29.18.2 Digital-to-Analog Conversion (DAC)

29.18.2.1 Conventional Ladder DACs

A means of turning the processed output signal from the DSP back into analog is necessary. These are described in Chapter 35 *DSP Technology*, Chapter 41 *Virtual Systems*, and Chapter 42 *Digital*

Audio Interfacing and Networking, but for completeness are outlined here. A DAC adds together voltages (or currents) of weightings corresponding to the importance of the binary bits. Fig. 29-109 shows a simplistic DAC. The required output digital word is applied and the most significant bit, if set high, sources a current of 1 mA. The next most significant bit sources half that or 0.5 mA, the next bit half that (0.25mA), and so on down to 7.8 µA increment for the least significant bit. In the 8 bit converter shown, the maximum output current is just short of 2mA (1.996mA) with all the bits set (one extreme) and none if all are low (the other extreme). Any current between those two, in 255 steps, which is the resolution of an 8 bit word, can be achieved by setting up a permutation of the input bits. This output current can be converted to an output voltage by a summing amplifier.

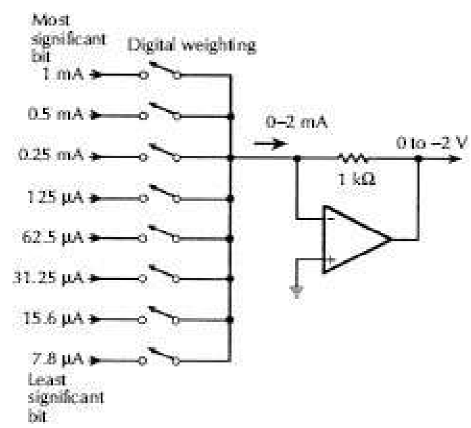


Figure 29-109. Simplistic digital to analog converter.

There are other kinds of D/A techniques, probably the most common being the R/2R ladder, Fig. 29-110. Much as the simple DAC, asserting a bit causes a correspondingly binary weighted current to be output.

29.18.2.2 Reconstruction Filters

All the earlier comments about anti-aliasing filters apply here, too. As well as the required audio—up to 20 kHz in bandwidth—coming out of the DAC there are a host of other products, the most unappealing and closest in spectral terms being a mirror image of the audio centered on the sampling frequency and descending in frequency; a 20 kHz audio signal sampled at 48kHz will be output from the DAC along with an image at 28 kHz (sample frequency minus audio). Heterodyning strikes again; there will also be an image at 68 kHz (sample frequency plus audio), and in all likelihood more sets of images centered on harmonics of the sampling frequency. The most dangerous sonically though is that first inverse image.

29.18.2.3 Oversampling

Enter a filter every bit as precipitous as the one needed at the front end. Every bit as nasty, too. Solutions other than good, well-designed filters come from the digital domain; oversampling, for one. One approach is to intersperse an interpolation filter between the processor and DAC. This digital filter reconstructs the audio but at a higher sample rate; the smoothing of the filter effectively creates more sample points between the few actually being issued by the digital source (DSP). If a guessed digital word is inserted between each of the real ones, the effective sampling rate becomes doubled and, in practice, the DAC is working twice as hard and fast outputting analog.

Here is the good part. If the sampling rate is doubled, the heterodyning images start that much higher up in frequency; following the earlier example through, a 20kHz signal's first inverse image is now going to be at 76 kHz (96 kHz – 20 kHz) instead of 28 kHz as before. The immediate benefit is in the relaxation of the reconstruction filter—it can be much less steep and pushed up in frequency somewhat away from the audio band.

The oversampling process can be carried on even further; four times, eight times, even sixteen times and more with greater oversampling rates commonly used, pushing the undesired products correspondingly higher in frequency and so dramatically relaxing reconstruction filter requirements. The fact that 15 out of 16 samples may be filter guesses belies the fact that it isn't those that improve the audible performance—it is the absence of brutal analog filtering that makes all the difference. Exactly the same conditions apply here to the application of higher (96+ kHz) sample rates, simply with the intent of pushing anti-aliasing filter effects out of audibility, as in A/D converters.

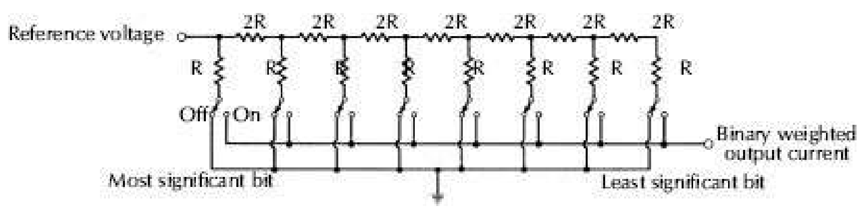


Figure 29-110. An R/2R digital to analog conversion ladder.

29.18.2.4 4 Sigma-Delta DACs

Sigma-delta DACs oversample to the same degree (64, 256, or beyond) as their previously described A/D brothers and the corresponding increase in frequency of the reconstruction filter dramatically simplifies their implementation. Most D/As in pro-audio are now sigma-deltas, although conventional ladder types are still in use, although not so much at baseband audio frequencies. Again, latency is the only major drawback to this type; the processing delay again depends on sample rate and the particular device and its filter length, but is generally around a millisecond. This, of course, means that a system using sigma-deltas at both ends (ADC and DAC) can potentially have a latency of a couple of milliseconds or so; this can be a bust in some applications.

29.18.3 Sample-Rate Converters (SRCs)

A big problem facing digital audio system designers in the early days was combining sources from different machinery that, unless heroics were performed and all the system's machinery was phase and word-clock synchronized, almost certainly were all running at ever-so-slightly-exactly-not-quite on the same frequencies, from their own independent clocks. Mixing such is a disaster.

SRCs allow sources with a wide range of sample rates to be reclocked to the console's master clock, allowing them to be processed normally. Being that they use internally very long FIR filters and of varying length depending on the ratio of input-to-output rate conversion, they not only have latency, but it changes too. Another slight shortcoming is the tendency to affect the ultimate dynamic range by leaving artifacts way down in level, but most current parts are excellent in this regard. All in all, they are a near miracle-cure for what was an intractable problem.

29.19 Digital Signal Processors

There are a number of features that distinguish devices specifically designed for DSP from the admittedly bigger and faster but generally dumber and seriously more expensive behemoths powering PCs and the like.

29.19.1 *Multiplier/Accumulator (MAC)*

The heart of a DSP device is the hardware multiplier in its arithmetic logic unit. This takes two full data-width numbers, multiplies them, and leaves the result in an accumulator, quickly. Further products of multiplications can be arranged by a software instruction called a MAC, Multiply/ACcumulate, to be added to results previously stored in the accumulator. The MAC is central to DSP. Nearly every manipulation of a signal in the digital domain is ultimately achieved by multiplying a sample by another value, called the coefficient. The simplest example is that of level control—in audio terms, gain control. If an incoming sample is multiplied by a value of 1, the result lodged in the accumulator is the same as the input sample. If the gain-defining coefficient is greater or less than 1, the accumulated result is correspondingly greater or smaller than the input sample.

The accumulator necessarily needs to be of a wider word width than the input byte width capability since a multiplication can end up with a much bigger or smaller number than the input sample; in the case of the very popular fixed-point Freescale 56 series DSP chips, the bus width (input-output word width) is 24 bits while the accumulator widths are 56 bits. It's a worthwhile rule of thumb—a multiplication results in double the bit width.

In this volume-control example, the input analog signal is sampled at the front end of the encoder, an A/D conversion is performed, and this value is deposited on the DSP chip bus at its command. The input word is multiplied by a coefficient, similarly picked up off the data bus, and the result left in the accumulator. Rounding off fits the possibly too long result to the width of the DAC (e.g., down to 16 bits from a possible maximum result of 32 bits from one 16 bit multiply). The answer is put on the bus to be picked off at command by the D/A convertor. The D/A performs a near-instantaneous conversion back to analog, ready for consumption by the real world. This whole routine is repeated 48,000 times a second; each operation has less than about 20 μs to take place. Congratulations! This is the digital replacement for a $5 potentiometer.

To get a sense of the great strength of the digital solution, Fig. 29-111A shows many A/D and D/A converters hanging on the DSP input-output bus. Each of these is independently addressable by the DSP chip; it can systematically pick an input signal word from any A/D, work on it, then deposit the result into any D/A converter. Further, it can take input samples from any or all of the A/Ds, multiply them in differing degrees according to differing coefficients, and add the results progressively within the accumulator. This accumulated result is then scaled and passed to a D/A. In effect, this is the digital equivalent of mixing a number of sources, all the sources at different gain settings, to one output.

The comparatively simple digital arrangement can be made to equate to an analog soft matrix, as drawn in Fig. 29-111B. It's starting to look more like a viable cost and space saving replacement; this small example of six-in and six-out is already equivalent to 36 VCAs.

More inputs and more outputs to and from the mix stage are of course possible. The principal limitations are accumulator width, which is taken care of by building in adequate head room just as one does in analog, but more importantly processor time; after all, it still has to do all the input-to-output multiplies within a 20.8 ms window.

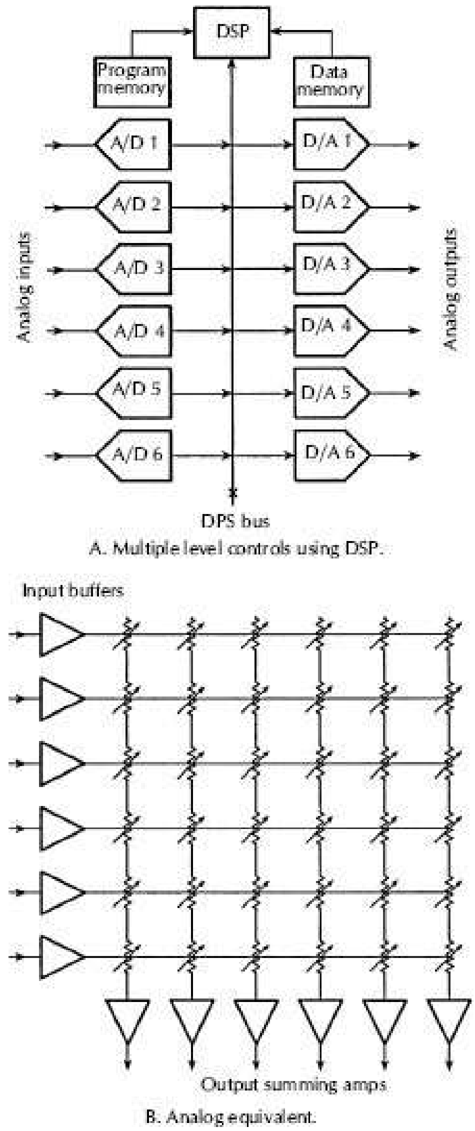


Figure 29-111. A digital mix of a number of sources.

29.19.2 Instruction Cycles

A processor instruction cycle is simplistically the time it takes to perform one single simple operation, such as a bus access (to acquire or dispose of data), an arithmetic function or a move of data from a register to elsewhere. Multiplies can take a bit longer, depending on the chip, but DSP chip architectures with hardware multipliers are very slick and time efficient. They need to be. The processor speed determines how many of these clock cycles are available for processing in a given time window and so directly limits how much work the processor can do. For example, a 400 Meg device has a processor cycle rate of 400,000,000Hz. Given a 48 kHz sampling rate, this gives a maximum of just over 8000 cycles per sample period. Some operations can take more than just one clock cycle, so this is an outside ideal figure. In practice, it works out at somewhat fewer. Although it looks like a big number, it seems vanishingly small as soon as anything clever is attempted with the DSP. This, above all else, is the primary reason why upping the sample rate above the bare possible minimum is a very unpopular notion in DSP circles. Cycle budgets rule.

29.19.3 Processor Types

Specific DSP devices are chosen for a wide variety of reasons, both real and perceived. Device flexibility, per-unit cost, and ease of implementation (in the forms of support from the manufacturer and quality of the design tools) all factor in. In very large run products such as consumer items, part cost will probably override everything else while ease and speed of implementation tend to be more important in lower-volume, high-tweak-factor arenas such as pro-audio. Rarely is there one overweaning performance feature that makes or breaks a choice. However, since in order to squeeze the most processing from each device a considerable amount of their programming can still be at the machine-code level, the designer's familiarity with a particular assembler language can have a strong influence—this definitely falls into the ease and speed of implantation department.

Perhaps the minimum for processing audio data is a 24 bit word width and correspondingly wider accumulators and registers. As such the Freescale devices just about fit the bill. They are fixed point processors, which directly limits their dynamic range to the number of bits (144dB for 24 bits, 336dB for the accumulators); fortunately, this is plenty for real-world audio processing. Some applications, like some filters, demand wider immediate dynamic ranges in their calculation and intermediate-value data storage, and for those instances long or double-precision arithmetic is used. The down side is that such filters can take up to twice as long (twice as many processor cycles*) to calculate as single precision.

29.19.4 *Floating Point*

Floating point processors (floaters) as exemplified by Analog Device's "Sharc" or TI 6000-series processors avoid this problem by representing numbers internally in exponent/mantissa format, having far more involved internal processing to handle the complexity of dealing with these numbers. These "Floaters" can be operated as either wide fixed point or 32 bit floating point. Since the dynamic range of a floater is as good as infinite regardless, none of the dancing around one sometimes has to do with a fixed point applies. On the other hand, the capacity to dig a big hole is as good as infinite, too.

29.19.5 *Parallelism*

DSPs differ from conventional microprocessors in that their architecture is contrived to make certain common processes as slick as possible and to be able to perform as much real data manipulation and housekeeping within each clock cycle as possible. This latter is called parallelism, and the degree of parallelism is what sets devices apart in capability. For instance, performing an FIR filter (or a mixer routine, for that matter) within one clock cycle, a DSP can.

- Do a multiply and accumulate (MAC).
- Fetch in the next data word ready for the next MAC, update data pointer.
- Fetch in the next coefficient ready for the next MAC, update coefficient pointer.
- Update the program counter.

In short, everything that needs to happen to ensure a filter point can be calculated within one cycle is done, ready for the next.

29.19.6 Multiple Memory Spaces

Memory allocation is nearly always somewhat different from ordinary processors, which usually have just one memory space shared between all functions; DSPs have at least two separate memory spaces; the Freescales have three, for example: one for the program information (program memory), one typically for coefficients, and one typically for the inevitable intermediate filter values etc. necessarily stored between per-sample calculations and for internally stacked-up audio data, if brought in enbloc from outside.

29.19.7 Real-Time Specific Peripherals

Additionally, most DSPs have convenient peripherals built into them to allow ready, seamless, and fast transfer of data in and out of the chip, either into memory-mapped data space and/or through a variety of serial communication formats. It is usually possible to seamlessly connect a DSP to a number of conventionally serially formatted A/D or D/A converters and to other DSPs. Definitely not least, a ready and fast means of importing fresh programs/coefficients with which to modify the data is always available via a host port.

A related major tool is DMA, or direct memory access. This allows the moving of considerable amounts of data into and out of the DSPs memories with little impact on the main cycle budget other than that required to set up the necessary pointers and to fire off the DMA activity at the required times. On very busy processors conflicts can arise (DMA does borrow some real processor resources while it's not looking, and under normal circumstances usually gets away with it) so it's not entirely free, but it is more than handy.

But what of x86 type PC processors? Although they don't have the separate memory spaces, slick parallelism and on-board peripherals, are comparatively expensive, demand large amounts of support circuitry and run very, very hot, they are blindingly fast. This blows past a multiple of sins.

29.20 Time Processing in DSP

29.20.1 Time Manipulation

Something what is very readily achieved in digital is storing information, either long term onto disks or flash memory, medium term in RAM, or short term within processor registers and internal RAM. Nearly all manipulations of data of any complexity greater than the soft matrix example above demand storage.

29.20.2 Delay

A stream of input data is written into RAM memory and subsequently, some time later, read out again. The length of time recordable (sample length) depends on the size of memory and the sampling rate—the faster the rate, the quicker the memory will be eaten. This memorized sample may be stored elsewhere then—say, on a hard drive.

Say a relatively short time delay is required for an echo. The input data stream would be written into RAM and read out at a fixed time (a certain number of samples) later. Sooner rather than later the memory would run out and the delay stop, so the memory is usually arranged as a circular buffer; when the buffer end is reached, the memory register leaps back to the start of the buffer and overwrites what was previously there, and so on. The buffer is read in the same manner, at a time after it has been written determined by the required delay. As long as the buffer is long enough to contain enough samples for the required delay, a continuous delayed output version of the input is available. The main advantage to the seeming complexity of the circular buffer is that only pointers, or indexes, are being changed and updated; the only audio samples being changed are the writing of the newest sample over the oldest one. What is important is what is not happening; huge amounts of data are not being read and rewritten somewhere else. The complexity is merely keeping track of those read and write pointers, which is in reality simple arithmetic and indeed an automated function in many processors.

29.20.3 Echo, Echo, Echo

Reentrant, or recursive, delay (spin echo) where a delayed signal repeats continuously until fading away is achieved by attenuating the delayed words (in the multiplier) and adding them in the accumulator with the concurrent new input words.

If the delay is made very short, and the delay is summed in the accumulator with the new, direct sample, something interesting happens—a direct parallel with the analog world. The direct and delayed signals sum and interfere. A 1ms delay corresponds to a half wavelength of 500Hz; in other words at 500Hz with 1ms delay, the delayed signal will be out of phase by 180° from the input signal. They will cancel and a notch at 500 Hz (and every 500Hz interval up the spectrum) will occur. Altering the delay time alters the frequencies at which cancellation occurs; studio people call it flanging; we'll call it a comb filter, our very first digital filter.

29.20.4 Reverberation

In real acoustic environments, reverberation is the summation of countless random time-delayed reflections and rereflections from floors, walls, ceilings, and obstacles. Complications set in with differing reflections having differing frequency aberrations due to varying surface absorption coefficients, but in essence it is an accumulation of time-delayed signals of various and decreasing levels. As such it can be reasonably well emulated in DSP by more or less complex variations on time delay; relatively long time delay loops are established to emulate major room reflective modes. Many short loops and all-pass configurations are used to emulate the decorrelation that occurs in an acoustic space by multiple short reflections and diffraction. The output-to-input feedback—terms for each of these elements is adjustable and equalization, typically in the form of simple roll-offs—are applied either after a loop or within its feedback path to mimic the typically higher absorption at higher frequencies in an acoustic environment. There are a lot of small and large elements, all with a lot of handles, or things that need to be fed parameters.

Basically, the number of elements and the skill in determining their convoluted interaction and parameters decide how convincing the reverberant effect is and its characteristics. Some astonishingly good results have been had from DSPs with quite small (64k word) external memories.

As DSPs become more powerful and much cheaper, becoming increasingly practical is a class of reverberation units that in effect perform a very, very long convolution of an applied audio signal with a digital recording of the reverberation tail of a real venue (see section 25.21.1 *Transversal Filters*, for the basic technique). This can involve hundreds of thousands of DSP multiplies (meaning lots and lots of DSPs) but is as one would expect highly impressive and flexible. Proprietary convolution algorithms can reduce the computational burden, but it is still nevertheless a big proposition.

29.20.5 Averaging

An average of a number of input samples is achieved by adding all the input word values for the period of time over which the average is required; this is normally figured out by numbers of samples—20ms worth of samples at a 48 kHz rate equates to 960 samples. (This would be a l-o-n-g train of samples.) These samples are all added in the accumulator and then divided by the number of samples—the result is an average value for that 20ms. If each sample is stored elsewhere, then a rolling average becomes possible; for each new input sample added in, the first sample of the 960 is subtracted, and a new average for that instant is calculated.

Division, as such, is something undertaken only under extreme duress in DSP; it is very thirsty and inefficient. A division in such a case as creating an average as here could be achieved by first arranging, if possible, that the average length is a binary interval (2, 4, 8, 16, etc.). Then the end result of all the additions could be bit-shifted right the corresponding number of times. An arithmetic shift right (moving a digital word one step to the right, filling in the now missing top bit with a zero) is the same as dividing by two; an average of 64 samples would thus need six right shifts. Alternatively, a single multiply by 0.015625 (1/64) (or the reciprocal of whatever the arbitrary number of samples may be) does the job. Either is an awful sight quicker to do in a DSP than a 24 bit division. Anything rather than divide.

(Along with any transcendentals, or doing a square-root, or many other common and seemingly “no-brainer” mathematical functions, division is expensive in DSP. This is a major reason code written in high-level languages often translate badly and seem very bloated when compiled for the DSP; simply typing “/” or exp() or sqrt() seems painless. DSP programmers have tricks to streamline or avoid such pitfalls.)

29.21 DSP Filtering and Equalization

29.21.1 Transversal, Blumlein, or FIR Filters

Yes, Alan Blumlein invented these, too. The train of samples concept becomes very valuable in DSP. This type of filter can produce a wide variety of time effects and frequency response shapes, particularly bandpass and cutoffs. While the determination of coefficients for the various filter types is beyond the scope and intent of this section, the underlying principle is shown in Fig. 29-112. For each sample period (i.e., every 20μs) a fresh input sample is inserted at the head of the train; all the samples move along the train and the oldest one falls out of the other end and is lost. Each sample is multiplied by a coefficient specific to that position and summed in the accumulator with other results from the other multiplied samples. Each pickoff is subject to a different coefficient and sum sense (normal or inverse). The accumulator value is the new output word for that particular sample time; 20.8 μs later the whole routine starts over again.

This passing of one set of data (in this case audio) through another set of data (coefficients) is also called convolution.

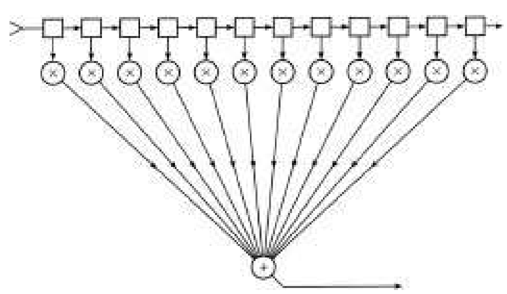


Figure 29-112. Transversal or finite impulse response filter (FIR).

29.21.1.1 Impulse Response

As familiar as we are with using the tool of frequency response measurement to analyze or describe the transfer function of a device or circuit, there is an equally powerful descriptor: the impulse response. Embracing the impulse response concept aids gaining a mental picture of how digital filters work. Fig. 29-113A is what the waveform of a large bell excited by an impulse could look like: a damped sine wave at the tone of the bell. (Hardly dissimilar to that from a damped oscillator, or bandpass filter. Hold that thought in mind. Actually, looking at the response, it would probably sound more like the dung of a lamp post, but please suspend disbelief for now.)

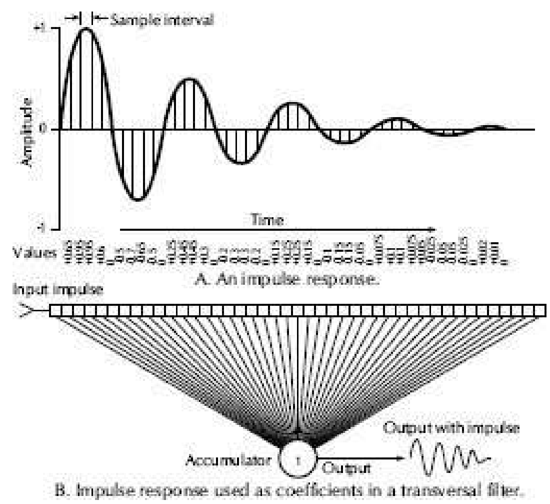


Figure 29-113. An impulse response becomes a filter.

Each of the vertical lines represents the instantaneous amplitude of the signal at each sampling period; this, if you like, is a sample-by-sample digital recording of the bell's sound. If we were to play back the samples at the rate they were recorded, we would hear the bell thunk again. We now use the bell's samples' numeric values as coefficients in a transversal filter, Fig. 29-113B, and send an impulse (one sample of full positive amplitude, the rest zero) into the filter; the effect is exactly the same. As the impulse passes each coefficient, the bell sound will be reconstructed once more.

There is nothing to stop us from putting real audio samples into the front of the transversal filter—the effect will be as if the audio is being played through a damped bandpass filter at the frequency of the bell. The bell's impulse response is impinging itself directly on the audio passing through the transversal stages. It will sound as though you're listening to the audio with your head stuck up inside that bell. Yes, it's a filter! In short, if we can describe a desired filter's impulse response and use its samples as coefficients in a transversal filter, any signal passing through the transversal stages will be filtered accordingly.

This kind of processing is commonly called FIR (finite impulse response) filtering. If a transient (impulse) were encoded and applied to such a filter, the samples describing it would enter the train of stages. Output summation contributions occur until they reach the end. When the last relevant sample has fallen out the end of the train, no further output samples that have anything to do with the originally applied transient are possible. The duration of the transient within the filter is limited to the lifetime of its samples in the train; they eventually all leave. The impulse's existence is finite. The filter's length is finite—hence, finite impulse response.

Intellectually, FIRs are very appealing through their very simplicity. Unfortunately, this genre of filtering is rather taxing in current DSP terms since it demands a lot of processor time for any useful audio filters. As a rough rule of thumb, to do anything meaningful at a given frequency the filter must be able to contain a full cycle of that frequency; to operate at 50 Hz an FIR would need to be at least 20 ms long, which (assuming a 48kHz sampling rate) would be about 1000 filter points long. A 200MHz part only has about 4000 cycles of processing available per sample—this one filter has just eaten about a quarter of a whole DSP! Suddenly, except for a few rather special and esoteric circumstances, such as phase-linear EQ and auto adaptivity, it becomes obvious why FIRs are not particularly popular in mainstream audio DSP processing. They are rather hardware and time thirsty.

Impulse response coefficient sets suitable for plugging into transversal filters may be either calculated (long-handed for the rigorously inclined, or within any of the many excellent filter design programs available) or, as in the only half-joking bell example above, recorded by issuing an impulse into a pet filter and using the resulting sampled output as coefficients—audio played through an FIR with those coefficients will sound just as if it was passing through the original filter. As earlier mentioned, there are reverberation units working exactly on that principle.

29.21.1.2 Windowing

Any attempt to generate a set of coefficients for FIRs will run into the problem that an ideal filter simply will not fit into the length of any practical filter. Obviously, the filter has to be long enough to realistically encompass the meat of the desired processing (a 99 point filter won't do 50 Hz, remember?), but this still leaves the problem that the filter is finite in length. A series of FIR filter designs showing the impulse responses and corresponding frequency responses of a 33 point (33 step long) nominally 12 kHz high-pass filter highlight the quart/pint-pot tradeoffs. Truncation, i.e., lopping the end(s) off to make it fit, leads to Gibb's phenomenon, in which the desired output frequency response of the filter is seriously compromised by large lobes, Fig. 29-114.

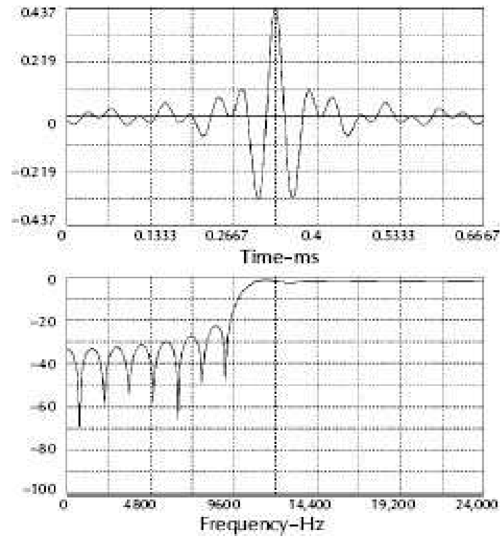


Figure 29-114. A 33-point unwindowed FIR filter.

Mr. Hann, Mr. Hamming, and Mr. Harris (among others) come to the rescue here, with a technique called windowing. These apply weighting to the values of the coefficient set, basically leaving the most significant elements (usually in the middle of the set) alone and tapering off the values toward the ends of the set. The taper that is applied varies according to the type of window, and the differing types are best suited to differing interests of compromise. Say a brick-wall filter had been described as in the figures; one window may optimize for stop-band rejection, Fig. 29-115, another may trade that against sharpness of the filter cutoff rate, Fig. 29-116, etc. Many thanks to Momentum Data Systems' software for the curves.

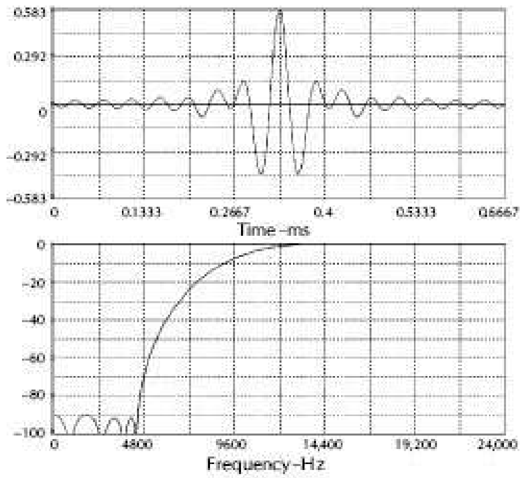


Figure 29-115. 33-point filter Harris windowed.

29.21.1.3 Symmetrical FIRs

There is an FIR implementation that has some quite interesting properties and as such is probably the most used, so much so that the majority of commercial design packages assumes as a default that one wishes to design symmetrical FIRs and that FIR has become almost synonymous with the symmetrical filters they afford.

These allow the imposition of a frequency response (in the case of a conventional-style EQ) without altering the phase response, unlike ordinary EQ (and nature) in which any frequency response change comes with a corresponding shift in phase response for free. Although this characteristic might at first blush seem ideal and a major leap forward for audio technology and world peace, in practice they are only rarely used; yes, Virginia, they do sound different to conventional EQ with equivalent frequency responses, but not necessarily better. (An odd effect is that one seems to need more phaseless EQ cranked in than conventional EQ for a similar subjective effect.) Certainly, it's not better enough to displace conventional EQ, which can be readily and far more efficiently created in either digital or analog form. The difference alone, however, is sufficient reason for existence in music production, and special-effects units and audio workstation plug-in software specifically to do symmetrical FIRs are available. There are applications where phase-linearity is essential, such as in airchain processing, and this is where they shine.

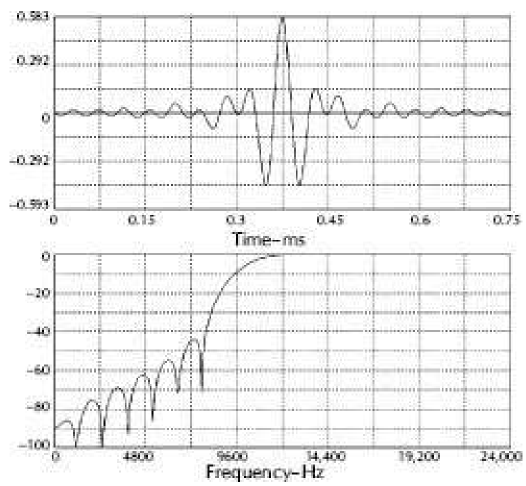


Figure 29-116. The same 33 point filter Hann windowed.

Symmetry refers to the fact that the coefficient set is arranged to be symmetrical about the center—the midpoint—of the filter; identical coefficient set-lets tail off toward the end of the set as tail back toward the front. The midpoint of the filter is regarded as the time center—in other words, a symmetrical FIR has an intrinsic time delay of a passed signal of half the length of time the filter takes to calculate; in our now-famous 50Hz capable, 960 point filter, the effective time delay is 10ms, or half of the time it takes for any one data sample to transit the entire filter, being 20ms. This inherent latency is another major downside to symmetrical FIRs; in order to keep everything in a multisource console time aligned, all other sources would have to be delayed by the effective time delay of just one FIR’ed source. Sometimes latency is a real issue just by itself.

Note that not only is half of the filtering done after to the time center, but, and this is the head hurter, half of it is done before the time center, leading up to it. The filter only remains causal because of the intrinsic time delay. That the ear can deal with filtering effects before something has happened and integrate it all into an acceptable sound is a true amazement.

29.21.2 Recursive Processing

This concept was approached in achieving spin echo and reverberation; feeding an already manipulated input sample back around in a loop to be reprocessed along with new, and/or yet other samples. Fig. 29-117A shows this diagrammatically. A time delay (a number of samples’ delay) is included in a loop and fed back at a level determined by the controlling coefficient. Picking off different samples and treating them with differing coefficients allows great control over the nature of the feedback and the dynamic nature of the loop. The most important thing to note is that once a signal has entered the loop, it just carries on going around and around, being summed with fresh input samples each time. The time taken for a signal to die away is determined by the coefficients in the feedback loop—this can very loosely be paralleled to the analog concept of Q; the more positive feedback in a filter the tighter its response, with the drawback that ill attention to its control can result in oscillation. Such is exactly the case with digital recursive stages. Even if controlled, the signal never actually dies away completely; in DSP this can result in leftover bits rattling around, manifesting as repetitive cyclical errors. Sufficient accumulator width needs to be available to round or noise-shape results off nicely.

The first big advantage of recursive processing is that significantly less memory accesses are needed than with FIR—history (equating to length of the filter and its temporal resolution) is built up within the loop rather than being necessary individually and sequentially. The second advantage is that far fewer coefficients and operations are needed.

29.21.2.1 IIR Filters

A filter built up around recursive techniques is known as an infinite impulse response filter or IIR, Fig. 29-117B, so called because once in the loop, an impulse just keeps trundling around indefinitely, infinitely. In practice it gets rounded off sooner or later, but it makes the distinction from the FIR.

Additionally, an output appears from an IIR at the same time as input samples are applied and the filter starts behaving as a filter immediately; the only delay is the group delay of the filter, exactly as in analog; there's no waiting for sufficient data to be affected by sufficient coefficients for the nature of the filter to become formed, as occurs in symmetrical FIRs.

IIRs are presently easier, quicker, and carry less time and memory overhead than FIR filters; consequently, they tend to be much more popular for audio DSP.

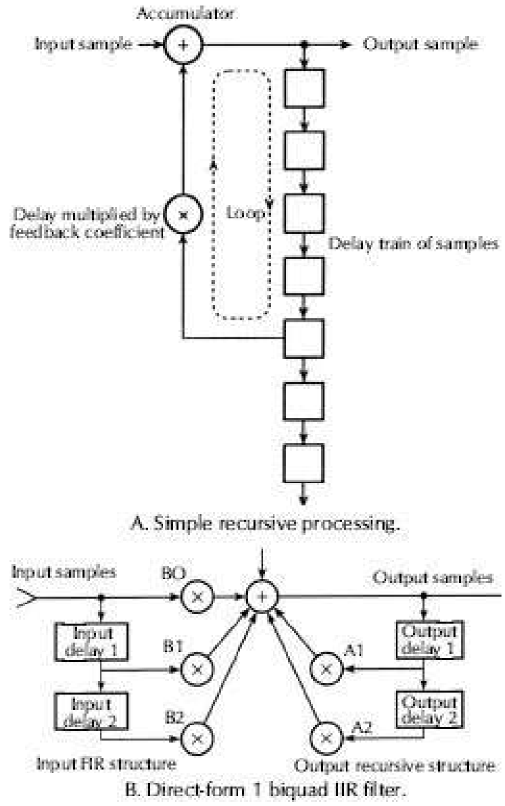


Figure 29-117. Recursive processing.

29.21.2.2 The Biquad

There are many different ways of implementing IIRs; the one in Fig. 29-117B is known to its chums as direct form 1 biquad and serves to illustrate the process well. (Others may use less memory space, or run a bit quicker, but they come up with the same results.) It's known as a biquad since it essentially calculates a biquadratic equation. There is no real limit to the number of input and output delays, multiplies, and summations; it's just that those longer than a classic biquad tend to be rather high-strung creatures that are far less easy to calculate coefficients for and to keep tame (read stable).

There are two halves to the filter: the input stage consists of a short, 3 tap, FIR. The applied input signal is multiplied by a coefficient (b0) and the result put in the accumulator; the two delay line outputs are multiplied by their coefficients (b1 and b2) and added to the accumulator. The output stage is a two-delay recursive section; each of the two delay-line outputs are multiplied by their respective coefficients (a1 and a2), the results adding to the accumulator, the total contents of which now represent the output. Once used in a sample time's calculation, the contents of the input and output delay lines move down one, such that the value in input delay 2 is displaced by the former contents of input delay 1, which is in turn filled by the last used input data sample; similar actions occur in the output delay, only the last calculated output sample enters the delay line.

In short, the biquad takes all of five coefficients, five MACs (multiply and accumulates) and a bit of concurrent shuffling of data to happily create a second-order high-pass, low-pass, or bandpass filter. They are very quick and easy to implement in DSP. That it is the basic building block of most digital EQ is hardly, therefore, a surprise.

29.21.2.3 Coefficient Analysis

A look at the coefficients for the input 3-point FIR can give a clue as to what class of filter the biquad is running. (Actually, complicated only by the higher number of coefficients, the same sort of analysis can be done to longer FIR coefficient sets, too.) Fig. 29-118 shows three biquad coefficient sets. Set A shows equal and opposite coefficients for b0 and b2, and none for b1. A very low applied audio frequency (dc, or practically so) will present substantially the same input signal level over the three samples in the FIR; this means that what is contributed to the MAC by the input sample by the b0 multiplication is going to be substantially canceled by a similar size signal from b2 (b1 not contributing at all, being zero). So low frequencies are not being passed. By a similar token, a signal of half the sample frequency (Fs/2) would have near identical values in the first and third positions, i.e., being administered by b0 and b2; since these coefficients are inverted this frequency gets nulled too. The highest valid audio frequency (Fs/2) is not being passed, and neither are very low frequencies. With a bit of luck something in the middle will be, so this is in all likelihood a bandpass filter.

On the second coefficient set, Fig. 29-118B, the effect of multiplies by b0 and b2 are entirely canceled by the effect of b1 at dc, yet by virtue of b0 and b2 not being inverted with respect to each other, Fs/2 is not nulled. The conclusion would be, if one didn't already know, that this was likely a high-pass filter.

The only apparently slightly misleading case is that of the lowpass filter, Fig. 29-118C, which one would expect to bung in a null at Fs/2, but doesn't at first glance seem to, being that the b0 and b2 coefficients are the same. Aha! though. If one imagines the positive peaks of an Fs/2 signal being coincident with b0 and b2, then the b1 coefficient is coincident with the negative-going peak—the b1 coefficient, being the sum of b0 and b2, neatly cancels their effect by creating a negative signal equal to their positive contributions at Fs/2.

48k	bpf	1000Hz	Q 0.707
b0		0.292543	
b1		0	
b2		−.292543	
a1		−1.98289	
a2		0.7074571	

A.

48k	hpf	1000 Hz	Q 0.707
b0		0.9115751	
b1		−1.82315	
b2		.9115751	
a1		−1.815318	
a2		0.8309824	

B.

48k	1pf	1000Hz	Q 0.707
b0		3.916071E-03	
b1		7.832143E-03	
b2		3.916071E-03	
a1		−1.815318	
a2		.8309824	

C.

Figure 29-118. Biquad coefficient sets for bandpass, high-pass, and low-pass filters. Note the similarity in the a1 and a2 coefficients; the class of filter is determined by the b0, b1 and b2 coefficients of the input FIR structure.

Fig. 29-119 graphically shows these analysis results. In short the input FIR is a dumb little filter of the same class of the overall filter —and actually is what determines its class—the output feedback IIR structure in effect determining the frequency and Q.

29.21.2.4 Filter Quantization Distortion

The output (recursive) stages of a biquad can cause some pretty wild signal levels to be achieved in the accumulator—they are, after all, little more than a slightly complex feedback circuit; the output signal is fed back in part through both delays in accord with their a1 and a2 multiplies and grows until it (hopefully) stabilizes. (This can be likened to operating a PA right on the edge of feedback at microscopically varying degrees; this is standard operating procedure with IIRs). Filters that are either high in Q and/or more importantly very low in frequency with respect to the sample rate (and that, unfortunately, means most EQ-type frequencies) exacerbate this effect. The cure is to only excite the filter to the degree that the desired output is unity with respect to the source; this is usually achieved by proportionally scaling back the coefficients in the FIR input chain. Looking at the coefficients in Fig. 29-118C, this can be seen clearly; the b0 and b2 coefficients are really quite small; the reciprocal of this value is the amount of gain being generated in the IIR output chain. Thankfully, commercial design packages and most cookbook coefficient calculation routines take this scaling into account. But the underlying issue is quite serious. Using, say, 0.0001 as a b0 coefficient (not unrealistic) and assuming a maximum input signal of 1 (the maximum signal range using the fractional arithmetic scheme in some fixed-point DSPs is 1 to −1), then a value of 0.0001 will end up in the accumulator, and despite the huge feedback-derived gain in the IIR output chain, the contribution to the output from the input signal is still only 0.0001; this corresponds to −80dB. If the output were to be truncated to 24 bits (144dB) the bottom 13 or so bits worth of the input signal would effectively, be sawn off and thrown away, leaving us with an 11-bit system. In numbers, this leaves a maximum signal to floor ratio of only 64dB; if the normal operating level of the system is −20dBFS (deci-Bels below Full Scale) (0.1), that is only −44dB signal to floor. Practically, with rounding, noise shaping, or dithering, things are worse.

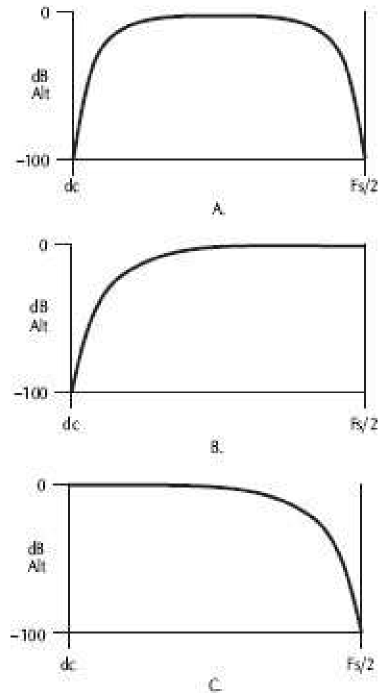


Figure 29-119. Three point FIR filters from b0, b1, and b2 of Fig. 29-118.

The good news is that such quantization noise in filters can sometimes be masked somewhat by the very signals the filter is passing; the bad news is that when it is audible, it is Audible. And even when not overtly audible, it lends itself to a disquieting roughness to the sound that is difficult to pin down. The accumulator has all the width needed for valid data; standard practice on any filter on which this is even likely to be an issue is to make the IIR chain delay storage wide enough to fully encompass the attenuated input signals. In the specific example of a 24 bit fixed-point processor, the IIR output delay chain is made long, or double width at 48 bits. It also means that nearly always the a1 and a2 IIR multiplies need to be long too—i.e. the lower 24 bits need to be MAC'ed in with the upper 24, which increases execution time of the filter.

Sharc and other floater programmers are permitted to smile at this point. It is a happy day to doff the shackles of fixed point.

29.21.2.5 Cascaded Biquads

Having previously noted that IIR filters with more than a biquad's worth of delays and multiplies are not attractive, there are approaches to coupling more than one biquad with the intention of making more complex or effective filters or simply those of a higher order. Better than just running one after another. Fig. 29-120 shows such an arrangement; the second biquad uses the output delays of the first as its input delays, and so on.

29.21.3 Parametric EQ

Raw biquads can take care of most traditional filtering. One approach to doing a console-style parametric EQ section, with independent control over center frequency and Q of the employed filter and of the amount of lift or cut introduced, is shown in Fig. 29-121. A standard biquad is fed directly from the source audio, which is also attenuated by (in this case) 12dB by the expedient of arithmetically shifting the data two bits to the right (down), or by multiplying by 0.25, in the DSP. The filter's output, fed through an attenuator, is summed with the attenuated direct signal, and the result arithmetically shifted (ASL) two bits to the left (up 12dB). This shifting up and down allows a correspondingly higher amount of the filter to be present in the output, which is required if high levels of boost are required. This example's 12dB allows a maximum boost of 13.8dB to be achieved, which happily encompasses the ±12dB control range often found in EQs; more boost capability would require greater shifting down and back up. One can in a floater (floating-point DSP) leave the straight signal alone and simply multiply the filter output up as much as is necessary (the attenuator becomes a gain stage) and avoid the shifts entirely.

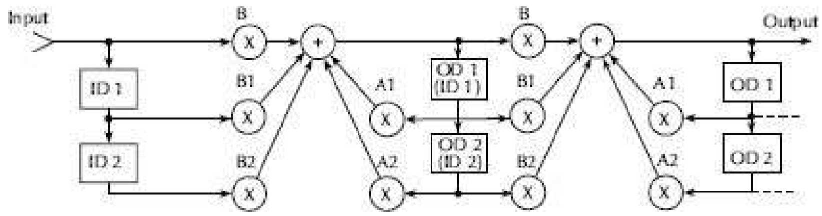


Figure 29-120. Cascaded DF 1 Biquads, sharing delay lines.

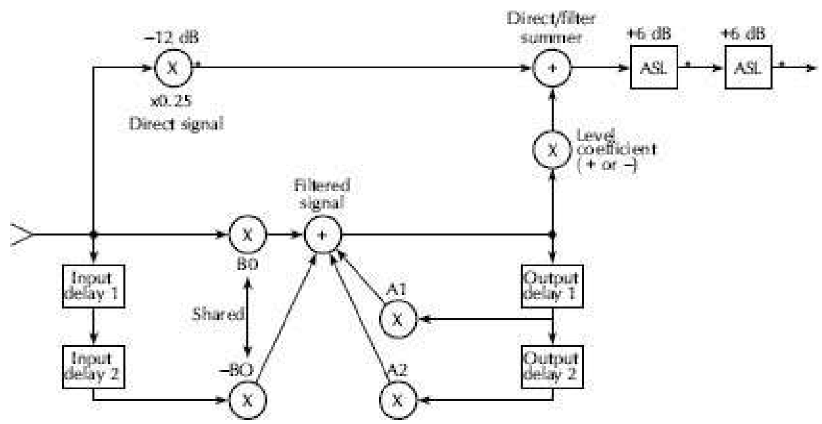


Figure 29-121. A bandpass parametric EQ stage.

EQ boost is achieved by adding in filter; cut is done by subtracting it away—a negative coefficient is thrown at the post filter attenuator instead of a positive one. There is a non-obvious criterion for cut coefficients—as one cuts, the effective Q of the EQ responder tends to sharpen; the frequency response of this arrangement at, say, 12dB of cut is not complementary to that at 12dB of boost; one needs to relate and modify the filter Q with cut level in order to take this into account and retain customary lift/cut symmetry.

Multiple sections of parametric EQ can be and usually are simply cascaded, although emulation of many classic analog designs has been better served by running the multiple filters in parallel and then adding their gained results all together with the straight signal. The band interactions are entirely different, offensive to a tidy mind, but far closer to the truth!

Given that most parametric EQs use bandpass filters only (at a push even shelving filters can be faked reasonably well using such) and that, as we've seen, bandpass filters have their b1 coefficient always at zero, it can make sense not to perform that multiply at all, thus saving data fetches and a multiply. Additionally, since the b0 and b2 coefficients are simply inverse of each other, only one need be sent from the host processor to the DSP, the inversion being simply achieved internally. This is welcome streamlining of the processing.

29.21.4 Shelving EQ

Real shelving can be achieved by using a full biquad in the EQ (as opposed to the simplified bandpass-only variety shown) with low-order high-pass or low-pass filter coefficient sets, or an even simpler structure as in Fig. 29-122. Much greater than a single-order response in the filter tends toward a frequency response with a "phase-bounce" in it near the turnover frequency, generally considered undesirable (except perhaps when one is being very picky emulating a Baxandall). The arrangement shown is a shelving EQ using very short filters. Advantage is taken of the fact that with single-order filters one can very easily create a high-pass filter merely by subtracting away a low-pass from a straight signal.

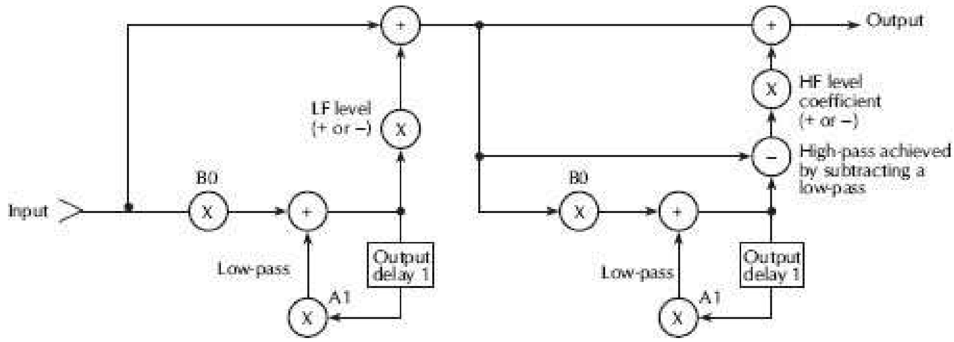


Figure 29-122. Shelving EQ using single-order filters.

29.21.5 EQ High-Frequency Response Anomalies

Odd phenomena occur when filters are attempted too close to half the sample rate, 24kHz in a 48kHz system for discussion here: partly as a result of the inevitable zero in response at half the sample rate with bandpass filters, as we saw, and partly an effect of the prewarping in the transform calculations used to create the filter coefficients. The effective Q of a filter as used in a parametric such as this appears to increase (become sharper) and become asymmetric (high side gets steeper) as its curve approaches Fs/2. Although this effect can be considered unimportant, occurring at the audible extreme as it does, this behavior can be improved by the expedient of applying a subsidiary correction to the desired Q value prior to warping, or by the more fundamental approach of oversampling. This basically means running the EQ (or at least the HF bits of the EQ) at twice the sample rate; upsampling to 96 kHz and downsampling (to get back to 48 kHz) are quite straightforward. This has the effect of pushing the squiffy zone up toward the new Fs/2 of 48 kHz, where it simply won't matter, keeping the normal audio-frequency range of EQ linear and tame. Under some conditions with some program material, upsampled EQ (even though subsequently brought back down again) can sound better. One has to be very careful with the nature of the reconstruction filters in the upsampling in order not to imbue even worse funnies in EQ frequency response than one is trying to fix.

Fig. 29-123A shows the squiffy effect on a 16kHz Q of 2 parametric EQ section; a similar Q of 2 filter at 200 Hz is shown for comparison. Correction (not oversampling in this case) results in the improved lower-frequency slope of the 16 kHz filter; this is now comparable to the skirts of the 200Hz filter, Fig. 29-123B. Unfortunately, there's not a whole lot one can do about that zero at 24 kHz without oversampling, so EQ close up to the band limit will always be a bit suspect.

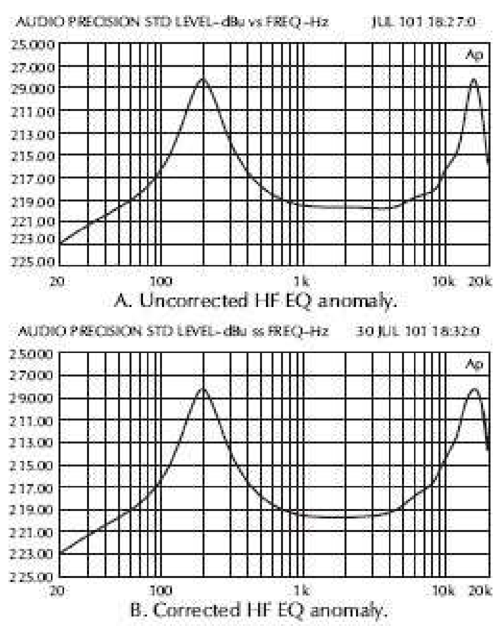


Figure 29-123. High-frequency anomaly.

29.22 Digital Dynamics

There are many approaches to dynamics processing in digital, but most fall under one of two categories: mapping and literal. Briefly, Mapping involves creating a plot, a table, or a map describing what the desired output level for any particular input level needs to be; an input sample comes along and based on its value, a gain-control value is picked out of the look-up table map that is to be applied to that particular value of input in order to create the desired output level. The map can contain the transfer values for many different sorts of dynamics processing simultaneously—say, compression and gate, limiter and expander, etc. Fig. 29-71 in the earlier discussion of dynamics gives a clue as to the structure of such a map.

Literal is building a digital processing equivalent of how one would literally achieve the processing in good old analog.

For all the fuss that is made about how DSP makes doing audio design harder, it's nice to come across aspects that make life so much easier and nicer. Here are two:

- ABS—a machine language mnemonic for ABSolute value. The effect is to look at a number and if it's negative, make it positive. It's the DSP equivalent of a precision rectifier, never the most trivial of analog design exercises.
- MPY (Multiply)—This is the most perfect, distortion-free, vice-free gain-control element. One will never, ever, want to play with VCAs or FETs again.

29.22.1 Mapping Dynamics

A look-up table is created in which the desired system gain is referred against the applied signal level. For a detected input signal, the corresponding gain value is "looked up". Simple indexing.

Look-up table dynamics have the strong property of being very fast in terms of processing cycles at the expense of the memory for an adequate map or set of maps. The precalculation has already been done, so all that needs to occur is the indexing of the look-up table from the value of the input signal—returned is the gain-control value—very short and to the point. Depending on the dynamic range over which the dynamics is to behave, the depth to which (how low in level) the dynamic behavior is to be adequately described (important for gates, expanders) and importantly the degree of resolution of the table (so that signal levels don't noticeably lurch from one value to the next), the size of the tables can get quite healthy. In addition, it is often convenient to actually run more than one table. Required memory usage may or may not be a problem—some DSPs have huge rafts of memory, while others, designed for less memory-intensive streaming audio applications, may have only just enough for the basics.

The map only describes the instantaneous gain value. Direct application of recovered gain values would result in awful distortion. Obviously some temporal constraints need to be added. Typically these are the classic dynamics values of attack and release and such. Where these time constants are applied is an interesting question. In order to use the usually relatively slow release time constant to smooth out the inevitable steps from the table quantization, this usually follows the look-up. If one were to be emulating a peak limiter, then one might well let the input signal directly pick its value from the table and then apply the usual short attack time constant to that. In other words, for a limiter, both attack and release would follow the look-up stage.

Compressors generally have a far more relaxed attack time, with the intention of deriving a signal more corresponding to the audio energy than its instantaneous peak. In this case the attack processing would take the form of short(ish)-term averaging or even rms-like detection; the result of this averaging would be used as the pointer into the look-up table. The release would be left on the output of the look-up, mostly for its role as janitor, tidying up the potentially ragged steps.

Assuming a compressor, a likely threshold range would be from −40dB below nominal operating level up to, say, 10dB above. Since nominal operating level is usually at or around −20dBFS, this implies that the look-up table has to encompass an input signal range of −10dBFS down to −60dBFS. It has to do this with sufficient resolution that no gain lurches are obvious. (Although most musical program material can withstand even comically large gain lurches under these circumstances, some—solo flute or a slowly decaying tremolo bass-guitar note spring to mind—will highlight painfully small ones.) Since the gain steps should almost certainly be dB linear or close to it, and the applied signal is linear, it is wise to perform a logarithmic conversion to the input signal to closer approach dB-to-dB mapping in the table. These tend to be computationally expensive or involve look-up tables themselves (!), but the penalties for not prelogging are either a look-up table to achieve adequate resolution at the lower levels (and −60dBFS is a long way down, to 1/1000 in fact) or reduced accuracy at the lower levels for a smaller map. A linear map for this compressor might need to be 2048 or 4096 steps deep to have nonembarrassing behavior near the bottom.

Big tables, actually big anythings, are bad news in the sense that if the parameters are changed (say the compression ratio is altered a notch) a whole whacking great new table has to be fed from the host microcontroller up to the DSP. An alternative, if the memory in the processor supports it, is to permanently have a suite of maps encompassing the range of parameters required. A different map is pointed to when a parameter changes. LOTS of memory!

A nice lateral thought solution to the really deep map problem is to create a good and concise map, or set of maps for various changes in parameters, and then to effectively move the threshold around by **scaling up and down the actual audio samples** accordingly instead. In particular, expansion curves can be created economically of memory; instead of moving the curves around, the audio is scaled instead to create the desired response. In a similar vein, only the audio within the sidechain need be scaled, the "real" audio being left alone except for the gain modifications determined by the scaled-sidechain lookup table.

A peak limiter, operating over a comparatively much reduced dynamic range, may possibly eschew a log convertor and just look up directly. On the other hand, an expander may have to adequately describe down to −90dBFS (or whatever the "don't care" level might be). Which brings to the fore another point, which is that if different time constants are required for different functions, as they certainly would be between a compressor and a gate, say, it could make sense to use a different look-up table for each.

29.22.2 Literal Dynamics

This is the technique of emulating (as close as one can) how an analog circuit achieves the required dynamic behavior. There is a bit more art in this approach, and although the algorithms tend to be longer and certainly more intensive than mapping, there is very little memory usage, and changing parameters just involves sending a handful of coefficients to the DSP from the host, rather than potentially thousands.

It is possible to emulate the rough-and-tumble free-for-all uncontrolled servo-loop behavior of a feed-back-style compressor/limiter, or alternatively plod through the tidy-mind deterministic feed-forward VCA approach, which involves division and/or much log'ing, antilog'ing, and untold processor time (transcendental functions are very long-winded in DSP), for ultimately a well-behaved but, frankly, bland result. (Guess which the author finds more fun?) Filling a whole DSP with such a VCA-like processor isn't difficult.

There is just as much latitude for approach with literal dynamics as there necessarily has been with analog design; indeed, if one's goal is to emulate classic analog dynamics this is really the only way to go.

29.22.2.1 A Simple Digital Limiter

Fig. 29-124 highlights how dynamics signal processing in DSP—in this case a simple peak limiter—can almost slavishly follow an analog architecture.

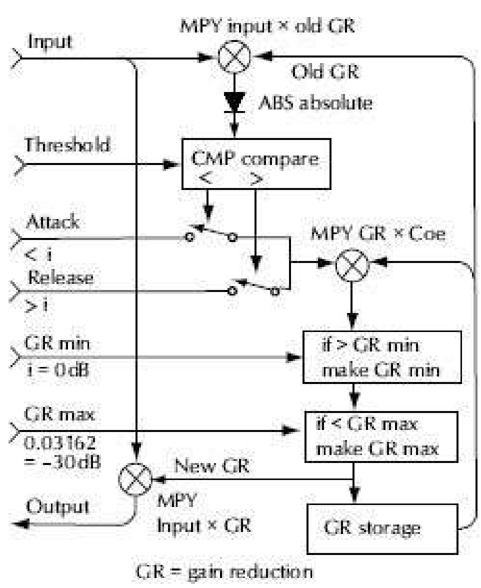


Figure 29-124. Flow chart of a simple digital limiter.

The key to the limiter's operation is the gain-reduction value—sorry, the author still thinks of this as a control voltage. Remember that multiplying a signal by 1 doesn't change the signal; multiplying by a fraction less than 1 reduces the output signal, i.e., affects gain reduction, which is what we need when a limiter is biting.

First, the immediate present (new) sample is multiplied (MPY) by the stored GR-generated last sample. This is necessary to judge whether and which way this last GR value needs to be adjusted for the present sample. The absolute value (ABS) of this modified input sample is then compared (CMP) to the threshold coefficient. If it is greater than the threshold the GR value needs to be reduced, and the program branches into attack, where the old GR value is multiplied by a coefficient usually just slightly less than 1. Likewise, if the threshold isn't breached the GR value can be relaxed, so it branches off to release, where it is in effect multiplied by a coefficient just ever so slightly greater than 1. This is shown in the diagram as switching between using the attack coefficients, or the release coefficients. Naturally, the modified GR value has to be clamped such that it can't rise higher than 1 (and so no longer be GR!) and that is also the normal unity "resting" case.

The coefficients for attack and release are in this simplistic case (it can get considerably more complex!) multiplicative—the GR value is changed by the same proportion or, in other words, the same number of fractional dB per sample. Running at, say, a 48 k sample rate, in order to have a 1 dB/s release rate the coefficient would have to represent 1/48,000dB increase in GR value, or about 1.00005. "Slightly" says it all. This dB-per-time gain trajectory works quite well in audio, emulating the dynamic response of many good analog systems.

The last things that happen are that the newly modified GR value is saved for use next sample and also used to multiply with the present input sample to create the gain-reduced output sample.

All in all, it is almost an exact parallel to a simple analog feedback-style limiter; complexity concessions exist for operating in a sampled-time system (such as the initial input/last GR premultiply), as opposed to relying on the always existent signals in the continuum of analog. On the other hand, one effortlessly achieves true dB/time gain rates for attack and release, usually a feature of posher analog designs and only ever approximated in simple systems.

29.22.2.2 Feedback-Style Limiting and Compression

Unlike most analog GR elements, the “MPY” in a DSP is directly linear in operation, i.e., a gain-reduction value of say 0.3 will cause the signal through the multiplier to be reduced some 10dB. It is not linear-by-dB, like a VCA, or a mangled exponential/logarithmic like thing such as from using a raw semiconductor element such as a transistor or FET. Yet, as has been shown above, linear-by-dB results can be achieved fairly simply. Emulating other laws can get rather interesting, but are certainly attainable, in pursuit of a sound. Similarly, the determination of the amount of instantaneous feedback in a feedback limiter depends on many things, not least the attack and release time constants necessarily applied to it. Another is whether the control signal is being generated all the time and only applied when the threshold is exceeded or alternatively if the control-signal determination is only woken up when the threshold is exceeded. Both can work well, and both sound utterly different.

Feedback-style compressors can use basic limiters as above as a starting point. The limiter (using the required attack characteristics of the compressor as its attack/release time constants) creates an overage signal implicit in its own control signal, representing the amount the input signal is exceeding the limiter threshold at a given moment. By manipulating this overage so as to create a control signal more in accord with a chosen compression ratio rather than the hard limiting, a suitable release time constant applied, the doctored control signal is used in a second multiply on the untrammeled input signal, outside of the feedback loop. As an approach, this combines the edge and sound of a feedback-style dynamics unit with a sane deterministic compression ratio. Fig. 29-125 shows a family of deceptively analog-looking input output curves from a digital soft-knee compressor using the described technique.

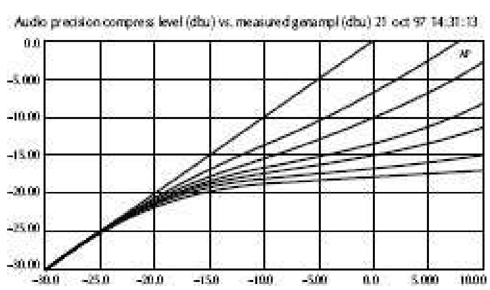


Figure 29-125. A family of compression curves from a digtal dynamics section.

29.22.2.3 Gating

The purpose of a gate is to attenuate completely or partly a signal that falls below a given threshold. Typically they should wake up (open) quickly, hang open for a while if the signal goes away just in case it really hasn't, and then close at a gentler rate. Also, to prevent "falsing," there are two thresholds, one for opening and the other, slightly lower in level to determine closure. Written as described above, it is about as "digital," yes/no, a set of conditions as one can ever hope to meet in audio and is a complete natural for the literal approach.

The absolute value of the input signal is compared to the open threshold; if tripped, a target control signal of 1 (unattenuated) is applied to a short low-pass filter bearing the attack (open) time constants feeding the attenuator multiplier (this will quickly ramp up the gate to open); at the same time a counter is initialized. The counter is the hang-time counter. It is reinitialized at every sample that the close threshold is exceeded such that it doesn't get a chance to start counting down unless the signal really has gone away. If that occurs, and the counter does count down to zero, a control signal value of zero (for off) or some other value representing an amount of off attenuation (depth) is applied to a longer release time constant low-pass filter, the output of which is applied to the attenuator multiplier.

Fig. 29-126 shows the dynamic transfer characteristics of a microphone input using a combination gate/soft limiter; this combination is used extensively, in this case for a stage backup vocalist. If the singer is making enough noise, the gate opens (actually it relieves 14dB of attenuation, which is enough to make stage spill go away enough), and the limiter almost immediately takes over, keeping the voice at a manageable, consistent level for the mix. Note the 3dB of makeup gain. Lest one is concerned that this combination is far too unsubtle for quieter songs, remember that this is a digital process and in the context of a programmable console can be (and is) reprogrammed to suit on a song-by-song or even section-by-section within a song.

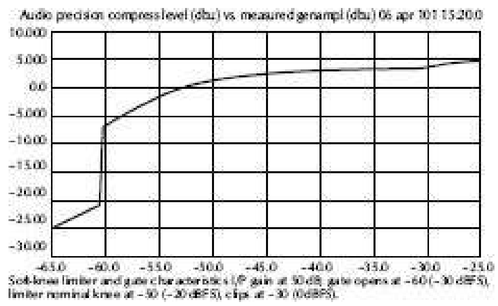


Figure 29-126. Soft-knee limiter and gate characteristics.

29.22.3 Inadequate Samples for Limiters

By and large, Mr. Shannon and Mr. Nyquist did great jobs. Digital audio works really well and it is indeed possible to reconstruct an indistinguishable result from a source through a system sampling barely twice as high as audio bandwidth. But there are a couple of places where a limited number of samples trips up—In particular, attempting to sense the peak of a signal, as one attempts to do with limiters and gates, both of which need to respond accurately to them.

Unfairly and unreasonably (since it isn't terribly relevant to most real audio) we'll consider first the example of Fs/2—half the sample rate. With only two samples per cycle of sampled signal it is entirely possible for the two samples to miss the signal altogether, if they happen to occur at zero-crossing points of the applied audio signal. But, then again, they might hit the jackpot at the crests.

More realistically, look at the two extreme cases of Fs/4, or 12kHz for a 48kHz sample rate, in Fig. 29-127. Nobody is arguing that 12 kHz isn't audible, yet here is a case where there can be as much as 3 dB error in sensing a level. There are similar, if far less serious points of error dotted throughout the audio spectrum (e.g., Fs/8, 6 kHz etc., or anywhere else where Fs is divided by an even number). Now, to be completely fair, under any reasonable circumstance this effect would not be excited and certainly not be audible, since an exactly 12 kHz tone in isolation is not by and large terribly common or useful—And would be a far-reached argument in support of a blanket increase in sampling rates for digital audio. However, in the specific case of attempting to peak-sense audio levels, one comes across these spot frequencies with too little effort. This is a reconstruction error, or more precisely, an error due to the samples not explicitly describing the signal but relying on a later reconstruction filter to fill in the gaps. Back to analog, the signal reconstructs just fine! This is just a hint of the occasional disconnect between the two domains.

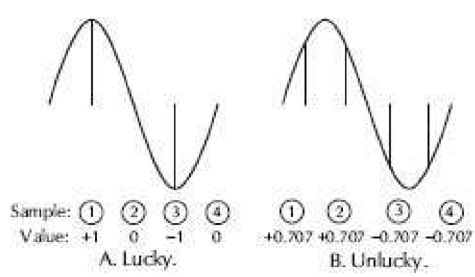


Figure 29-127. Sampling Irregularity at Fs/4 (12 kHz); possible 3 dB error.

A second—and actually practically more worri-some—reconstruction error effect is if through clipping or heavy dynamics processing a pair of adjacent-in-time samples are made full-scale, a downstream reconstruction filter will cause a significant overshoot beyond full scale in the recovered analog signal. This can cause clipping in the following analog stages if insufficient head room is allowed, and a nasty surprise to anyone who thinks full scale is the most one can see out of a digital system!

By way of a slightly different practical example of sampling oddities, it was noticed that digital limiters seemed to respond differently each time to a snare-drum impulse; sometimes they'd catch it hard, sometimes they wouldn't. In comparison an analog limiter just, well, caught it. Now, snare drum is pretty evil, a nasty big initial short spike. On analysis, sometimes the spike was adequately described by the limited number of samples, sometimes it blew through the gaps. There wasn't much difference, between 3 and 6dB in captured peak level, but that is plenty enough a difference to be audible and, more to the point here, plenty enough to invalidate the devices as peak limiters!

Oversampling—i.e., making the sample rate twice or even more times higher—has the effect of pushing the worst of the reconstruction-error potholes out of the relevant audio band. Even though one is operating on exactly the same audio data originally sampled at the lower inadequate sample rate, peaks are captured accurately enough for all practical purposes. The missing peaks of Fig. 29-127 are actually being filled in by the reconstruction effect of the low-pass filter employed by the upsampler, in exactly the same way as a D/A's reconstruction filter would.

As such, there is a very strong argument for oversampling in peak-limiter and other fast dynamics processing. Since it has one tightly defined purpose, the impact of doing the sample-rate conversions does not have too great an impact on DSP processing cycle budgets.

29.22.4 Predelays, or Look-Ahead

Applied only in rare cases in analog because of the difficulties in providing for good audio delay, predelaying is eminently achievable in digital dynamics sections. Predelay is the technique where the main signal path through the dynamics is delayed for a short period (1ms, 2ms or so) to allow the side-chain processing to determine the right amount of gain reduction to be applied; this value is then applied to the main signal path in a gain-control element discrete from that used in the sidechain, Fig. 29-128. The prime use is in peak limiters (which are nearly always feedback style, even where the other sections may be feed forward), where overshoot, which can occur during this onset settling period, can be completely avoided. An improvement in sound results, too, since very hard and brutally short attack times can be mellowed out knowing that overshoot is not going to increase as a result. A relatively soft attack time (for a peak limiter) of 1 ms combined with a comparable predelay captures the peaks without the need for subsidiary clipping and yet is sufficiently aggressive that it retains its loud characteristic but without the usual telltale ripping hard edge."

Look-ahead limiting is extensively used in broadcast air-chain processing, and especially on feeds to streaming compression codecs (AAC, MP3, or HD radio, for example), which generally do not react well to the artifacts generated by more conventional clipping or unavoidable transient escapee overloads from ordinary limiters.

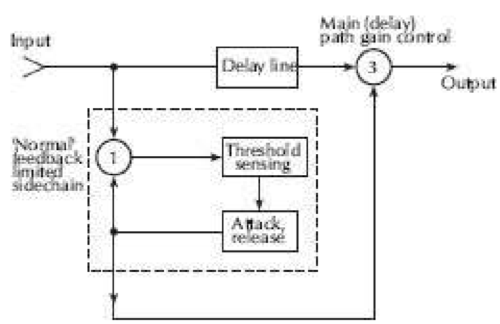


Figure 29-128. Dynamics section pre-delay system.

Only occasionally would such processing be done in a console channel, but when it is, it should be remembered to apply equal delay (whether limiting or not) to other contributory channels in the mix.

29.23 Digital Mixer Architectures

Two distinct approaches seem to be taken to the signal-processing architecture in mixing consoles, probably stemming from how deeply steeped in traditional computer science the designer is.

29.23.1 The Sea of DSPs Approach

In this, a large enough array of DSPs for all envisaged processing is closely coupled to enable the rapid transfer and sharing of data between them. The "tank" is fed with all sources, and all destinations are taken from it. In a telephone-exchange kind of approach, signals requiring processing are farmed out to other processors in the tank and the results returned; it could be regarded from the outside as one big processor. The main advantages of this approach are that not being physically constrained to a particular organization, reconfiguration is straightforward; any signal can go anywhere at any time for any purpose. If more processing is needed it is merely attached to the busing system, growing as required. The major downside is that all the flexibility makes programming such a beast very difficult (the word nightmare has been bandied around).

The "one big processor", is of course a reality for many modest-sized applications, in the form of the humble, or increasingly not-so-humble, PC. These have increasingly faster and more capable processing cores, and multiples of those, too. Although they have their own issues as far as processing audio (e.g., they generally don't do DSP terribly efficiently), brute force, speed, and might make up for those. A major positive is having everything under one roof, dispelling the problems of multiple-part interconnectivity.

A major drawback to the Sea-Of-DSPs (whether DSPs or PC) is a lack of scalability. When it's full, it's full, and there is a high incremental cost to any expansion, which may even be practically insurmountable. Given the nature of all products is to fill available capacity +10%, this is ignored at peril. PC-based DAW users in particular are quite used to having to keep an eye on the "gas gauge".

The second approach very much follows the signal-flow approach of a conventional analog mixing console, using multiple DSPs as required, processing being applied in-line as and where it is needed. On very large consoles the mixing processing itself can take on the look and feel of a tank-style array, but other than that the layout of the signal paths has remarkable parallels to analog.

29.23.2 A Practical Digital Mixer

As with the discussion of analog consoles, which revolved around the description of elements of a particular design, so this section uses as its basis the architecture of a real digital mixing console. It is shown in its basic form. As the reader is well aware, pin-by-pin details of implementation, bells, whistles etc. can rapidly mushroom and the weight of resulting detail tend to obscure; as it's not too difficult to figure out how most of this is done, they have been omitted for clarity. It is important to remember that in terms of lines carrying audio signals, it is accurate, due to the use of the serial audio format outlined below. Shown here in a mid-size 64-by-24 format, this particular design's premises were simplicity and scalability (it can be readily made bigger or smaller) and has proven to be robust and reliable, using no scary technology and with nothing running on the edge. Over the years this basic architecture has grown and evolved through generations of increasingly powerful DSP and support devices with the odd effect in this blighted world that it has actually become progressively simpler and cheaper to build with time. It is actually at the heart of many, many thousands of consoles. Also the steady and welcome improvement in integrated converters has resulted in the overall performance blossoming to the extent that this, along with other digital mixer designs using comparable technology, owe nothing to analog in performance whatsoever.

It is assumed, of course, that the control surface has been undertaken as the separate design exercise that it largely is; this discussion concerns the signal-processing side of things.

29.23.2.1 Serial Audio Format

Nearly all converters and like peripherals such as AES/EBU format transmitters and receivers use in common a serial digital interface; this is usually set up as to be two sets (left and right) of 32 data bits per sample frame (64 total), meaning a data rate of 3.072MHz (for a 48kHz sample rate). This is a very tame and robust rate and can be run around quite happily without fear of corruption, and as such is used as the nearly sole means moving audio data around in this console. Adopting this serial format also minimizes the amount of data format changes required.

29.23.2.2 Inputs

Input signals are applied to whatever form of convertor or interface is required: microphone amplifiers or line-level inputs into A/D converters, AES/EBU into AES receivers, and subsequent sample rate converters. Sample-rate converters (SRCs) are necessary since it is unlikely (unless a whole amount of trouble has been gone through to synchronize the whole system of which the console is a part) that other digital sources will be and remain in word/data-rate synchronization with the console. The recorder may well be, but typical AES/EBU devices, such as outboard effects, or remote sources, rarely will be. If it is considered necessary (on the basis that anything that messes with data unnecessarily is a bad thing), the SRCs may be bypassed for sync'ed sources, but frankly SRC's today have artifact levels so low as to be considered quite blameless.

At this point, all the data is in native format (the convertor serial standard), travelling in pairs—mono signals (microphones, say) in pairs and stereo sources as left/right pairs per data line. For a 64 input console, this means 32 data lines.

The channel signal processing is done four channels (two pairs) per input DSP, Fig. 29-129. The DSPs used here very conveniently have native format inputs and outputs (being designed to work with normal converters), making interfacing really simple. They are also easily powerful enough to do four well-featured channels worth of signal processing, more modern ones can do many more. Typically, this would be high-and low-pass filters, a four-band parametric EQ and limiter/compressor/gate dynamics, and delay (memory is attached to the external memory interface of the device to support this if required). The channel DSP has spare input and output capability, which can be implemented if required as selectable direct channel outputs, keying inputs to dynamics, etc.

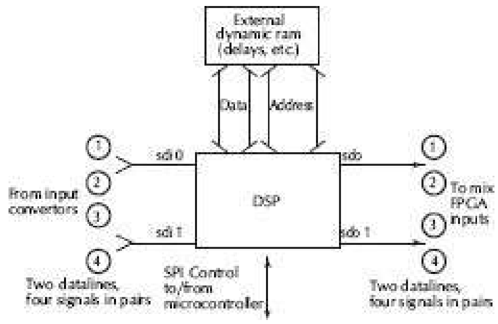


Figure 29-129. An input DSP (one of 16).

29.23.2.3 Mix Stage

The 64 channel outputs are taken from the 16 channel processing DSPs as 32 output lines and applied to the mixer stage(s), Fig. 29-130.

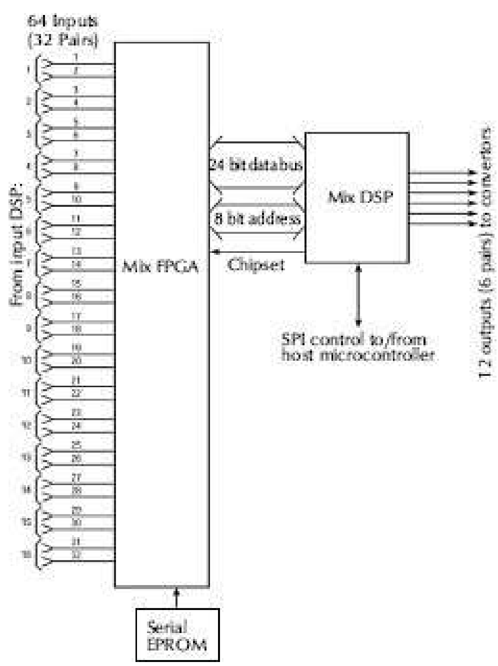


Figure 29-130. Mix stage.

The ominously large device labeled FPGA (field-programmable gate array) into which all those lines disappear is programmed to be merely a (large) collection of serial-to-parallel data converters—no voodoo. A slightly simplified version of its contents is shown in Fig. 29-131. The FPGA takes each data line and puts it into its own 24 bit long shift-register; when it has counted that the necessary 24 bits have arrived, it seizes the data and tells the mix DSP with which it is associated that the data is ready for harvesting. (A long-ago prototype of this design actually used discrete logic shift registers. Lots of them. It was huge. FPGAs are much better.) To the DSP, the 32 shift register outputs are arrayed and addressed to look exactly like memory, and indeed, the FPGA sits on the 24 bit wide external memory bus of the DSP, with enough address lines to uniquely address each shift register location. Once informed the data is ready, the DSP copies the data values down into its own internal memory, from which the mix code accesses it. Although this can be done in real DSP software, it is usual to invoke a DMA routine (direct memory access) that, depending on the sophistication of the chip, can transfer data quietly in the background of normal processing from one area or peripheral into/out of internal memory with minimal impact on normal operation. In practice, it always seems to slow things up a bit (background is a relative term, it seems), but overall, DMA is slicker. The FPGA/DSP DMA combination does this transfer operation twice per sample, once for left data, the other for right. These two sets of data are held in buffers in DSP memory so as to be in time alignment ready for the next pass of the mix code.

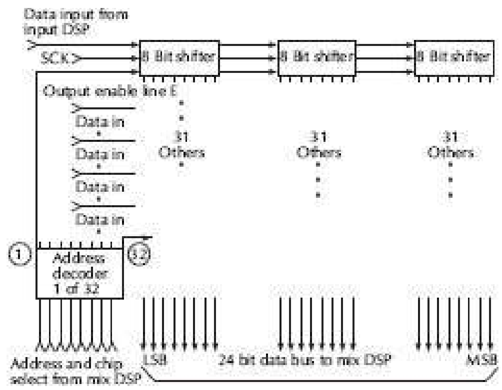


Figure 29-131. Simplified contents of FPGA.

29.23.2.4 Mix Code

As earlier mentioned, DSPs are designed to do some functions really well and one of those is the FIR filter. This involves multiplying a piece of data by a unique coefficient, adding the product into an accumulator and then rapidly moving on to do the whole thing all over again (next data point, next coefficient), and again, for as long as the filter may be. Well, from a mixing point of view, a group output multiplies an input channel sample by a unique coefficient, adds the product into an accumulator, and then rapidly moves on to do the whole thing all over again (next input sample, next coefficient), and again, until it's done all the input channel samples. Got it? A mixer and an FIR are as far as the DSP is concerned chemically indistinguishable. For each mix group in turn the DSP addressing runs through all the input data in turn, multiplying each by its appropriate gain coefficient. Automated addressing in the DSP keeps track of which coefficient goes with what sample. It is about as efficient a processing operation as it can possibly be.

Although bigger and juicier DSPs and FPGAs can radically increase the input, output and mix capabilities, the following description better highlights the practical design considerations:

Crosspoints. Each channel-to-group calculation is a crosspoint (from the concept of the mixer being a big soft matrix). A small 150 MHz DSP has a little over 3000 processing cycles at 48kHz, but necessary programming overhead precludes the use of all of these for mixing. Around 2000 crosspoints is perhaps more realistic. So for 64 inputs, 32 output mixes per DSP is theoretically possible; 32 sources, 64 outputs; 128 sources, 16 outputs, and so forth. (Naturally, bigger, faster DSPs expand these capabilities, but if enough is enough, why bother?) This input vs. output scaling elasticity has limits, however, as follows:

Input Bounding. This is the limit on how many sets of input data one can realistically capture and shovel down into the DSP and still leave it with processing time to do any mixing. Even DMAs have some time impact—in general, DMA notwithstanding, the more time taken dragging data around, the less time available for processing. External memory fetches take time (access to such memory is much slower than to internal memory—hence the need to bring the data down from the FPGA rather than access it directly for the mix). Even though it is theoretically possible to spend an entire sample period wheeling in new data in the background for calculation in the next, data management can get pretty hairy. In comparison the actual mixing is a doddle.

Output Bounding. This is less of a problem, since by and large there are fewer of them. But expecting a large number can lead to an issue of how to get all those mixes back out into the world.

Since simplicity was a major aim of this particular design, the outputs of the mixer stage are taken as six pairs (twelve groups) from the DSPs in-built serial interface; these group outputs are applied in the same mix'n'match fashion as the inputs, directly to whatever class of output device is necessary—D/As for analog outputs, AES transmitters for digital, etc. Deriving more outputs from the DSP involves getting those mixes back up into the FPGA—again by DMA—and doing a parallel-to-serial conversion of each there, in reverse fashion to that done on the inputs to the mixer stage. By such means, the modest processing power in this mixer core can easily handle a 64-by-32 console. Here, the twelve buses not coming directly serially out of the DSP are done in this manner.

Increasing the number of mix buses yet further would be achieved by using another FPGA/DSP core, allowing a further thirty two mix buses. The second mix stage's FPGA is simply parallel-fed from exactly the same thirty two data lines from the input channel processing DSPs as the first mix stage.

The fact that the mix outputs are in serial native format dramatically facilitates the dropping in of a further DSP for post mix processing if required for some applications—graphic EQs and group dynamics sections, for example.

As can be seen from the diagrams and the description, the signal flow of this console is in such striking accord with an analog implementation, one can rightly wonder what all the fuss is about.

FPGA's are becoming increasingly faster, more capacious, and capable with the addition of on-chip RAM, dedicated multipliers, instantiated microcontrollers and such. A mixer of modest proportions, such as this design, is implementable directly into an FPGA alone with no need for the DSP. Currently, it is a cost-benefit design exercise, deciding whether a capable enough (and more expensive) FPGA is worth it over the low-cost DSP and cheaper FPGA. But the trend is clear.

29.23.2.5 Universal Mix Buses

The described design provides a large number of raw mixes, with no mix-specific hardware or code. It may be noticed that apparently an otherwise vital subsystem to the mixer appears to be missing—monitoring. Well, actually it isn't, and the fact that it is implicit in the design as it stands points out an approach and attitude to mix buses that would be hard to maintain in analog where every bus is a significant expense: in digital, buses come cheap.

Monitoring in this case commandeers a pair of mix buses (assuming stereo); think PFL bus for now. Any input to the mixer can be monitored on this bus by applying an on coefficient to the appropriate crosspoints to bring the source(s) onto the bus. So far so good. But for monitoring output buses (stereo group, auxes, any of them actually) rather than apply the analog solution of a selector to switch between those existing groups, what one can do is exactly recreate the mix to which one wishes to listen; if one were to apply the same coefficient set that is making, say, auxiliary bus 5 to the monitoring bus, too, then one will exactly recreate what is happening on aux bus 5 in the monitoring.

Where this approach really shines is main mix bus monitoring—one can mess about with the monitoring bus as much as one likes without affecting the main mix at all—nondestructive soloing becomes a reality, simply implemented at that.

Talkback is simply treated as one of the sources to the mixer; it can get routed into any of the mix outputs with no necessity of creating a separate subsystem, with IFB (Interruptible Foldback) talkover ducking or muting deriving from modified coefficients.

Note there is no distinction made between the group outputs as to what their ultimate purpose will be, group, aux, cleanfeed, etc. All that distinction is done at the control surface and the interpretation of its requirements by the host microcontroller—in other words the differences are all in the controlling software and not in the hardware that implements the mixes.

29.23.2.6 Coefficient Compounding

This is a rather fearsome title for a rather nice concept. This is how master fadering and group fadering are achieved. Rather than have a separate downstream gain stage after a mix has been achieved to effect overall level control, a convenient approach with a soft matrix mixer such as has been described here is to take the sensed level of the real, physical, group fader and then multiply each of the coefficients feeding that particular bus by its value. This is a direct analogy of VCA grouping, where one fader actually modifies the level contribution of each source to the mix bus, rather than gain changing the mix after the fact. Since all of these numbers (source contribution coefficients and group fader) exist in the host microcontroller, the arithmetic manipulation is quite straightforward. The database management aspect of this on a large console can get quite interesting, but this pseudo-VCA grouping approach is widespread and very powerful.

29.23.2.7 Coefficient Slewing

Rapidly altering coefficient data in a DSP runs into exactly the same tone click problem as do MDACs in analog; even small transitions made when the audio data sample is nonzero stand a very good chance of being heard as a click. A fader swipe can generate the all-famous zipper noise, and just as with MDACs, without care and attention the effect in EQs is little short of comical.

Sensing zero-crosses in digital is practically impossible, since particularly at high levels of high frequencies there may well not be any samples anywhere near zero—remember this is not a continuum like analog, the samples are just a regular set of stabs in the dark. A wide enough window to capture enough zero-crossings would probably be wide enough to still allow some transitions to be audible. Never mind the fact that the processing overhead for doing a window compare and decision on each and every coefficient would be overwhelming; it would probably cut the potential number of crosspoints in a mix stage down by an order.

A good solution is to allow the DSP to ramp relatively slowly between its present value and the new desired value, creating its own interpolating steps on a sample-by-sample basis small enough that each is inaudible. (This, by the way, is one of the necessary processing elements that eats up a chunk of mix-DSP cycles, limiting the maximum number of crosspoints available to significantly less than the raw cycles availability of the device would suggest.) A slightly different approach is to “pre-slew” the coefficients in an intermediate processor (often also a DSP) to offload the effort from both the host and the target DSPs. The inter-DSP communications can start to get a bit fierce, however.

It is a nerve-wracking moment when first trying on-DSP slewing. After all, the coefficients for IIR filters such as in EQs can be very, very touchy and have little tolerance for error before doing very odd things most unlike the filters they were intended to be. Amazingly though, it seems as though provided the filter set is stable where it starts, and stable where it ends up, it stays stable in between as the coefficients are slewed; it might get just a little wonky, but not enough to cause any serious sonic issues and certainly not enough to explode into what has been charmingly called “screeching cats from hell” (DSP audio guys and gals hear lots of them).

29.23.2.8 Clocking

A major subsystem within a digital mixer is clocking—making sure that each of the various circuit elements get the necessary hard, clean clocks required to operate properly. In this design alone there are six clocks for processing: 12.288MHz master clock (actually divided down from 24.596MHz to ensure symmetry), 6.144MHz used as a master clock by AES/EBU transmitters, 3.072MHz as the main serial bit clock for the standardized native serial data format, an inverse of that used by some A/D or D/A converters of less serial format flexibility than others, then 48 kHz, which of course is the data sample rate and house left/right clock.

Although there is from a component-count standpoint a tendency to want to include the clocking generation in with an existing FPGA, say one from a mix stage, it can be beneficial to have it stand alone in a smaller FPGA or CPLD package. Generally, each clock feed to each device should be individually buffered and be as close to its target as possible. Needless to say, this takes a lot of FPGA/CPLD pins, and a single-purpose device starts looking like a good idea. The major benefit is that one can physically locate it where it can do the most good; this is as close as one can get it to the A/D, D/A and sample-rate converters. Ideally (but rarely is it possible) these should all be clustered in a “convertor ghetto” to keep the clock lines really short and tight from the clock generator, which minimizes noise and slewing on the various clocks, which can directly affect convertor jitter noise performance.

29.23.3 Signal-Processing Control

Fig. 29-132 outlines a typical control architecture for signal processing, or the processing end. It should be considered along with Fig. 29-106, which shows the control-surface end. The separation reflects that often the processing and the control are, indeed, in separate places interconnected by a network.

29.23.3.1 Controlling the DSPs

Each of the DSPs has an SPI (serial peripheral interface) port, an industry-standard means of device intercommunication. This consists on each device of a serial clock line, which synchronously clocks data in or out, and a serial data in line; these may be paralleled around all the DSPs. A serial data outline needs to be selected in a multiplexer for feeding data (such as metering information) back into the host processor. There is also a chip select line, which needs to be run individually back to the host; when yanked, a particular DSP knows that the data being clocked out on the serial data line is for it.

It is down this SPI interface bus that the DSPs receive their boot code at turn-on (the program code which it will run), a set of working coefficients (usually those that were current when the console was last turned off), and any changes to those coefficients as the console is being operated and parameters changed.

29.23.3.2 Metering

The indication to the user of the various channels' and groups' signal levels, dynamics gain reduction values, etc. is performed by the control-surface host, driving the appropriate indicators. How the data gets to that micro from the DSPs that are doing all the work can vary widely in implementation depending mostly on the physical configuration of the console. If it is a single box, with the signal processing under the hood of the control surface, then metering data can best be taken simply and directly from GPIO (general-purpose input and output) pins on the actual DSPs. Alternatively, but it is giving up a major advantage of the one box, the host micro could recover all the metering data from the DSPs and distribute it all accordingly. If, though, the console is split, then a means of harvesting all the metering data from the DSPs, squirting it all up to the control surface, then disseminating it appropriately definitely has to be devised. This one sentence describes something that has many, many times been hopelessly underestimated, and at least in one historical case required a whole separate Ethernet run back to the control surface purely to handle metering.

In this design's case, the host micro polls each of the DSPs in turn, clocking back the metering information from each through the return path of its SPI; a packet is created of each complete console wide set, which is then delivered back to the control surface.

The good news is that metering data is not needed at anything like audio data sample rates. The feeds have been prefiltered in the DSPs with appropriate time constants and updating the relatively small data (8 bits is plenty) relatively slowly (better than every 25 ms or so) is adequate. Nevertheless, unless the polling by the host is under rigorous and deterministic control and the total bandwidth of even this fairly slow, small data set is carefully considered, the metering can start to be a major burden.

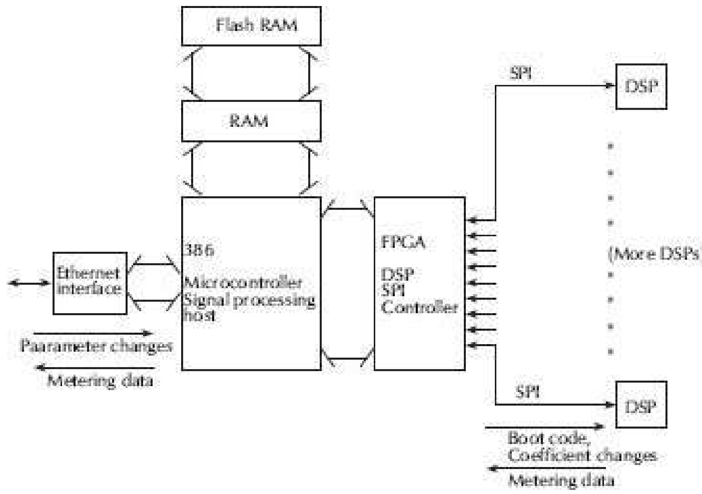


Figure 29-132. Signal-processing control architecture.

29.23.3.3 Host Microcontroller

This is usually a fairly fast and meaty micro, often of the ×86 persuasion. In the case of a self-contained console (control surface and signal processing all being in the same box) this will in all likelihood administer the control surface and displays in addition to the relevant function here, which is riding herd on the DSPs.

The host's job is to turn control parameters (as generated by the control surface) into coefficient sets that the DSPs can understand to perform the effect of those parameter changes. This would be by the coe-gen, or coefficient generation, software (typically written in C) and is equal in importance—a fact little appreciated—to the actual DSP code the DSPs are running. It is the coe-gen code that just as much determines how a console runs, feels, and sounds—after all, the DSPs are just doing what they're told and running code sent to them, by this host. By ways of example, the coe-gen code looks up what mix DSP crosspoints need to be modified to what coefficient values in response to a given fader being moved to a certain level and to take into account any mastering overlays, pseudo-VCA subgroups, etc.: what coefficient values to create for a parametric EQ section changed to differing parameters of frequency, level, and Q. In addition to being a fraught exercise in database management, there's some pretty good math in there too. (In DSP software design, one strives to keep the actual DSP algorithms as straighforward [fast] as possible, leaving as much squirrely and calculation-extensive stuff as possible to the coe-gen code in the host.)

Since a major part of the thrust toward digital consoles has been their promise of storage and recall, statically (snapshot) or dynamically (as in real-time automation), it is beholden to the host to manage the data transfers involved. Everything that may need to be stored is already in the host, but the software routines and hardware to facilitate storage/recall need to be present. A console can be quite self-contained in this regard if the data set is relatively small; on-board flash memory may suffice. Otherwise, whirling and whining hard drives may well be necessary. In the event the console is integrated reasonably closely with an audio recorder, hard disk, or otherwise, the automation data may get squirted onto that as a sideband, a concurrent rolling EDL (Edit Decision List).

In a split console (control surface is separate from the guts), the host also has to manage intercommunication with the control surface; typically this is done with an Ethernet variant, which demands the existence of a TCP/IP stack for the communication protocol, and hardware to terminate the Ethernet.

29.24 Digital Audio Workstations (DAWs)

Fig. 29-133 shows the major (usually indivisible) elements within a digital console system and their relation to each other.

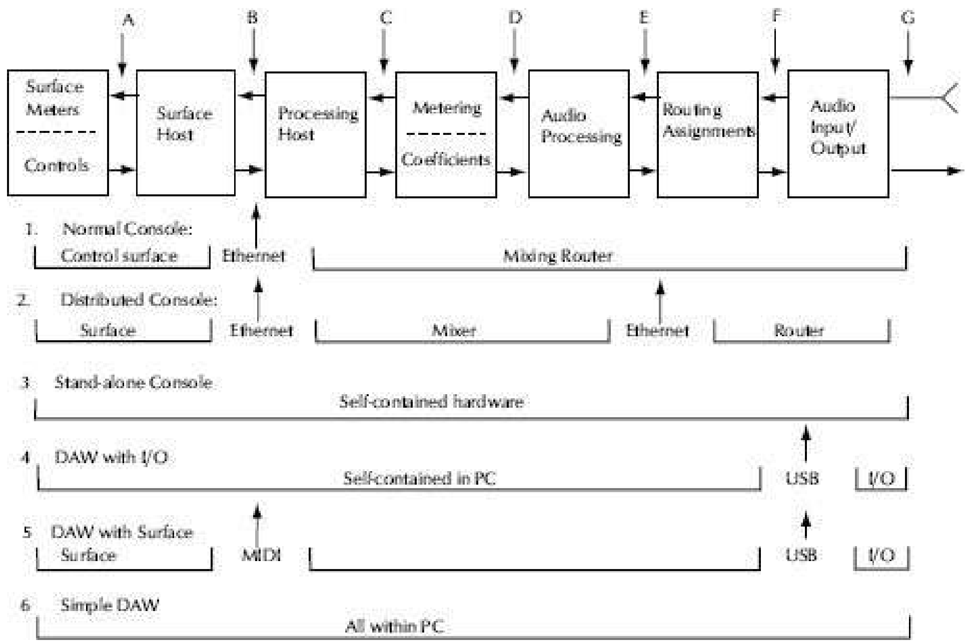


Figure 29-133. Process distribution in various styles of digital mixers.

User Surface. This has indication of control positions, metering, and means of controlling the audio processing and can range all the way from the sea-of-knobs large-format console-style surface to graphics on a PC screen with a mouse or touchscreen.

Surface Host. A micro to look at, make sense of the controls, and drive the metering. This can vary from being a small embedded micro to a large PC-like processor, depending on the size of the surface and if it is expected to do high-speed communication, should the control surface be remote. It can also not exist, its nonetheless necessary functions being subsumed by a PC's CPU processor.

Processing Host. Takes care of looking at the data being passed to it from the surface host, creating the necessary coefficients for the audio processor, and also looks at the various and many metering returns from the audio processor, rendering them down into a form useful for the control surface to display. This is usually a sizable and capable processor, unless its functions are being done by a PC's CPU.

Audio Processing. This uses the coefficients from the processing host to modify the audio path(s) as desired, usually within a raft of DSPs, either aided by or supplanted by FPGA's, unless, of course, the audio processing is within a PC's CPU.

Routing Assignment. Decides what audio path is going to get routed through what path within the audio processing and so what controls apply to it. Also decides on what output ports processed audio appears. Often an additional soft matrix done in a DSP mixer (particularly when integrated with Audio Processing), or a hard switch in FPGA, and often a stand-alone product but entirely within the capabilities of a PC's CPU.

Audio Input Output. This is the termination of audio sources and destinations and their conversion into a form of digital information the Audio Processing can assimilate. Examples are A/D and D/A converters, usually many, and SRCs (Sample Rate converters) to seamlessly integrate external digital audio sources. Often stand-alone, in cages dedicated to their purpose, locally or separated from the rest of the system. Often integrated with the Routing/Assignment, and sometimes along with audio processing. And sometimes just plugged straight into a PC. (Bet you can't guess where this is going...)

The paths between any of these blocks may be broken and subject to transport if need be, but it is far more likely by way of practicality at labeled junctures B, E, and F; for instance, it is generally easier to transport the rendered, lower data concentrations of control parameters and meter data between the Surface and processing hosts at (B) than it would be to try to move the raw preprocessed metering data and coefficient sets, as would be the case if a split were made at juncture (D).

Nearly any audio console with digital control fits into this loose model and contains all these elements in some form or other; even a DCA (digital control of analog) console, with a control surface separated from the processing electronics at interface B, is similar to case (1), a normal console. So it can be seen that the lineage from analog consoles, through DCA, pure digital consoles (which at least initially mirrored analog consoles almost totally), through to DAWs is quite plain.

The subdivisions of these processing blocks for different classes of console, and with likely transport means, are shown in Fig. 29-133:

1. A normal control-surface/mixing router arrangement is shown with a single major transport requirement at juncture (B).
2. A distributed mixing-on-the-network style console is shown with network links inserted at (B) and (E).
3. An all-in-one box arrangement (emulating standalone analog consoles and simpler digital consoles) is shown with no transport insertions.
4. A DAW with an external input output unit has a connection at E.

5. A similar DAW but with an add-on physical control surface has links at B as well as E.
6. A simple DAW with limited I/O output having no need of external interconnection.

This latter uses the host PC’s GUI for control and display, the PC’s CPU to do all the hosting and audio processing, and internal converters to get the audio in and out. A surprise may be the extensive use of semi-pro or domestic communications schemes in the DAW contexts—for instance, MIDI (musical instrument digital interface) for the control surface interconnection, and USB as the audio transport to the external audio I/O interface.

The major underlying message from all this is that DAWs are consoles, too! In broad-brush architecture as shown in Fig. 29-133 they are—since they have to perform all the same functions—indistinguishable from “real” consoles, which is actually an understatement, since in many respects DAWs are more versatile and powerful.

One major distinction though is that unlike most consoles, DAWs have tight, immediate integration with the recording medium, being in most cases hard drives or SSDs (Solid State Drives) which are mounted as part of the PC system on which the DAW software is running.

29.24.1 The PC

A decently fast and capable central processor(s); a reasonably easily crafted and programmed graphical and user interface; fast, inexpensive, and capacious memory; and omnipresence all afford the PC an enviable basis for audio production. It is and nearly always has been a more cost-effective platform than any purpose-built digital audio system of comparable facility. All the technical advantages made it a natural basis for initially fairly elementary audio functions such as a hard disk recorder/stereo editor, up to today where entire multitrack recording/editing/processing systems readily fit on a laptop—the like, of which would have been the envy of major studios but a few decades ago. Despite the best efforts of operating system manufacturers to make real-time audio streaming into and out of PCs problematic, the PC is a formidable tool.

29.24.2 MIDI Sequencing—Where It Began

An early PC application was in the recording, storage, manipulation, and automation of MIDI-encoded musical parts, to facilitate the assemblage of songs. This did not involve any audio, per se, merely the management of streams of MIDI commands against time. These were then issued in sequence down a MIDI path to attached music synthesizers that played the music itself.

The desired ability to compose, rearrange, copy, and time-slip parts in relation to others in synchronization gave birth to extensive and powerful automation, which largely outshone concurrent traditional console automation schemes.

Recording and manipulating audio on a PC occurred when processor speed and disk drive size and access speed allowed (two-track editing became commonplace, resulting in stereo tape recorders plummeting from hallowed possessions to doorstops virtually overnight). Although the means of getting multiple simultaneous live audio streams into the systems lagged, it was certainly possible for multiple tracks to be recorded sequentially so building up a true multitrack recording, and this was exactly the mode of operation prevalent in basement studios anyway.

And so it was not the least bit surprising that the major exponents of MIDI sequencing software became the major exponents of PC-as-studio, and their approaches from MIDI world translated over into audio world reasonably well, despite significant differences in philosophy. This does explain why those previously steeped in traditional recording find the assumptions, methods of control, and even terminology of sequencer-studio tools quite alien, while those who have grown up with it regard traditional techniques (and terminology and assumptions) to be, well, odd and quaint. MIDI sequencers have cast a long shadow over today's audio processing.

Many of the strengths of the sequencer applied to audio readily in ways unthinkable before—time-slipping or copying individual tracks or segments, unlimited takes of tracks or segments being treated as related parts rather than completely separate tracks, as examples. As the recording hardware (PC) became more powerful and the number of instantaneously available tracks increased, a deliciously ironic approach has come to the fore: originally, the sequencer shuffled MIDI elements around, in the hybrid audio-plus-MIDI the two were treated in parallel yet separately, but now it is common for all the MIDI tracks to be rendered as audio onto audio tracks just like live sources, and the audio control and automation methods rule.

29.24.3 DAW Audio Ins and Outs

Means of moving audio around are covered in more detail in section 29.25. DAWs just like any other console need to get audio in and out, and from the lesser to the greater this can include:

- PC’s built-in sound card. (It had to be mentioned, and besides, who honestly has never used one in a pinch?) Typically analog-in, analog-out (sometimes S/Pdif), at very low domestic signal levels, and of generally indifferent to awful quality. But convenient.
- USB/Firewire. Links to external sound card converter boxes from stereo in/out up to as many as 16 in/out on Firewire or USB2, much greater capacity on USB3, see Fig. 29-134.
- ADAT. 8-in or 8-out via fiber-optic cable.
- MADI. Up to 64 ins or outs via coax and more recently RJ45 type connectivity.
- Ethernet. Either true TCP/IP Ethernet or audiospecific UDP variants using the same hardware, typically 64 I/O bidirectional for UDP at 100 MHz.

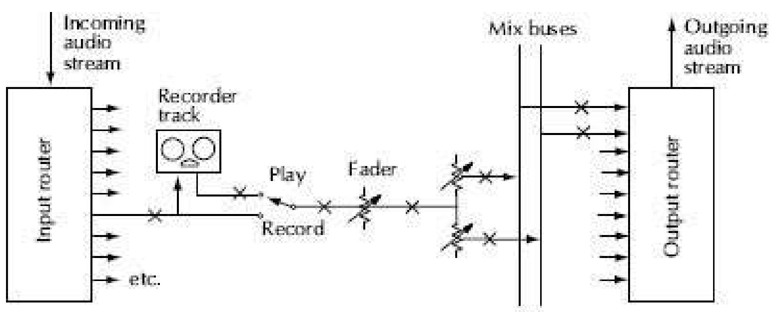


Figure 29-134. Simplistic DAW audio path.

Specific drivers—and in the case of ethernet and variants whole suites of interface code—need to be installed on the PC to deal with the audio on these various schemes. The DAW software has both input and output routers that can pick which incoming sample within a stream goes to what input, thus track, and which DAW output gets sent out what slot.

So far, the PC-based, sequencer-modeled audio control approach looks like a multitrack recorder (of virtually unlimited scope) with wicked automation and editing. But what of console-style audio processing?

29.24.4 DAW Internal Audio Paths

The audio routing within a DAW tends by design to be quite basic: Fig. 29-134 shows this as being essentially a route from input to a recorder track, thence from either before or after the track to a mix bus (or buses), thence to an output(s). Recognizable console like features such as a fader and panning are included just to show a typical starting environment.

What all the Xs mean is that it is possible to “drop in a plug-in” (translated: apply an instance of a signal-processing software module) or apply the signal at that point to anywhere else there’s an X. An output bus, for example, can and frequently does get routed back to be a recording track source (bouncing in old speak); many modules may be inserted concatenatedly at each X. There is considerable flexibility. This approach—nearly everywhere being an insert point and only providing access for processing—as opposed to the “everything’s in there in case” traditional console model allows what processing power there may be to be applied as and where it is needed while leaving all other paths unfettered.

29.24.5 Plug-Ins

A Plug In is a collection, library, of disparate software programs that variously (a) actually processes (or generates) the audio in some form or other—e.g. EQ, dynamics, delay, reverb, etc., or a MIDI musical instrument; (b) provide a graphical module for display on the system’s GUI, replete with knobs, buttons, dials, gauges, meters, and blinky-lights, (c) calculate the conversion of the parameters from those controls into coefficients that the actual signal processing can understand; and a render metering data in the reverse direction, from audio to GUI. All the “handles” typically become available to the system’s automation system, either directly or by being MIDI addressable—in other words, the module can be thought to look to the system as just yet another MIDI slave device, and which can automate it accordingly.

Standards have evolved (in the form of requirements by major DAW players) for plug-ins; VST (virtual studio technology), AU (Apple) and DirectX (Microsoft) stand out as the more popular nonproprietary schemes affording wide interchangeability between different flavors of DAW. Some DAW vendors have proprietary plug-in standards which are less (if at all) transportable outside their own ecosystems.

As mentioned, many plug ins are MIDI musical instruments or devices in their own right, but the widest variety is in audio processing. Most DAWs come with a decent suite of generic modules that allow all the traditional functions such as EQ and dynamics, plus some others that had hitherto been rack-box fare, such as reverberation units, flangers, etc.

There is a huge variety of modules available, some being specialized, and many that emulate, with greater or lesser degrees of success, existing real-world boxes, either contemporary or classic. These vary wildly, from being merely a pretty face (GUI) controlling a set of disappointingly banal cookbook algorithms and being passed off as something special, to exceptionally and painstakingly crafted emulations of existing products, accurate even down to little-known quirks. Emulations aside, DAWs have reached such a level of acceptance and usage that there are module manufacturers for whom plug-ins are their sole business.

In the “mixed blessing” department, two of the most over-worked plug-ins are “Beat Detective”, which can regularize errant timekeeping by musicians, and “Auto-Tune”, originally intended to subtly correct occasional vocalist pitchiness, but is most commonly abused as a sonic effect in its own right.

29.24.6 DAW Limitations

Any description of DAW limitations is doomed to become laughable as their underlying power inexorably increases. A solution to the lack of signal-processing horsepower in the earlier days was to offload the audio signal processing onto DSP farms, either on slot-in cards that fit within the PC’s box itself or in an external frame. This afforded far superior overall performance than was then possible from the PCs CPU alone and is still an approach taken by DAWs aimed at the professional area. The downside is that it can lend itself to creating a proprietary technological island, exacerbated by the use of nonstandard file formats, making interchange between other types of systems difficult.

Clever approaches to make the best use of limited processor steam revolve around using otherwise dead time and the almost limitless ability to store recorded tracks. As an example, if an EQ is applied to a track, rather than run that EQ in real time each time it’s played (along with possibly dozens of others, which may very well drown the system), it can be run very quickly and quietly, once, across the length of the track, which is then saved as another track. That way the system just plays back a pre-EQed track rather than having to run an EQ—a huge saving in resources. If a change is made to the EQ halfway through a playback, the EQ runs in realtime from the change but at the end of the playback the resultant overall EQed track is saved as yet another track. The system keeps track of which track is the most current: this is also key to how DAWs can seem to have boundless ability to roll back or Undo changes—in addition to the automation remembering all the changes, all the older tracks are still available for instant application. Effects tracks, reverb passes, etc. need only be striped once, and never need to eat PC power again. The tracks displayed to the user can sometimes be the mere tip of an iceberg. As PC power has increased the need for such clever tricks has receded.

Reference to Fig. 29-134 shows that it is, theoretically, possible to avoid the use of the recorder altogether and simply use the DAW as a straightforward mixer. However, one has to remember two things:

1. Every instance of every plug-in will use juice, and one will eventually find out how many plug-ins is too many—the resource mitigation dodge of prestriping effects on tracks doesn't work mixing in real-time, where everything has to be happening at once. As processors inexorably improve, this limitation pushes out further but will never disappear—remember "Work fills resources allotted plus 10%". That said, large-scale live recordings can very successfully be done on such a DAW—the sources typically all go straight to track with little on-the-fly processing being necessary, and it can all get fixed in the mix. It is nowadays common for each channel/track of live shows done on DSP or DAW-based live consoles to be recorded each night as a matter of routine, just as once a stereo show cassette once was.
2. There may be excessive latency (input-output delay), mostly from the acts of getting the audio into and out of the boxes and

sometimes from cumbersome internal routings; this may well be audible or annoying in some circumstances, or an absolute bust if excessive in say a stage environment. It should be remembered that if one path in a console/DAW has added latency, all the other related paths should be arranged to match that latency, too.

Fig. 29-135 shows a typical DAW main screen B most follow this general format to a large degree. Most notable is the horizontal disposition of recorded tracks showing amplitude against time. This particular work area allows direct editing on individual (or multiple) tracks, even down to moving chunks around in time, crossfading between tracks ("comping") and otherwise processing to microscopic degrees. Along the bottom are more conventionally vertically disposed fader blocks, most commonly used in the mixing phase. Indeed, it is common when recording to use these as one would have used the monitor section in a traditional pre-inline console. Plus ca change. The beauty of the DAW approach is that the workscreen can be morphed at will to emphasize the most useful facets for any particular phase of workflow, maximizing use of the display.

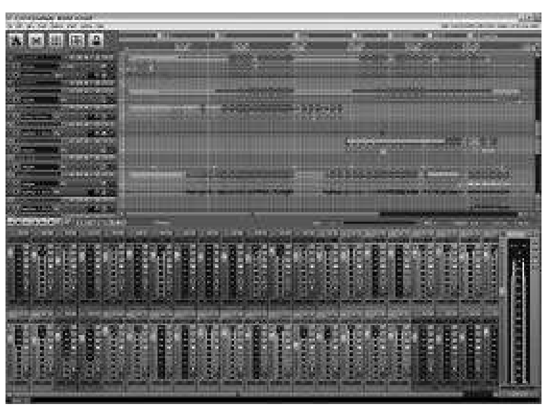


Figure 29-135. A Typical DAW mixdown screen. Reaper, courtesy

Cockos Inc.

29.24.7 DAW Irony–The Console Lives

Most professional users of whatever DAW add external control hardware to make life easier. (This can be achieved by mechanisms as simple as MIDI commands tying specific hardware controls to virtual screen actions, or via more complex an proprietary Ethernet based control.) The two main areas are in monitor control (selecting into and controlling the level etc. of the control room monitors), and of course fader-based surfaces on which to mix. Fig. 25-136 shows a culmination of this, the Yamaha Nuage control surface for physical control of Steinberg "Nuendo" DAW software, in this instance obviously intended for an audio-for-video editing suite. This approach, with surfaces of similar complexity, is common in workaday facilities where complex productions are mixed. Indeed, the console lives.

Figure 29-136. Nuage DAW Control Surface. Courtesy Yamaha/Steinberg.

29.25 Moving Digital Audio Around

As is plain from the earlier discussion of digital audio mixing and processing systems, and in particular that there are few constraints on where the constituent bits are physically in relation to each other, there can be an awful lot of audio to shuffle around between them. The term intra-console is used to distinguish this interconnection between bits of a console system, as opposed to moving lumps of audio around in a facility. Often, though, this gets blurry!

Sometimes all that is required is the movement of some audio from one place to the other, but increasingly there is a requirement to have all—or some—sources available at all—or some—destinations in a free-grouping arrangement. This takes on a life of its own as a network. Most end-to-end signaling types as described here can be made to become part of a such a network if they are arranged to have one of their ends terminated in a hub, in star configuration with other end-to-end links; the hub—router—has the intelligence to route the signals telephone-exchange-like accordingly. Some described transport mechanisms are designed to be networks, or network like, in their own right—the 800lb. gorilla in this world being TCP/IP.

29.25.1 Moving Audio—Small-Scale

29.25.1.1 AES-3

Stereo pairs (or pairs of monos) have been catered for by the venerable AES-3 standard; this is a Manchester encoded stream of two (up to) 32 bit audio words, some informational tags for link status and format, and a number of user bits that may be used for anything from turning stuff on and off remotely to serially carrying metadata (program-specific information) or more complex real-time control data. It was designed to be robust, simple, and as usable as possible in the predominantly analog world into which it was born, even down to using the familiar 3 pin XLR connectors in its usual implementation. With minor updates, mostly concerning connection variants and data rate (it now handles the once-unthinkable 192 kHz with ease), it still serves well. It is a very close cousin (indeed the underpinnings are chemically indistinguishable, and use the same chip sets) to the domestic S/Pdif (Sony/Philips Digital InterFace). The audio is treated identically, but the format and informational tags differ. It is common for AES-3/-S/PDIF receivers to be set up to strip these off so as to allow universal connection, but obviously this is at the expense of any metadata that may accompany the audio stream, and, if this is of the least concern, any digital rights mismanagement flags.

A performance downside to AES-3, particularly with early implementations, was recovered clock jitter. Best performance is achieved by reclocking at the receive end, either by SRC (Sample Rate Conversion) or the use of very good flywheel phase-lock loops to reestablish solid, quiet clocking. If the facility is homogenous with everything running off a master clock this is less important; “bits is bits” and as long as they arrive within the same framing period (e.g., 20.8 μs at 48kHz) and sample clock period, and any D/A is done with the same pristine clocks as any A/D, transmission jitter is irrelevant.

29.25.1.2 AES-42

As will been seen in the later mention of USB microphones, there is a drive to push digital as close to the source as possible, in that case for simplicity’s sake, in proaudio for performance. The concept of putting mic preamplifier, A/D converter, and processing inside the microphone itself is at one and the same time seductive and puzzling. The idea of simply taking a digital stream (possibly in AES-3 format) straight from a microphone into a digital system holds strong sway; reflection shows that this—in any meaningful system—means either the addition of a plethora of hitherto unknown knobs and switches on the microphone itself or the means of remote-controlling all those functions and takes the shine off the idea somewhat, particularly to those to whom a microphone is something one simply plugs in and uses.

As has been made clear, there is little that binds a particular function to a particular physical location or piece of system hardware or software. Given that, some mouse-and-screen GUI widgets to control the microphone parameters, or indeed a physical set of hardware knobs and switches to do the same, don't care whether the target is in the same box, another processor, or even on the same continent. That, in this instance, the target is on the top of a shiny microphone stand in the studio is irrelevant. So, not only is a means of getting digital audio from the microphone necessary, but means of getting the control parameters or coefficients up to the microphone, as well as a synchronizing reference clock so that the microphone's pristine audio doesn't have to suffer the immediate indignity of a sample-rate conversion to match the rest of the system. And, of course, a means of powering all this.

And so was born AES-42, in an effort to standardize all this before multiple incompatible approaches dissipated the concept's appeal. Fig. 29-137 shows in outline form its scope.

Many hitherto console functions have found their way into microphone control via AES-42. Although the scheme is not limited to these, the Neumann TLM-103-D digital microphone, for example, allows gain, microphone pattern, absolute phase, high-pass filter, an in-built compressor/limiter/de-esser, and a peak limiter's parameters to be controlled. It's easy to see where that's headed; no need for console channels as we've known them.

The normal connectorization is via the old familiar XLR, although the XLD is suggested for circumstances where confusion with other XLR-using systems could potentially result in damage. As would be expected, signal formatting owing much to the familiar AES-3 is used to retrieve the audio, which ordinarily comes differentially down a shielded pair; user bits in the data stream relay fixed data such as the microphone's manufacturer, model number, and available controls; variable data such as instantaneous parameter value are also available by this means. Now the fun begins—power is sent phantom style (common mode and with reference to the shield) back up the line; instead of its merely being regulated down to power the microphone and its electronics, it is also modulated with control data and a synchronizing word clock, which are filtered off and used to instruct the microphone's processing. The microphone's sample rate may either free run, in which case it is a master (but will probably need SRCing to work in a system of any complexity), or it can be slaved to the synchronizing word clock. The latter is favored, if available. Note that the power scheme is not normal 48V phantom; this would be wholly incapable of powering all the necessary electronics in the microphone body; it is 12V at significant current capability.

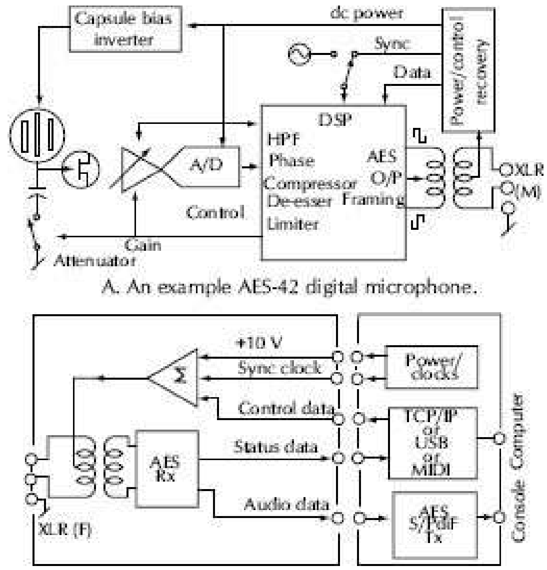


Figure 29-137. AES-42.

Present digital microphones using AES-42 have a choice of termination, depending on whether the system into which it is plugged already speaks AES-42 in that particular microphone's dialect (and so control from that system is implicit), or via an external interface box which permits a computer running the appropriate and proprietary control software to talk to the microphone and audio data recognizable as AES-3 or S/Pdif stripped off for use.

An interesting side note is that Neumann, a major influence over the scheme and early adopter, make claims that such an arrangement results in better overall dynamic range than traditional microphone connections; this is presumably down to two major aspects, being the minimal amount of interfacing between microphone capsule and convertor, and the use of an overlapped gain-ranging convertor scheme—this in addition precludes the need for a gain control element, since that can be achieved in purely in digital scaling with this method.

29.25.2 Moving Audio—Multiple Paths

More than a stereo pair calls for more radical answers, and as is typical with fast-moving development tended to outstrip standards-making—never mind the commercial impetus to try to capture users within a proprietary format. Two formats, one from pro audio, the other from semi pro, stand out from the earlier days of multitrack recorder/console interconnection:

29.25.2.1 AES-10–MADI

This format is very common for the interconnection of digital reel-to-reel recorders (whoever thought we'd be weeping nostalgic for those?) and older large-format digital consoles. It carried up to 56 audio channels (now 64) originally down inexpensive TV-style 75 Ω coax (plenty enough for a 48 track recorder) and owed a lot to FDDI, an older communications network backbone format. Being unidirectional meant that a MADI link in each direction to a recorder was necessary. Latency was quite low, it was deterministic, and implementation fairly straightforward. An oldie-but-goodie, it is still used to an extent in the pro audio world by some manufacturers for overall system interconnection, intra console knitting, and as a common-ground form of interconnectivity between different proprietary digital networking schemes.

29.25.2.2 ADAT

ADAT is a simple (both in hardware and signal format) unidirectional fiber-optic interconnection, originally, to get 8 audio signals into or out of the once highly popular Alesis ADAT VCR-based 8-track recorders (which can be thanked as being the likely tipping point of recording from uptown to basement). It is still an interconnect of choice in semi-pro recording equipment, where "pieces of eight" is adequate or sensible, as when additional functionality is marketed in such a modular fashion.

Being a strictly hardware interface it is wholly deterministic (audio arrives exactly when expected) and with very low latency. Although inexpensive chipsets are available, the format lends itself to low-impact implementation in (possibly already existing) FPGAs (Field-Programmable Gate Arrays) within a product design, so incurring near-zero add-in cost.

An ADAT frame, which can carry up to eight 24 bit audio words, is 256 bits long at a clock rate of 12.288MHz for 48kHz. There is a 16 bit preamble containing a 10 bit frame-sync period and four user bits for control/messaging. (The arithmetically astute will wonder where the other 46 bits went; they are used throughout the frame after every 4 bit nibble—except in the frame-sync period—as synchronization zero-value bits). The bits are scrambled (Manchester-encoded) to non-return-to-zero to remove any tendency to have a dc component. Some of this—in particular, the syncing and NRZ—had a lot to do with coping with the vagaries of VCR tape transports, but as a long-standing standard with millions of installed instances it holds up very well and doesn't warrant the potential confusion revisitation and redesign would incur. It is hard to envisage a simpler robust multichannel self-clocking interface, and its designers deserve full credit!

29.25.2.3 USB

A somewhat surprising development has been the adoption of the humble USB serial connectivity of PCs to move moderate amounts of audio about. This reflects the massive shift over recent years from the large-scale studio-as-shrine approach of the recording business to small-scale home or demo studio recording becoming the new mainstream.

Conceived as a replacement and expansion for RS-232 serial connections (and related mouse/keyboard interfaces) for PCs, the early USB implementation (e.g., v1.1) was hard pressed to reliably move a stereo pair of 44.1kHz about, but the upgraded USB2 with its nominal 480 MHz data rate changed all that. As an example, Fig. 29-138 shows a 1U rack-unit box by Tascam (beneath the laptop, above the mega mic preamps) that readily simultaneously transports 16 audio paths to, and 4 back from, a PC running DAW software, all via USB2. USB2 has eclipsed Fire Wire (IEEE 1394), a similar-speed (if network-capable) interconnection that hitherto briefly reigned in the sphere of small-scale PC audio transport to external A/D and D/A boxes and such.

Figure 29-138. A modest-sized DAW running on a USB link between the audio interface unit (center) and the laptop. The (almost free) DAW software allows for 48 simultaneous recorded tracks, with significant audio processing. Complete with shown external mic-amps and computer, the cost of this outfit is about that of a decent microphone; its 1990 equivalent in facilities and performance would have cost the same as a decent car, while the

1970 version would have equated a decent house, which would have been needed to fit it all in, too.

It is not at all uncommon for small mixers—analog or digital—to present their outputs and accept a returning pair of inputs via USB; small hand-held recorders likewise; microphone preamplifiers; even USB microphones, which only work in that context. The PC to which they are connected can recognize such simple interconnections as just an external Sound Card, and apply built-in generic drivers to cope. Zero-effort connectivity. Better performance and more advanced control and features can be achieved with special drivers installed in the PC (ASIO), but the instant connectivity thing is hard to beat.

One rightly has to be circumspect about jitter performance on a transport mechanism that certainly was never characterized for digital audio streaming with very stringent clock recovery requirements; in the case of the small DAW setup described, since A/D and D/A are in the same box, and assuming the clocking is done conscientiously within, the overall performance is limited by that, any vagaries of the link, its latency, or computer timing being irrelevant as long as the clocks remain synchronous with the data. Such are the dangers of any transport scheme where the clock is solely implicit to the data, with no external reference.

Of far more concern, however, is the computer's operating system's handling of audio, which can create a land of horrors transcending any worries about link jitter; this is typically addressed by the loading of unit-specific drivers (ASIO in this case) into the computer, which blow right by the operating system's clunky hardware abstraction scheme, and instituting delay buffers capable of absorbing most temporal irregularities. Nevertheless, a

USB (and perhaps more so FireWire) link's performance is often dropout-limited by the host PC's handling of DPCs (Deferred Procedure Calls), the stacking up of interrupts and time-related routine calls that take longer to address and clear than the link's buffers can sustain. Sometimes a lot of effort has to be put into disabling features/programs/peripherals (like wireless networking in particular), updating or finding the right/better drivers, optimizing this and that, installing replacement hardware, and general hair tearing, just to get a PC to adequately pass/process meaningful amounts of audio. The PC really is not a shining beacon of a streaming-audio-friendly environment!

29.25.3 Digital Audio Networking

29.25.3.1 CAT-5/RJ-45 Interconnectivity Types

The following communication schemes all typically use the widely available (even from the local stationery store) networking style cabling, exemplified by CAT-5 or CAT-6 cable terminated in the little plastic RJ-45 phone like connectors, and indeed often share the same terminating PHY (PHYsical Layer) electronics. What actually goes through them can differ wildly though. As will become plain, this enabling technology has also expanded the notions of what can be done in the context of moving large amounts of audio around, blurring the distinctions of transport, mixing, and processing.

29.25.3.1.1 UDP

User Datagram Protocol (sometimes called User Defined Protocol) —essentially uses the same (fabulously inexpensive and readily available) Ethernet-style connectivity, hardware, and chip sets but with a far simpler messaging protocol than TCP/IP can often be more efficient and better suited to the application at hand. The complexity and low overall capacity of TCP/IP in the early low-speed days drove most effort then to mass reinvention of the network wheel in UDP. There became countless varieties, all proprietary and utterly incompatible, of course, since there is a strong commercial impulse to keep everything in house. A standard, AES-50, attempted to bring sanity and compatibility into the gigahertz realm, with the myriad 100MHz schemes already considered a lost cause. Paradoxically, 1 GHz finally made audio over TCP/IP (below) viable, resulting in few AES-50 implementations.

At its simplest, a big packet consisting of a header and then however many audio words of whatever width is constructed, then sent down a dedicated Ethernet hardware circuit, every sample period; at the receive end the simple format is readily decoded and the constituent samples recovered. It is a reasonable assumption that on a dedicated line the packet will be unimpeded and arrive intact, and such links typically run raw with no mechanism for error trapping. In short, these systems run just fine without and typically do.

Audio is only part of the whole picture. Metadata accompanying it, logic switching contingent on control, control data and metering data all have to be considered and accommodated within the link for it to be a fully usable system in any meaningful context.

Such UDP links are typically bidirectional (but sometimes unidirectional) end-to-end closed links. In and of themselves, they don’t constitute a network, which can loosely be described as anything to anywhere—where any source connected to the network may be picked up by any destination. There are two general schemes for turning these one-to-one links into networks: cascading them node-to-node, with a modified signal passing along each link (Serial), or arranging them all to radiate from a central hub (Star).

29.25.3.1.2 Serial or Loop Networking

In this methodology a single unidirectional line is run passing through each area that needs access to the network; access is achieved by nodes or breakout boxes of varying complexity depending on the requirement, each of which has a unique address for programmability purposes. At the simplest, a small fixed number of inputs to the network and outputs from the network may be offered at the node, along with unique control data. These may be analog ins/outs or digital ins/outs or a combination, and each may either look at (in the case of outputs) any of the (say) 64 program slots, or select a slot into which to place their audio (in the case of inputs).

More advanced nodes could, by way of example, look at many slots, mix them, mix that with local input material, and even place that composite mix into further slot(s) in the network. A frequent application is to retrieve audio from a slot and replace it with local input material. Signal processing specific to a local need (e.g., crossover/EQ for a speaker cluster) can be done within such a node; indeed, such products can be thought of primarily as a processor that just happens to have wide connectivity through the network and is marketed as such. Third-party or multiple vendors can be interoperable, providing they're all licensees of the same networking protocol.

This modified stream is then sent downstream to the next node, and so forth. The stream can be unidirectional (serial) or looped back upon itself (surprise, loop), whereupon the originating node seemingly perversely sees as its input the stream after it has passed through all the other nodes.

Such networks tend to be quite efficient, since they are able to reuse slots along the way. Disadvantages are:

- That serial anythings tend to be badly affected by single-point failures—in other words, one node failing makes orphans of all the others downstream of it, bisecting the net.
- It takes an appreciable time to receive the packet of slots, disassemble it, modify slots to whatever degree, reassemble it, and send it on its way, and such processing latency is obviously cumulative with that from previous and successive nodes. That said, latencies can be low, in the handful-of-sample-period range, trivial compared to those of 100MHz TCP/IP systems.
- The network cabling routing has to be carefully thought through and follow a logical progression of where the audio needs to go next. This sometimes isn't easy.

29.25.3.1.3 Star Network Topology

In contrast to relying on the free-form packet addressability of an IP-style Ethernet network—which belies the need for centralized command and control-CUDP-style networks which dumbly if faithfully and with low fixed latency propel a fixed amount of audio from one end to the other of a straight pipe, requires a central switch: this unpacks each incoming stream, decides which elements within them need to go where, and assembles outgoing streams appropriately. Such routing systems were long commonplace in broadcast installations, and the morphing of the concept to using high-bandwidth UDP pipes for transport was a natural and welcome progression from running lots of discrete signal lines. Such UDP pipes are often described as TDM (Time Division Multiplexed) even though that isn't strictly true with packetized systems (meaning having a preceding descriptive header combined with the many audio samples, and sent as a whole). TDM implies a strict time-slot in which a particular audio path is made to be present within a large cyclical number, which is inflexibly repeated. Packetized, it just means that the audio samples are concatenated in the same order each packet.

Having a central hub from and to which all network runs are connected is a concept as old as the phone system. As such, it has similar strengths and weaknesses in that cable runs are logical and obvious, but reliability hinges on that of a central server. Although this would seem a vulnerability, a single point of failure, parallel and redundant methodologies are common, as will be seen.

The switch in audio terms is in fact an audio router, which accepts large numbers of sources and can redistribute them in any combination to large numbers of destinations. Oft-times many sources and destinations are local to the router, but more often high-density audio pipes as described above of, say, 64 discrete signals bidirectionally, spur out to remote locations, where these pipes are terminated in input and output terminations of whatever nature and complexity are desired. (If all the outputs need to be analog on XLRs, so be it. AES pairs, no problem—termination styles are easily accomplished to suit the application. If local-specific signal processing is desired, no problem.) These pipes are simply arranged to look like multiple sources and destinations to the router and are treated as such. The router also parses any metadata, logic control, or metering that accompanies each audio path and routes or deals with each of these accordingly. (Losing the meta data or sending it to the wrong place is like the airline losing your bags: it isn't the end of civilization, as you've arrived, but you're nowhere near as equipped to accomplish what you have to do.)

This centralized router model works well in the broadcast environment, radio or TV, where much of the engineering work is clustered in a central racks room anyway; likely many of the sources and destinations of the router would be local to it in that room, easing interconnection, with spokes of high-density audio transport issuing out to each studio and production area—Something of a natural for the star topology.

Live sound benefits from this method more than others (such as serial), too. A typical setup is for two consoles (house and monitors) to each be recipients of all stage audio sources or at least major overlapping subsets of them. In this instance the router would receive the outputs from the active stage boxes as sources and distribute them as required by transport links to the two consoles.

Returning from the consoles to the router are:

1. House—main mixes and/or loudspeaker processor outputs.
2. Monitors—many, stage monitor mixes.

These enter the router and are then sent by further links to the desired amplifier racks for the flown and/or stacked PA and sub-low cabinets, or indeed straight to the powered speakers themselves; and to the amplifier racks for the stage monitor speakers and the transmitter rack for in-ear monitoring.

Additional feeds, such as for recording, are assembled in the router and sent to the band's DAW or the recording truck and terminated accordingly.

29.25.3.1.4 Mixing in the Router

A progression from the notion of a router being a simple crossbar or summing switch is that it becomes a soft matrix, where the relative levels of sources and destinations may be varied. In other words, a mixer. Or a number of smaller mixers than the total capability described by possible numbers of inputs and outputs.

And, taking a step further forward, given the already signal-processing-intensive environment, the console-style signal processing on mixer inputs, outputs, and submixes becomes relatively easy to implement.

It is common for the router to have console-type signal processing capability, indeed for it to be where the mixing/processing parts of consoles reside; it makes sense, since all the possible component signals to be mixed exist within the router or can be got there expeditiously enough through one or several links. Fig. 29-139 shows a large-format TV audio mixing console in which there is not a shred of audio—it is merely a control surface, controlling the signal processing/mixing elsewhere within such a processing router. That the console system is or is part of a router means that the number of available mix sources is limited only by the size of the router and can so extend to the thousands. Unusually, this particular system is also capable of using AoIP-connected I/O units too, and form part of a larger AoIP networked system. (See below for AoIP.)

From an operating perspective, however, the console is limited to mixing only as many sources instantaneously as it has faders; this can be multiplied by paging the surface, such that it can flip-flop either entirely or on a fader-by-fader basis to control multiple other channels; instantaneous channel counts can thus run into the hundreds, and on big live shows (such as election night) often does.

Figure 29-139. The Wheatstone Dimension 3 large-format TV audio console, a control surface which has no audio in it at all! It merely controls a remote mixing router, and connects with its I/O via either UDP or AoIP.

There are two other major applications for this router-as-console solution. One is radio studios—several, if relatively small, consoles within a single complex: they can all share the same hardware and resources of a single router, indeed may all be in the same box, yet to all intents and purposes be discrete operationally.

It is a natural live sound solution, too, where as described above there may indeed be multiple fair-sized console systems (house, monitors, recording) but that all share common sources from the stage and yet have very separate destinations (PA, monitors, recorder). The mixing router not only performs the signal routings, but it is also home to all the console-type signal processing required for all three operations.

A valid concern for each of the above applications is that of single point of failure. Large routing mixers typically address this with fail-safe measures, meaning each host microcomputer has a hot standby ready to take over if the main one should hiccup, and spare signal processing/mixing DSP boards equally stand ready to be reassigned on-the-fly to take over from one that may have halted. Some designs even have an entirely separate router, operating in parallel to the main one, ready to take over in the case of a failure. Although they could be perceived as expensive precautions, they look really inexpensive in relation to dead air.

29.25.3.2 AoIP—Audio over Internet Protocol

Ethernet, using its handmaiden TCP/IP platform, is highly popular —ubiquitous—and there is a large base of skill in operating and maintaining networks based on it. This has lent impetus to trying to use it for things for which it wasn't really intended and is not particularly apt, such as moving professional audio around. Immense effort and marketing has gone into making it work adequately.

Any networked computer user knows how facile it is to move audio around, either in chunks as files or drip fed as streaming, either on a local network or the internet itself. A seemingly sensible follow-on would be to wonder if the self-evidently already existent methodology could be used for moving large quantities of audio around in a digital audio network.

Well, the short answer is "Yes." Given the existence of Ethernet connectivity and a good IP stack (network operating firmware) in each of the required connected units, a significant number of uncompressed audio channels can be moved around successfully by such a network.

Now for the long answer. Aspects of its performance at best are eclipsed by alternative methods. The only real advantages are the previously mentioned wealth of user familiarity with, and labor skilled in, TCP/IP networking, and the freedom from having to predetermine cable runs. As long as a desired area is connected to the network by any means, audio transfer can take place—there is no need to provide a "home run" directly back to the main router, or a dedicated sub-router. There are though significant drawbacks.

In the Internet example, the audio is almost always compressed by MP3, WMA, AAC, Ogg-Vorbis, or whatever format to radically reduce file size or to fit within a required streaming rate. Except for a few arguable examples such as news, remotes, or spot insert distribution for radio broadcast, compressed audio has no place in professional audio. So suddenly the necessity of uncompressed audio payload can be ten times or beyond the size of domestic audio. Network congestion effects loom that much closer, that much sooner.

TCP/IP is a packetized system and incurs at best a minimum packet assembly/disassembly time at the ends in addition to the relatively quick transmission times. The packets are ordinarily comparatively small—in a streaming sense—multiplying the processing/deprocessing overhead and bandwidth wasted in sending packet headers. (This can be tweaked, however, trading packet size vs. flexibility to minimize the overhead proportion. Ideally audio "streaming" packets are made long to minimize such overhead.)

This all incurs latency (i.e., a delay between input and output) that may or may not be acceptable. Live applications—say, broadcast or sound reinforcement—might well have issues, particularly if many links' latencies become cumulative. These are relatively minor latencies, though, compared to what's to come in a real environment. Oh, yes. It gets worse....

Congestion is paradoxically key to TCP/IP's main limitation for audio, considering it was designed to—and does—handle congestion superbly for its intended use. In the absence of any other network traffic that may contend with a primary audio stream, the packetized audio will likely arrive unmolested and in order, and a fairly high density (lots of audio) may be passed from a point A to a point B. In short, in a point-to-point dedicated link, audio via TCP/IP can work at its best and reasonably well.

Unfortunately, that's not what the network concept promises: multiple independent streams from multiple sources and with multiple destinations sharing the same wire infrastructure. As soon as other traffic hits the network—say, another audio stream from point—to point D—try as the carrier-detect collision avoidance mechanisms inherent to IP might, packets from one stream will unavoidably tread on those from the other. One of TCP/IP's great strengths is that it recognizes such events and deals with them handsomely; each stream gets the opportunity to resend its broken or unacknowledged (i.e., lost) packets, and the receive stack knows to reassemble the stream in the correct order from the now possibly out-of-order and certainly delayed packets. Jolly good. So what's wrong with that?

29.25.3.2.1 Buffering Latency

The network has lost any tenuous claim to determinism—predictability—it may have shown, since collisions and recovery therefrom are unpredictable both in frequency and recovery time.

Determinism is, in short, knowing exactly and consistently when recovered audio is ready for use—absolutely essential for streamingtype or real-time audio or dropouts occur. A pure, isolated, low-density point-to-point TCP/IP link can be close to deterministic and with a relatively short latency, predictable from the above-mentioned packetization, framing, and transmission times. Even so, it's only close—other traffic still exists on the link, in the form of ACK (acknowledge) replies for each sent packet: Yes, collisions can occur between the real data and its own ACKs! Real-world, where multiple paths on the network are in use, significant collision-recovery times get thrown into the mix and this now unknown added time becoming even more so and approximately geometrically longer as the amount of traffic increases. There is also the very strong likelihood, nay—certainty, that packets that have been stepped on and repeated will arrive out of sequence, the repeats only getting through sometimes many frames after several in-sequence packets have progressed.

The workaround—trading a fixed, known, longer latency for a shorter but unusably unpredictable one—is by instituting a buffer at each receive point. It has to be longer than the longest expected oopsy-recovery on that particular network configuration. In order to allow time for packets to be eventually received and juggled back into order, this fixed deliberate buffer latency has to be incurred; the more traffic, the more latency is required, and on a busy network this can be in the tens or hundreds of milliseconds to encompass worst-case congestion effects. Although acceptable in some circumstances this is difficult to swallow for many audio applications—particularly those where there is a requirement for humans to listen to themselves live through such a system.

Consequently, lowish-latency and pseudo-deterministic audio links are usually recommended to be placed on discrete one-to-one links with little risk of contention. Which really rather begs the rationale behind using TCP/IP, and the promise of “networking” upon it. Oh, well. All of these ills are exacerbated if any other traffic is permitted on the same network—which finishes off the naive notion of running significant amounts of audio on an existing office network, or allowing facility enterprise traffic anywhere near the audio network.

Worse yet is expecting sensible behavior if incoming or outgoing real-time audio is expected through the outside-world Internet—build-out latencies may have to be far, far longer (sometimes seconds) to absorb the hairiness of the unknown out there! Again, this may be acceptable in some circumstances—after all, if one is using the Internet, the likelihood is that the audio is going a long way away, where no frame of reference in time exists to its source.

One saving grace of the general move to gigahertz Ethernet (as opposed to the once commonplace but now obsolescent 100MHz variety) is that everything happens much quicker, and that for normal practicable amounts of traffic the collision rate and recovery times go right down—or are eradicated completely—and so the build-out buffering latency can be radically reduced; TCP/IP as the basis for an audio network reaches a lot closer to the promise, as opposed to the highly marginal on-the-edge behavior of any meaningful size system at 100MHz. The advantage is not so much that ten times the traffic could theoretically be handled, but that a similar amount of traffic can be handled well; latencies in the low single-digit milliseconds are readily achievable, which, if not too many passes through the system are attempted (remember, the per-pass latencies add up), is a generally acceptable performance.

Any meaningful size network soon demands the use of a network switch. A handful of units may work perfectly happily tied together through an office-supply router, but beyond that some muscle is required. Routers merely buffer and ensure that data passes to the right destination from the right source. A switch handles things more intelligently, by stratification—allowing certain links to be deemed inviolable (audio streams) at the expense of lower priority data (control, metering, texting), and by ensuring data from one point to another passes through the minimum amount of network infrastructure. What this latter means is that if a stream (heavy data usage) can be kept off wires, the unused bandwidth in those wires can be used for additional traffic. They also handle distribution chores, such that one source being sent to multiple destinations need only be sent once total, as opposed to once for each destination; a major resource savings. On large systems, the Big Switch can be both highly expensive and invaluable.

29.25.3.2.2 AES67

Quite predictably, there are a number of commercial AoIP (Audio over Internet Protocol) systems in the marketplace which have carved out market niches; some are strong in broadcast, others in commercial audio, yet others in live sound. It is unlikely during the technology's likely lifespan that any one flavor would predominate. The need for at least future systems from different manufacturers to interoperate was the impetus behind the AES67 standard, and at has at least provided a common bridging format between them until then. The main thrusts are:

- SampleRate—48 kHz required, with 44.1kHz and 96kHz capability suggested.
- Low Latency—a packet of 64 24-bit words will end-to-end within 1 ms at 48kHz.
- Clocking—should use the IEEE-1588 synchronization method.

Most AoIP systems use a relatively simple metronome timekeeping system (sending out a clock packet at regular intervals) which can be arranged to keep all the audio sampling on all the nodes of the system in very good agreement. IEEE-1588, which owes much schematically to the Internet's hierarchical NTP (Network Time Protocol), uses a high-accuracy traceable realtime clock which is distributed to all relevant points on the network—each node knows the exact time-of-day to very fine degree, and uses this to regulate audio packet transmissions.

Not defined in the standard were control protocols (the forms of metadata and remote controlling data, etc.) nor Discovery. Discovery is how a network figures out what is initially connected to it, and determines if something wants to join, or has dropped off the edge of the world. It can get very involved and is a key element to determining a system's operational robustness.

29.25.3.2.3 Latency—How Much Is Too Much?

To a greater or lesser degree Audio-over-IP systems incur significant latency. Despite many learned researchers' effort, most data concerning the audibility of latency is based on the anecdotal and apocryphal. But there is no substitute for being on the wrong end of a broadcast presenter ripping off his headphones and spewing invective as establishment of an incontrovertible benchmark.

We won't even discuss delays that are long enough to be discernible as a delay, or a discrete echo; that is obviously way too long, and everyone, trained or not, has a hard time speaking normally when fed such into headphones or monitors. No, it's that mushy area less than, say, 50 ms delay—a period of time below which the ear/brain attempts to integrate all correlated sources into one—that is of concern.

Latency is an issue where a performer is listening directly to a delayed version of him or herself; two situations to keep in mind are a DJ wearing headphones or a stage performer with in-ear or conventional floor/side-fill monitors. An important thing to note is that very different answers from these people as to what is noticeable, annoying, or untenable are garnered depending on whether they are introduced cold to a system with delay, or are steadily introduced to it, particularly in the cases of headphones/in-ears, or it changes.

Talking, one hears oneself not only by what's coming through the headphones, if they're open-frame headphones (i.e., not enclosed), by room spill, but also by bone conduction within one's own head. This latter is distinctly band limited, and what is passed is usually just the fundamental and possibly early harmonics of vowel sounds. Interference (comb filtering) between this and what is being stuck in the ear causes a nonflat perceived frequency response, with cancellation notches and corresponding reinforcement summations. (It is the same mechanism as the audio effect flanging.) This is in general no real problem—one quickly accepts that sound as being normal, the sound of oneself wearing headphones. Deliberately introducing a different delay by even only a millisecond or two is immediately perceptible—the interference cancellations/summations change—the sound changes. This is why many tests attempting to establish acceptable latency by steadily increasing delay have resulted in unrealistically low values; the relative changes in coloration with even small changes in delay are very easy to perceive, even by the unskilled—and immediately flagged as a problem.

Conversely, if one were to present a subject with a delayed headphone feed even quite a bit larger than this (without previously having had chance to establish a reference), the interference-related sound would readily be accepted as normal.

In daily use on countless radio stations are airchain processors with delays in the 10–15ms region; this, in addition to other latencies in the loop path from microphone to headphones listening off-air, means delays approaching 20 ms are commonplace and to a greater or lesser degree, accepted. Much more than that, though, engenders complaints of the sound being disconnected or hollow and distracting.

Time-alignment experiments conducted on large-scale rock’n’roll sound systems reached broadly similar results; 20ms monitor delay was as much as could be tolerated by most performers, although some could detect far less, but most readily acceded not to be too bothered by it. Delay between the performer and the PA, particularly in a large venue, proves relatively unimportant for two reasons: firstly, the performer has much more present (louder) monitoring to which he’s likely paying much more attention, and, secondly what scatters back from the PA is quite diffuse and decorrelated anyway. In all cases, the threshold of unacceptability is very crisp—definitely a straw-that-breaks-the-camel’s-back situation.

The main thing to be considered in all this is that latencies add: each pass of a signal through a signal link or network; each piece of gear or processing to which it is subjected; each propagation delay adds up to often be significantly bigger than one might expect. Just one more teentsy-weensy little few link milliseconds through an AoIP pipe might just break it.

AoIP networks do afford the low-impact distribution of many and various system elements to where they are best located. For instance, a console's mix engine may be a stand-alone unit, separate from but controlled by a surface, yes, over IP. In particular Input/Output may be distributed; microphone amplifiers to the studios to minimize microphone cable lengths; AES I/O in a rack room for interconnection to satellite feeds and such; AoIP-to-MADI convertors for multitrack recorder interfacing. The routing can be just as flexible as any centralized router—any microphone source, say, can be picked up and used by any console sitting on the network. From an expandability standpoint, AoIP systems are very powerful: Instead of having to pre-determine likely usage and expansion up-front as is necessary with a UDP system, extra units or indeed additional consoles may be added piecemeal to the network as demand requires.

29.25.3.2.4 An Example AoIP Network

The elements of an AoIP system are usually broken up into bite-size chunks, allowing greater flexibility and lower incremental costs for expansion. In Fig. 29-140 a moderate-sized radio broadcast configuration is shown, being three separate broadcast studios and basic support equipment. Each 1U rack unit (sometimes called

“Nodes” or “Blades”) is AoIP capable, as are interface cards in each of the four PCs along the bottom. Each of the “consoles” is in fact just a control surface (familiar?) and each controls its own mix engine (again, in a 1U box) via a local switch. Also attached to the local switches are input/output Blades, with a mixture of analog and/or digital inputs and outputs as required for that particular studio. Each studio is similarly but not identically equipped here—normal. The three local switches are connected to a main central Big Switch, to which other Blades are attached, say for I/O in the rack room or a shared continuity space.

Similarly the PCs attach to the Big Switch. Their purpose is mostly as playout systems (all the program material, music, liners, commercials etc. are recorded as files on a PC—long gone are the days of a Disk Jockey actually playing Disks!), as hosts for the streaming codecs feeding the Internet, and decoding feeds from outside the facility. More basic AoIP systems require a centralized PC to administer the system on a continual basis, while this one (WheatNet IP) only requires a PC for initial setting up and not for operation.

Any of the I/O sources regardless of where they may physically be are available on the network anywhere else. Better executed systems carry the parameters of a source with the audio routing; a microphone source, for example, will let a newly chosen destination know what gain it has had set up for it, any filters applied, and any EQ and dynamics settings etc., so radically simplifying studio reconfiguration.

29.26 In Conclusion…

…the battle between analogue and digital signal processing is no longer; the technical descriptions and images in this chapter adequately convey that the control surface, optimized for a given application, is everything; it is almost invariably rationalized, and the underlying signal processing—analogue, specifically DSP, or DAW-based—is almost immaterial. The best man among many has won—all of them.

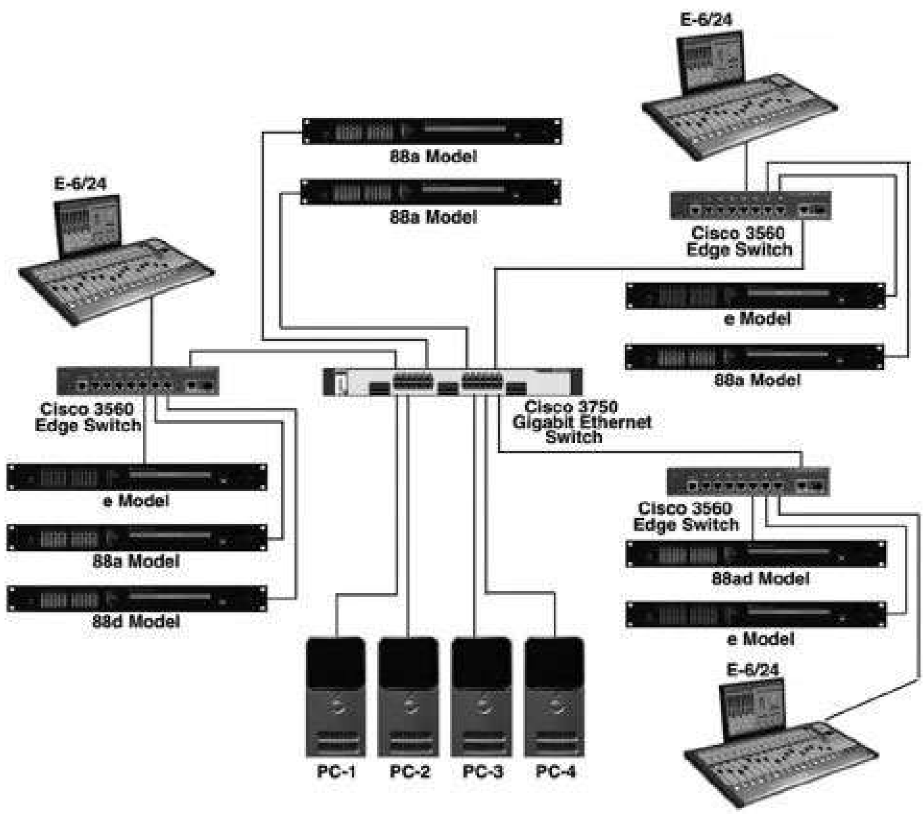


Figure 29-140. A typical AoIP installation.

---

* Note: Processor cycles, not samples; the audio does not take any longer to pass through the DSP just because some parts of the processing are done double-precision.

Chapter 30

Audio Output Meters and Devices

by Glen Ballou

30.1 General

To operate a sound recording or reproducing system properly, some method for determining the signal levels in different parts of the system to avoid overloading, noise, and distortion is required. This is the purpose of the *standard volume indicator (SVI) meter*. A VI meter, more often called a VU meter, is a meter used to measure levels of audio-frequency signals. Until recently, volume unit (VU) meters were devices to measure power with respect to 1 mW of power across a 600 Ω line. Today VU measurements are made with respect to many different bases.

VU meters were first used by the telephone company. They were used to measure the level of the signal being sent down the line. The lines were open wire pair of AWG #6 wire spaced 12in apart, which translated to a characteristic impedance of 600 Ω as determined with the equation

$$Z = 276\log\left(\frac{2D}{d}\right) \tag{30-1}$$

where,
D is the spacing of the two wires,
d is the diameter of the wire.

Today, most amplifying devices have a high-impedance input and a low-impedance output as specified by a 1978 I.E.C. standard requiring the output impedance of a device to be less than 50 Ω and the input impedance to be greater than 10Ω. Since very little power is transferred between 50 Ω and 10kΩ, it makes more sense to make measurements as voltage gain rather than power gain.

It is important to know what kind of measurement reference is being used. The following are some of the common references:

dBm. It is power level in dB referenced to 0dB or 1 mW and a 600 Ω load.

dBW. Power referenced to 1 watt.

dBf. Power referenced to 1femtowatt (1×10^{-15}W).

dBV. Voltage referenced to 1Vrms. dBV is used in voltage coupled systems while the dBm is used in power coupled systems.

−10 dBV. A voltage reference level used by many consumer products and is equal to 0.316 Vrms.

dBu. Voltage referenced to 0.775Vrms. It is not affected by impedance. The u stands for unterminated.

+4 dBu. The pro-audio voltage reference level for most countries of 1.23Vrms. Germany, Austria, and Switzerland use +6dBu and Great Britain uses +8dBu reference.

dB FS. Digital audio reference level equal to full scale, which is the maximum peak voltage level before digital clipping of a data converter. Reference level of a sine wave peaking at digital full scale can be found with

$$\text{Signal level (dB FS)} = 20\log(A/B) \qquad (30\text{-}2)$$

where,

A is the amplitude of the signal whose level is to be determined,

B is the amplitude of a sine wave that corresponds to full-scale amplitude.

dBA. An unofficial method of stating loudness measurements using the “A” weighted curve on a sound level meter.

dBC. An unofficial method of stating loudness measurements using the “C” weighted curve on a sound level meter.

dB-SPL. Sound pressure level referenced to 0.0002 μbar where 1 μbar = 1 dyne/cm^2 or the threshold of hearing.

dBr. An arbitrary reference level that must be specified. It can be used for many different references as long as it is specified.

dBTP. Meters that use an oversampled sampling rate of at least 192 kHz, should indicate the result in the units of dB TP (dB True Peak). This designation signifies decibels relative to 100% full scale, true-peak measurement. (ITU 1770-3).

DIN Scale. The DIN scale as used in Germany and Austria uses +6dBu as the reference level for the 0dB mark. This is equivalent to 1.55Vrms.

30.2 Standard VU Meters

A *volume unit (VU) meter* is a special form of VI meter used for monitoring broadcast, recording circuits and sound reinforcement systems. Such meters employ special ballistics that average out complex waveforms to properly indicate program material that varies simultaneously in both amplitude and frequency. For complex waveforms, such as speech, a VU meter reads between the average and the peak values of the complex wave. No simple relationship exists between volume measured in VU and the power of a complex waveform. The indicated reading will depend on the particular wave shape at the moment. For sine-wave measurements, a change of one VU is numerically equal to a change of 1dB.

VU meters are designed to have a dynamic characteristic that approximates the response of the human ear. When a speech waveform is applied to a VU meter, the movement will indicate peaks and valleys in the signal. The average of the three highest peaks in 10 s (disregarding occasional extremes) is taken to be the indication of the meter movement.

Many meters marked as VU meters are not actually such meters, since they do not have the special ballistics and characteristics of the standard VU meter.

The *VU* meter is a device whose standard has remained the same since 1961. The meter consists of a $200mA_{dc}$ D'Arsonval movement fed from a full-wave, copper-oxide rectifier mounted within the meter case. VU meters are calibrated in reference to 1 mW of power into a 600 Ω load. A typical moving coil VU meter is shown in Fig. 30-1.

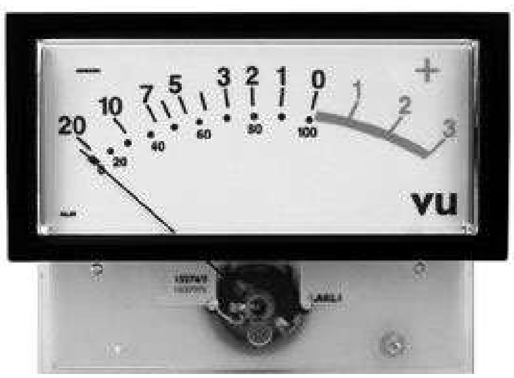
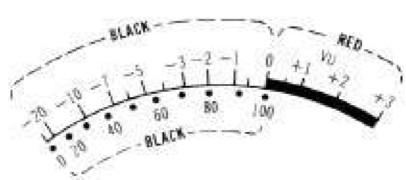


Figure 30-1. Moving coil VU meter.

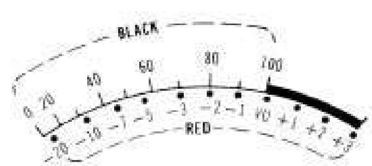
In the 1920s and 30s copper-oxide rectifier power-level meters were inaccurate and not satisfactory for program monitoring. The development of an entirely new meter was jointly undertaken by the Bell Telephone Laboratories, Columbia Broadcasting System (CBS), and the National Broadcasting Company (NBC). The results of this research were not only the development of a new type VI meter but also the standardization of a new reference level of 1 mW, a unit that was adopted by the electronics industry in May 1939. The current standard is ANSI C16.5-1961, formerly the Acoustical Society of America (ASA) C16.5-1961.

The characteristics of the dBm VU meter are as follows:

- **General**. The meter consists of a dc meter movement with a full-wave, copper-oxide rectifier unit (mounted in the instrument case) and responds approximately to the root-mean-square (rms) value of the impressed voltage. This value will vary somewhat depending on the waveforms and the percentage of harmonics present in the signal.
- **Instrument Scale**. The face of the instrument may have either of the two scale cards shown in Fig. 30-2. Each card has two scales: a VU scale ranging from –20 to +3 VU and a percent-modulation scale ranging from 0 to 100%, with 100% coinciding with the 0 point on the VU scale. The normal point for reading volume levels is at 0VU or 100%, which are located to the right of the center at about 71% of the full-scale arc.



A. Recording and test equipment.



B. Broadcast monitoring.

Figure 30-2. VU meter scales.

- **Dynamic Characteristics**. With the instrument connected across a 600 Ω external resistance, the sudden application of a sine-wave voltage, sufficient to give a steady-state deflection at the 0VU or 100 scale point, shall cause the pointer to overshoot not less than 1% or more than 1.5% (0.15dB). The pointer shall reach 99 on the percent scale in 0.3 s.
- **Response Versus Frequency**. The instrument sensitivity shall not depart from that at 1kHz by more than 0.2dB between 35Hz and 10kHz, or more than 0. 5dB, between 25 Hz and 16kHz.
- **Impedance**. For bridging across a line, the volume indicator, including the instrument and proper series resistor (3600 Ω), shall have an impedance of 7500 Ω when measured with a sinusoidal voltage sufficient to deflect the meter to 0VU or the 100% scale point.
- **Sensitivity**. The application of a sinusoidal potential of 1.228V (4dB above 1mW in a 600 Ω line) to the instrument in series with the proper resistance (3600 Ω) will cause a deflection to the 0VU or 100% point.
- **Harmonic Distortion**. The harmonic distortion introduced in a 600 Ω circuit, caused by bridging the volume indicator across it, is less than 0.3%, under the worst possible condition (no loss in the variable attenuator).
- **Overload**. The instrument must be capable of withstanding, without injury or effect on the calibration, overload peaks of ten times the voltage equivalent to a reading of 0VU or 100% for 0.50s and a continuous overload of five times that voltage.

30.2.1 Meter Ballistics

Meter ballistics are the mechanical and electrical characteristics built into the meter movement. A given characteristic may be obtained by shaping the pole pieces and counterweighting the pointer mechanism. Shunts are sometimes used across the meter terminals, but this use will reduce the sensitivity of the movement.

The ballistics characteristics of a typical old-style VI meter or voltmeter and a standard VU meter, when a 1000 Hz signal is applied for a period of 1 s, are shown in Fig. 30-3. Note the VU meter comes to a steady state at the end of 0.30s, while the VI meter continues to oscillate showing peaks and valleys over a period of 1 s. An ac voltmeter would be even worse than the old style VI meter as it would never settle down and would constantly overshoot. This clearly indicates why the ballistics of the VU meter are desirable for monitoring program material containing complex waveforms.

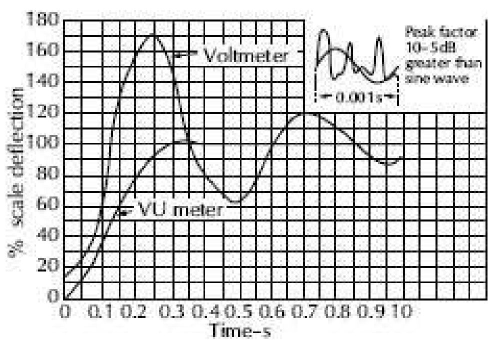


Figure 30-3. Comparison of the original VI meter and the present VU meter ballistics when a 1000Hz signal is applied for 1 s.

A VU meter reads the rms value of the waveform. On a sine wave, the rms VU indicator of the peak is only 3 dB above the reading; however, on voice or music, the peak may be 10 to 12dB above the VU reading. This difference is called the *crest factor* and is illustrated in Fig. 30-4.

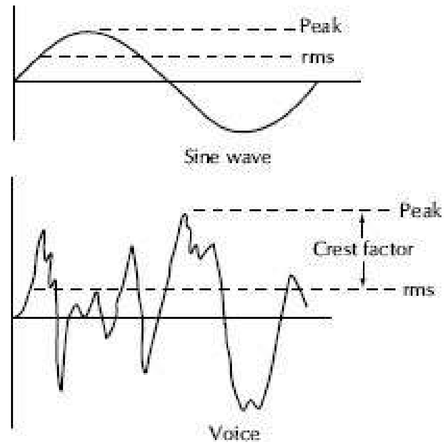


Figure 30-4. Crest factor caused by the peak of music or voice being greater than $\sqrt{3}$ rms.

Because of the meter ballistics, a VU meter indicates somewhere between the average and the peak values. Program material is of a complex and transient nature; therefore, the VU meter reading is considerably under the instantaneous peak program level. This means that 8 to 14dB peaks present in the program material are not indicated by the meter because the meter movement cannot follow small instantaneous peaks. Even if they could be seen, it would be too late to reduce the level. Therefore, the meter must either be set or caused to indicate in a manner that will not overload the system in which it is operating.

Since VU meters do not include the true peak values of program material (complex waveforms), it is quite easy to overload a recording system. To protect against these unseen peaks, a lead or margin of safety is inserted in the VU meter circuit.

To insert a lead into a VU meter circuit, the VU meter is connected across a bridging bus with a sine-wave level of +14dBm. A 400Hz or 1000Hz signal is sent into the input of the recording console. The mixer control is set to its normal operating range, and the signal level is adjusted to bring the bus level to +14dBm (the VU meter reads 100% or 0dBm).

Remove the input signal and return the VU meter attenuator to its +6dBm position. This inserts an 8dB lead or margin of safety in the VU meter by making it 8dB more sensitive, protecting the system against unseen peaks up to 8dB. The program material is now mixed in the usual manner. Some recording activities, because of the heavy peaks and overloads encountered in some types of music, use a 10 to 12dB lead in the VU meter.

Radio transmitters are adjusted in a similar manner. However, in this instance, the percent modulation indicated by the VU meter indicates the percent modulation of the radio transmitter.

30.2.2 Reference Levels

In the early days of broadcasting and recording, both 10mW and 12.5mW into a 500 Ω line were used as a reference level. However, later this was changed to 6mW. In May 1939 the present standard of 1mW into a 600 Ω line was adopted. This reference level was selected as a level that would conform to the telephone company's standards of limiting the signal level on a transmission line to a value that would produce a minimum of crosstalk and still provide a satisfactory signal-to-noise ratio (*SNR*). The 1 mW reference level is a unit quantity and is readily applicable to the decimal system, being related to the watt by the factor 10^{-3}.

Zero level is a reference power level of 1 mW of power into a 600 Ω load. This is equivalent to a voltage of 0.775 V.

30.2.3 VU Meter Impedance

The VU meter and its attenuator impress a 7500 Ω impedance onto a circuit. The VU meter system consists of an indicator movement, a variable attenuator, and a series resistor of 3600Ω, Fig. 30-5. Meter manufacturers supply only the meter movement; the external circuitry is added later. A 200 pA D'Arsonval meter movement with an internal resistance of 3900 Ω and a full-wave, copper-oxide or selenium rectifier are contained within the meter case. The attenuator is variable in steps of 2dB, presents a constant impedance of 3900 Ω to the meter movement, and prevents the ballistics of the meter from being affected when the attenuator setting is changed.

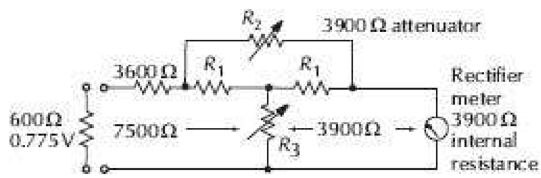


Figure 30-5. A 7500 Ω VU meter, calibrated for 1 mW reference level or 0.775V across 600Ω.

Standard VU meters are designed to read 0VU, or 100%, with 1.228 V (+4dBm) applied to the instrument. If the meter is used with the attenuator but without the 3600 Ω series resistor and is connected across a 600 Ω load in which 1mW of power is flowing, the movement will be deflected to the 100% calibration point. This method is not recommended since the impedance looking back into the meter is only 3900 Ω and loads the 600 Ω circuit. It is the usual practice to keep the impedance of bridging devices at a ratio of 10:1 or greater.

Increasing the input impedance of the VU meter from 3900 Ω to 7500 Ω creates a 4dB loss across the 3600 Ω resistor. If a signal of 1mW (0.775V) is impressed across the input terminals of the circuit in Fig. 30-6, it will not deflect the meter to the 0VU calibration but only to the −4VU (or decibel) mark, or approximately 65%. This means that if the meter is to be deflected to the 100% point, the input signal must be increased to a +4dBm. This is the reason why 1mW of power will be indicated at the −4 dB calibration mark.

Attenuators used with VU meters start at a +4dBm. The bridging loss caused by the VU meter being inserted into the circuit is the drop in signal level caused by the absorption of power by the meter circuit. As a rule, the power absorbed is quite small and may be ignored. However, at high powers, it may become important. Bridging loss may be calculated by the equation

$$dB_{loss} = 20\log\frac{2B_R + Z}{2B_R} \tag{30-3}$$

where,

B_R is the VU meter input impedance,

Z is the line impedance.

A 7500 Ω VU meter has a bridging loss of 0.34dB.

30.2.4 VU Impedance Level Correction

VU meters are calibrated for 1 mW of power across a 600 Ω load as −4VU, therefore when a VU meter is connected across any other impedance, a correction must be added to the indicated reading to give a proper VU reading. The equation for the level correction is

$$dB_{corr} = 10\log\frac{Z_2}{Z_1} \tag{30-4}$$

where,

dB_{corr} is the decibel amount added to the VU reading,

Z_2 is the impedance for which the meter is calibrated,
Z_1 is the impedance of the circuit bridged.

A typical example of applying a correction factor is as follows: a VU meter calibrated for a line impedance of 600 Ω is bridged across a 16 Ω loudspeaker line and indicates a level of +1 dBm. The true VU would be

$$VU = 1\text{ dBm} + correction\ factor \qquad (30\text{-}5)$$

The correction factor from Eq. 30-4 is

$$\begin{aligned} dB_{corr} &= 10\log\frac{600}{16} \\ &= 10 \times 1.574 \\ &= 15.74\text{ dB} \end{aligned}$$

The correction factor of 15.74dB is added to the meter reading of +1dBm for a true level reading of +16.74 dBm. Typical correction factors are shown in Table 30-1.

Table 30-1 Correction Factors in dBm to be Applied to a VU Meter When Connected Across an Impedance other than 600 Ω

Line Impedance-Ω	Meter Cal 600 Ω-dB
10,000	−12.22
5000	−9.21
2500	−6.20
1000	−2.22
600	0.000
500	+0.791
250	+3.800
200	+4.770
150	+6.020

125	+6.810
100	+7.780
50	+10.790
30	+13.010
16	+15.740
15	+16.020
8	+18.750
4	+21.760

If a VU meter is connected across a line impedance different from that for which it was originally calibrated, the voltage supplied to the meter will either be lower or higher than the original calibration; therefore, the meter would indicate incorrectly. Two circuits are shown in Fig. 30-6, one a 600Ω circuit and the other a 16Ω circuit. Both are dissipating the same amount of power; yet the voltage for the 600 Ω circuit is 0.775 V, and for the 16 Ω circuit it is 0.127 V. As can be seen, if a VU meter is connected across the 16 Ω circuit, it will not deflect the same amount as for the 600Ω circuit, although the same amount of power is flowing in each circuit. To arrive at the correct power level in the 16 Ω circuit, a correction factor must be applied to the meter indication.

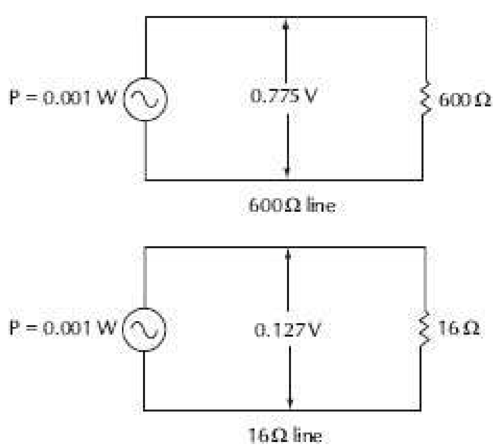


Figure 30-6. Voltage across lines of different impedance but with

the same power in milliwatts.

30.2.5 Voltages at Various Impedances

If the line voltage for a given level at 600 Ω is known, voltages for other line impedances may be calculated using

$$V_x = V\sqrt{\frac{Z}{600}} \tag{30-6}$$

where,
V_x is the unknown voltage,
V is the voltage for 600 Ω,
Z is the new impedance.

As an example, assume voltage V_x is required for a line impedance of 150 Ω at a level of +4dBm. Referring to Fig. 30-7, the voltage for a level of +4dBm at 600Ω is 1.23 V. The new voltage may now be calculated using

$$V_x = 123\sqrt{\frac{150}{600}}$$

$$= 0.615\text{ V}$$

Voltages for a line impedance of 600 Ω for levels between 0 and +50dBm may be taken from Fig. 30-7.

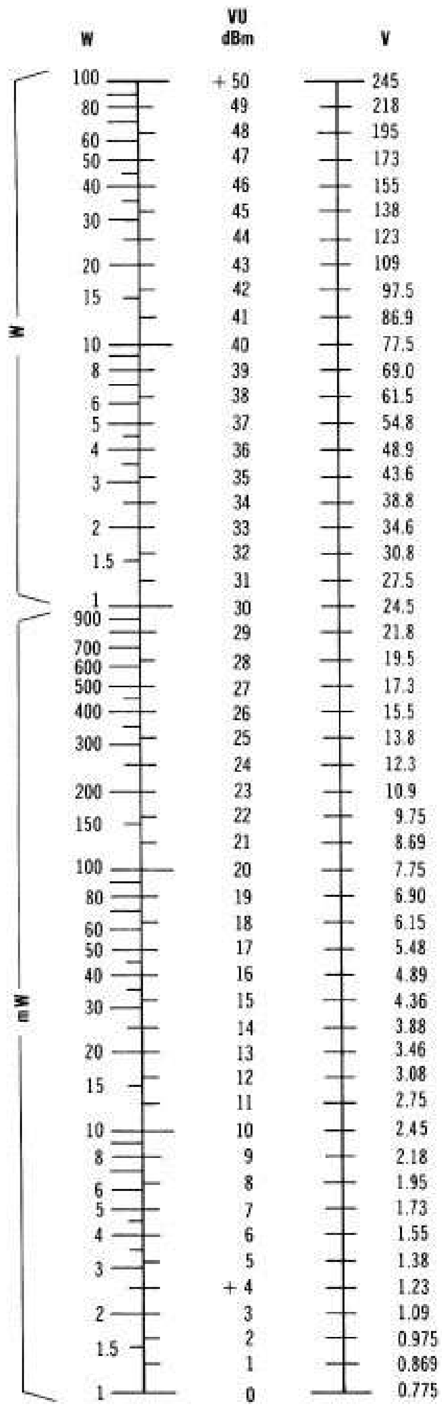


Figure 30-7. Relationship of VU and dBm to power in watts and voltage in a 600 Ω line.

Voltage across 600 Ω can be calculated from dBm with the following equation

$$V = 0.6 \times 10^{\frac{dBm}{10}} \qquad (30\text{-}7)$$

30.3 Wide-Range VU Meters

Standard VU meters measure only the upper 23 dB of the signal level. From the practical standpoint, this limits the display to about 20dB below the reference level of the 0 indication.

This short range of operation limits its usefulness, particularly when it is connected across a bridging bus for monitoring program information. A *wide-range program-monitor meter*, Fig. 30-8, displays the program information over a 60dB meter scale, spread from −57dB to +3dB. The large spread of program material permits the very low-level signals to be observed as well as the noise between program pauses. The wide-range VU meter was not designed to replace the conventional VU meter; however, its characteristics are compatible with the VU meter. In addition, a dc output is provided for connection to a linear tape recorder for logging program levels over a range of 60dB. The 0dB indication may be set to represent a reference level from −22dBm to +18dBm.

The basic component is a logarithmic amplifier, Fig. 30-9, with a nonlinear feedback circuit, a preamplifier, a 15kΩ bridging input transformer, a reference-level selector switch, and a sensitive indicating meter movement.

30.4 Bar Graph VU and Spectrum Analyzers

The United Recording Electronics Industries (UREI) Model 970 Vidigraf is a bar graph display generator that operates any National Television System Committee (NTSC) standard video monitor or (with an inexpensive accessory) black-and-white television receiver.

The system provides both a VU level display and the frequency-spectrum-level information. It is designed primarily for multitrack recording studio applications. However, its dc to 20 kHz input capability suggests its use for a wide range of dc or ac analog voltage measurements.

The 970 Vidigraf's modular construction provides users with complete flexibility to adapt the system to their specific needs. A maximum of four 16-channel input display modules may be installed for VU level, automation control voltages, or frequency-spectrum viewing. Each module may be individually switched to the video generator in the single mode. In the dual display mode, the screen is split vertically to accommodate the information from any two input modules simultaneously. Instantaneous identification of the input channel sources and/or frequencies, as well as vertical scaling indices are automatically provided by the built-in programmable character generators. This eliminates any need for screen overlays or masks and ensures accurate positioning of the alphanumeric information regardless of screen size or width and height adjustments.

Figure 30-8. Wide range VU meter. Courtesy Dorrough Electronics.

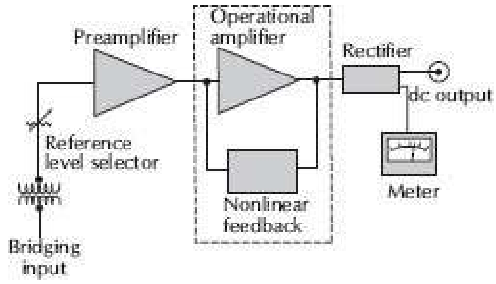


Figure 30-9. Block diagram of a wide-range program monitor VU meter.

Some typical displays are:

- 6 or 32 simultaneous VU channels.
- 16 or 2 × 16 bands of frequency spectrum (1 or 2 channels).
- 16 VU channels, plus channels of automation control voltages.
- 16 VU channels, plus 15 bands of frequency spectrum and 1 composite level.

One *VU module* provides 16 bar graphs with standard VU ballistics over a display range of 30dB. Each bar has two shades of gray, with the lighter shade above the 0dB reference. When a signal is applied to any of the 16 inputs, a bright bar moves up and down with the signal level. The 0dB reference point can be continuously adjusted to any standard from 0 to +8dB. The VU module is user programmable to display a logarithmic scale of −20dB to +3dB when measuring audio signals or to read linearly from 0 to 10 for display of ac or automation dc control voltages.

The *spectrum module* provides a visual real-time display of VU level versus frequency of an audio signal. It is useful for setting equalization and adjusting frequency balance. This module provides 16 bar graphs with visual characteristics similar to those of the VU module. One bar is assigned to the full spectrum of the audio signal, and the other 15 channels display increments of the frequency spectrum, centered on standard ISO 2/3-octave filter frequencies. Two independent controls adjust the level of the full spectrum bar relative to the spectrum analysis bars.

30.5 Power-Level Meters

A *power-level meter* is a VI meter calibrated in decibels. As a rule, this type of meter is normally used with test equipment for steady-state measurements and is not used for monitoring program material because its ballistics are more like those of a voltmeter.

30.6 Power-Output Meters

A *power-output meter* is used for measuring the power output of audio amplifiers and other devices. It may also be used to determine the characteristic and internal output impedance, the effect of load-impedance variation, and other applications involving the measurement of output power and impedance with respect to frequency. The power output meter may be calibrated in watts and/or dBm. The power output meter is a test instrument and not used for monitoring program level because of its ballistics.

30.7 Peak Program Meters

The *peak program meter* (PPM) is used extensively in Europe and falls under four standards, the DIN-type DIN 45406, the BBC type, the EBU type, and the Nordic N9 type. These meters measure the peak program signal, which is usually +6dB to +20dB above the readings seen on the VU meter.

30.7.1 DIN 45406 Standard

The PPM is popular in Europe. It is designed to have a fast rise time, 30 times as fast as a VU meter, and a much slower fallback or decay time.

The DIN 45406 and the IEC 268-10 have an integration time of 10ms and a decay time of 1.5 s for 20dB of fallback and 2.5 s for 40dB of decay. The indicator range is −50dB to +5dB. The scale marked for 100% reading is 0dB which is the reference level of +6dBu or 1.55Vrms, Fig. 30-10A.

The RTW 1019GL analog peak program meter + loudness meter + phase correlation meter is shown in Fig. 30-10A. The 127mm (5 inch) 201 element bar graph display has a 1 dB per division scale from +5 dB to −10dB changing logarithmically to −50dB. Roll-off above 20kHz is 12dB/octave. The meter includes a +20dB gain increase and a peak memory/reset circuit. The integration time is selectable between 1 ms and 10 ms. The balanced input is transformer isolated.

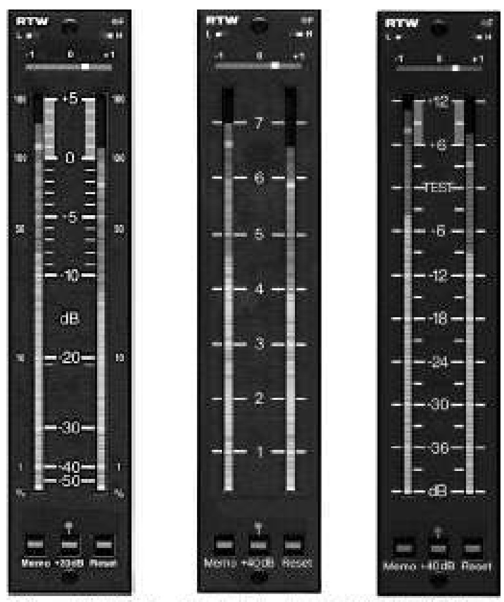
The meter panel also includes a three-color phase correlation value display with memory. The correlator indicates the phase correlation r, of stereo signals. If both channels are in phase, e.g., a mono signal on both channels, the reading is +1 r. With only one or no signal at the input, the meter will read 0r.

30.7.2 British Broadcast Standard

The British Broadcast Standard, BS 55428 Part 9, has an integration time of 12 ms and a decay time of 2.8 s for decay from 7 to 1. The indicator range of 1 s to 7 is equivalent to a −12dB to +12dB. The scale mark for 100% reading is 6 and is referenced to

+8dBu or 1.95 Vrms.

Fig. 30-10B shows an analog RTW 1034GL British standard scale IIa analog peak program meter + loudness meter + phase correlation meter. The 127 mm (5in) 201 element bar graph display measures from –12dB to +12dB. The meter includes a +40dB gain increase and a peak memory/reset circuit. The integration time is selectable between 1ms for digital audio and 10 ms for analog audio. The balanced input is transformer isolated. The meter panel also includes a three color-phase correlation value display with memory.



A. Stereo DIN 45406 peak program meter plus loudness meter plus phase correlation meter.

B. British standard stereo peak program meter plus loudness meter plus phase correlation meter.

C. Nordic N9 stereo peak program meter plus loudness meter plus phase correlation meter

Figure 30-10. European VI standards. Courtesy RTW GmbH & Co.KG, Cologne.

30.7.3 Nordic N9 Standard

The Nordic Recommendation N9 has an integration time of 5ms, a decay time of 1.7s for 20dB, and 3.4s for 40dB of decay. The indicator range is from −42dB to +12dB. The scale mark for 100% reading is 0dB and is referenced to +6dBu or 1.5Vrms, Fig. 30-10C.

Fig. 30-10C shows an analog RTW 1039GL Nordic Recommendation N9 analog peak program meter + loudness meter + phase correlation meter. The 127 mm (5in) 201 element bar graph display measures from −42dB to +12dB. The meter includes a +40dB gain increase and a peak memory/reset circuit. The integration time is selectable between 1 ms for digital audio and 5 ms for analog audio. The balanced input is transformer isolated. The meter panel also includes a three-color phase correlation value display with memory.

30.8 AES/EBU Digital Peak Meter

With the advent of digital equipment, new meter standards are being written to work with the AES/EBU digital format. This requires being capable of sampling 32kHz, 44.056kHz, 44.1kHz, 48kHz, and 96kHz with an AES/EBU digital format.

The attack time is one sampling period and the decay time is 1.5 s for a change from 0dB to −20dB. The indicator range is from 0dB to −60dB.

The RTW 11529G digital peak program meter + loudness meter + phase correlation meter, Fig. 30-11A, has a 127 mm (5in) 201 element bargraph display. Its sampling rates are from 27kHz to 96 kHz and it includes a dc filter and has indicators for 44.1kHz, 48kHz, and 96 kHz, emphasis, error, and overload. The meter also includes peak memory, peak hold, +40dB gain, and a three-color correlation correction value display.

The RTW 1 1528G AES/EBU Digital PPM, Fig. 30-11B is especially useful for radio and TV broadcasting applications. The meter features AES/EBU inputs and outputs. The digital signal can be displayed once as it is without any weighting (sample precise display), which corresponds to the digital standard that has a scale range from −60dB to +9dB but with a fixed head room of −9dB FS, which is marked 0dB, and highlighted and superimposed with an integration time of 10ms. It can also be displayed with a superimposed and highlighted loudness display. Finally, it can be shown as a 10 ms integration time–only function, as a quasi analog display. Its sampling rates are from 27kHz to 96 kHz.

A. RTW 11529G Digital Peak Program Meter + Loudness Meter + Phase Correlation Meter.

B. RTW 11528G Digital Peak Program Meter + Loudness Meter + Phase Correlation Meter.

Figure 30-11. An AES/EBU peak program meter plus loudness meter plus phase correlation meter. Courtesy RTW GmbH & Co.KG, Cologne.

Both the RTW 11529G and the 11528G include an over indication with a selectable overload detector range, 9 to 24 bit overload response word length, and number of overload samples.

The trend of excessive levels near 0dBFS in combination with data reduction systems revealed the problem of distortion and intermodulation insertion in digital audio streams which cannot be foreseen or prevented by the use of digital sample peak meters.

Fig. 30-12 displays a 3kHz analog signal sampled with 48kHz. Zooming in it can easily be seen that the sample points are not taken at the true peak point of the waveform but instead 0.17dB below. With higher frequencies this is even worse. Such a signal normalize to 0dBFS would cause overload distortion at least when it comes to a sample rate converter or D/A. To prevent such artifacts ITU recommends the use of a true peak meter measurement. The unit for such measures is dBTP.

The meter consist of an oversampling stage of at least 4 times the basic sample rate, (for SR = 48kHz → 192kHz), Fig. 30-13.

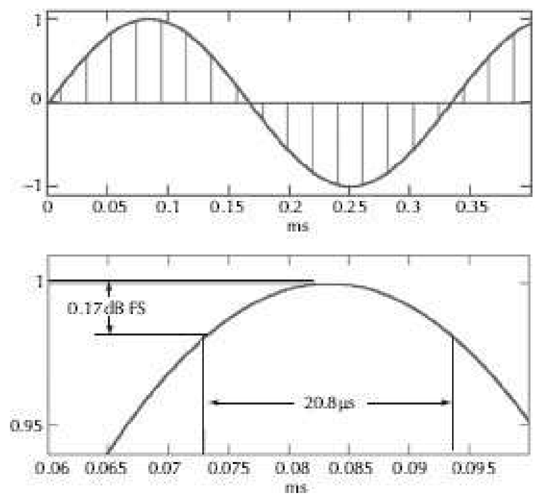


Figure 30-12. Example of a 3kHz sine wave sampled at 48 kHz. Courtesy RTW GmbH & Co.KG, Cologne.

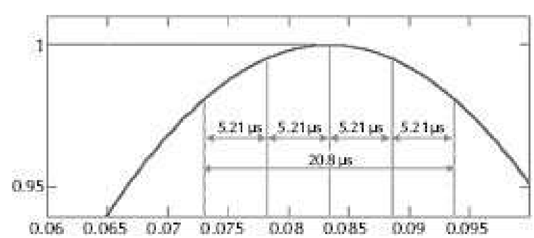


Figure 30-13. Oversampling 4× the base rate. Courtesy RTW GmbH & Co.KG, Cologne.

This leads to a better description of the waveform and a better assumption of the peak value. 4-times oversampling was selected as being the best compromise between preciseness and required resources for the measurement.

To be on the safe side recommendations advice limits like −1 dBTP for linear systems and −3dBTP for data reduction systems and for heavy data reduction even higher like −5 dBTP. Companies like RTW have added this measurement to all actual devices. Meters with this behavior are indicated by the max upper value of +3 dB.

It is required by the ITU standard that such a measurement must be measured in a synchronized environment without an input sample rate converter of the measurement device as that would already cause measurement failures.

30.9 Loudness Meters

Legacy loudness meters placed VU and PPM meters on a single panel, providing an indication of the entire dynamic condition of the signal. It also eliminates the condition that eyeball wobble could develop in the attempt to follow two adjacent meters with differing ballistics. The use of two pointers with such differing ballistics on a single scale would demonstrate that the PPM would read consistently higher levels than the VU meter, and the large differential of decay with respect to rise time of the PPM in comparison to the equal rise and decay times of the VU meter would also be difficult to interpret.

Three types of scales were used on loudness meters:

- Based on +14dB of headroom.
- Referenced at 100% for broadcast transmission.
- Based on 20dB of headroom.

The head room available to mixers in postproduction is not the same as allowed in broadcast. The U.S. standard in digital (SMPTE) is 20dB below FS (full scale) and the EBU standard used in European and many Middle Eastern countries is −18dB below FS. When film and post material is sent to the broadcast facility, the peak shall not exceed +12dB analog or −8dB digital.

The music and recording industries do not have these requirements for their products, and therefore use the full dynamic range. Commonly, this material will peak fairly consistently at −1 dB on a digital reading meter, with the bar graph fairly consistently four or five LEDs under the peak. If this material makes its way to broadcast, it will be louder than audio on video by possibly as much as 8dB. The result might be rejection and need to redo the material using the guidelines required, or quality control at the broadcast facility will make a judgement on the loudness and turn it down accordingly. *Dialnorm* on HDTV was designed for these irregularities.

Observation of complex audio signals with an oscilloscope indicates the peak excursions of the program material. The use of an oscilloscope with a long or variable persistence CRT will show additional information relative to recurrent amplitude displayed by the persistence of the screen as a concentrated band of energy about the center of the CRT. It is these two pieces of information that provide the composition of acoustically related peak to quasi-average information. The meters shown in Fig. 30-11 feature a loudness display and peak holding display on each bargraph.

Fig. 30-14A is an analog reading meter, Fig. 30-14B is a digital reading meter, Fig. 30-14C is a remote control unit to access features from both. The remote push buttons control the following functions:

- Left/Right.
- Sum/Difference.
- Phase.
- Overs Display with Overs Reset.
- Three second Peak Hold.
- Peak Hold Permanent.
- Reference Mode.

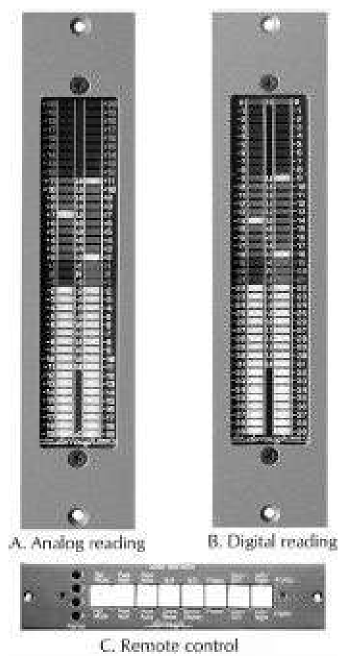


Figure 30-14. The meter scale for a gain-riding device with both a normal persistence range (much like VU readings) and a normal peak range. Courtesy Dorrough Electronics.

The alarms on the left side of the remote are for Phase Error, Bit Stream Corruption, and Full Scale. Fig. 30-15 is a block diagram of the loudness meter of Fig. 30-14.

With signal input both meters read the audio in the same dynamic way. Each meter displays 20dB of Peak Amplitude above the 0 VU persistence reference level, and therefore, 0 reference is the same on both.

The use of a dot display for Peak information and a bar display for Persistence information allows a single display for both ballistics. Each lamp in the display is therefore driven by two drivers, one for peak, the other for persistence. This representation presents a display of a dot riding on top of a bar graph. In order to make this display useful, there has to be a meaningful relationship between the two ballistics. The peak display has a rise time of two time constants, or 10μs, which is 1000 times faster than the PPM. The decay time for the peak display is 18ms/dB.

Equal energy, properly weighted for program material, will be discerned as equal in loudness. Since energy can be displayed as a function of amplitude and time, an oscilloscope can be used to confirm that large amplitudes of short tone bursts can be equal to longer tone bursts of lower amplitude.

Average power is defined as equal to the area under the curve divided by the time interval. Since the area is equal to the input energy, W or watts, during the interval, Thus, the averaging-type metering can provide an indication of power.

$$P_{ave} = \frac{W}{t} \tag{30-8}$$

where,
W is in W,
t is in s.

The persistence display has a time constant of 270ms and a rise time of approximately 600ms or twice as long as that of a VU meter.

Perceived loudness to the ear from source to source is determined by which circuit (peak or persistence) is first to illuminate its respective set of red LEDs. Program adjustments for equally perceived loudness should be holding either the peak or persistence excursions to its corresponding red LED area.

The relative loudness characteristic of the Peak to the bar graph has been retained by way of red LED reference points on both meters. This is the window of 12dB of separation of the Peak from the Average for maintaining equal loudness. The +12dB analog and −8dB digital are the same scale points on both meters.

30.10 Surround Sound Analyzer

The surround sound analyzer method translates the important details of surround signals into a graphical display suited for instant evaluation. Successful mixing of surround signals is important. Besides the artistic and aesthetical aspects, there are fundamental technical preconditions for obtaining professional results.

Reality is often far from the ideal, particularly during live broadcasts and in audio production for video or TV. This makes it even more important to know, even in the most hectic working environment, how the surround mix will be perceived by the listener.

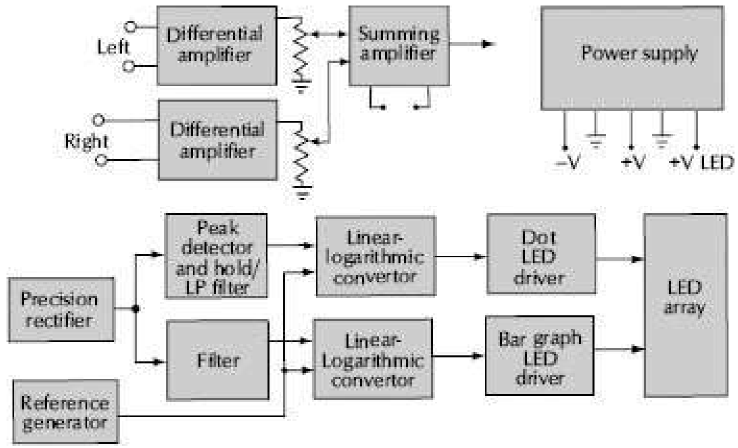


Figure 30-15. Block diagram of a Dorrough loudness monitor. Courtesy Dorrough Electronics.

The RTW Surround Sound Analyzer which is integrated in the RTW SurroundMonitor, Fig. 30-16, is a unique tool showing all the important parameters of a surround signal at a glance. It gives detailed information for all individual channels as well as the overall effect of the mix.

Figure 30-16. RTW SurroundMonitor 11900 for AES/EBU digital and analog standards with Surround Sound Analyzer. Courtesy RTW GmbH & Co.KG, Cologne.

The visual display of the Surround Sound Analyzer provides level and phase relations of all channels. The dynamic response of the display elements is a direct representation of the acoustic image so the balance of the surround sound can be observed.

The volumes of the four channels L, R, sL, and sR are displayed as diagonal white level bars originating from a common center point. Their tips are connected through cyan lines. The square formed by this figure, the total volume indicator (TVI), is a direct measure of the total volume and the balance of the acoustic image.

The curvature of these lines shows the channel correlation, positive values through an outward deflection (roof), negative values through an inward deflection (funnel).

The volume of the Center channel is indicated by another upwards-pointing level bar with yellow connecting lines, showing the perceptibility and dominance of the Center in relation to L and R.

Direction and width of front, side, and rear phantom sources are represented by lines between the loudspeaker symbols, called the Phantom Source Indicators (PSI). Their color changes with channel correlation. A separate correlation indicator for the two surround channels is available at the bottom of the display.

A cross representing the dominance vector indicates the position of the subjectively perceived center of gravity of the mix.

The Surround Sound Analyzer displays a correctly-scaled graphical representation of the relative volumes in the surround sound field. The interaction of levels (volume or sound pressure level) and the correlation of all channels in the production of the overall surround sound are displayed graphically.

The display in the Surround Sound Analyzer can be set to correspond to the volume or the reference sound pressure level by calibrating the instrument and the studio monitoring equipment accordingly. The axes of the 45° coordinate system use a dB volume level or dB-SPL scale and have a reference mark that is also displayed in the volume level and SPL displays in the peak program meter of the instrument. The balance between the Center channel and the L and R front channels is critical for all surround sound productions. The Center channel is displayed with its own display elements to show the volume differences between the Center channel and the L and R channels.

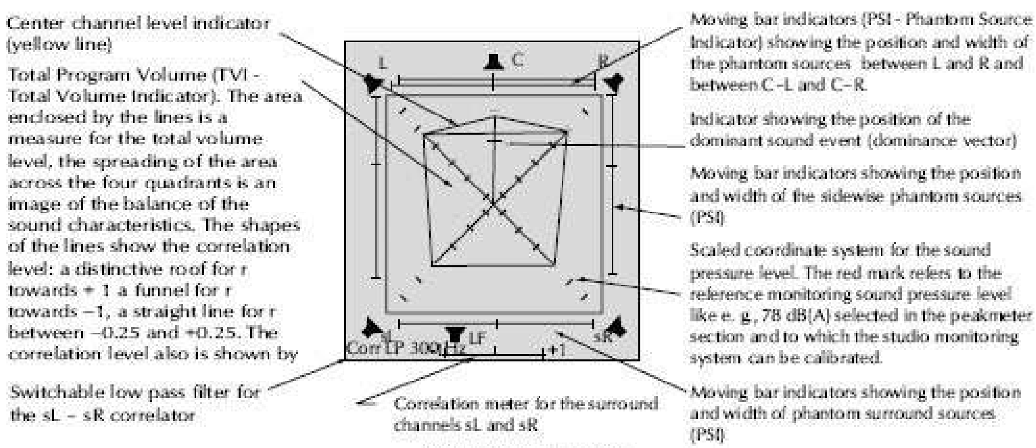
In addition to the signal level, the correlation level (and aligned to that the generation and the location of phantom sound sources) is important for multichannel sound productions, primarily in relation to downmixes or possible sound-faking erasements when generating a mono signal. The correlation level of the sL and sR surround channels is also important. Highly anomalous frequency-dependent correlation levels induce an unimpressive envelopment effect of the sL and sR surround signals. For monitoring this, the Surround Sound Analyzer features a correlation meter for the sL and sR surround channels. Fig. 30-17 shows the display and examples of various patterns.

30.11 Loudness Metering and BS1770

Legislative interest in the USA over widely varying TV program and commercial segment loudnesses, resulting in the restrictive CALM act, broadly coincided with industry efforts to simply yet reasonably accurately quantify loudness in broadcast and production environments. The resulting similar specifications ITU-R BS1770 and EBU R 128 addressed this; meters to these standards are now de rigueur in TV sound suites, and on sound-for-TV audio consoles.

30.11.1 Long-term Indication

The basic intention was to provide a single-number indication in Loudness Units LU, or LUFS signifying “Full Scale,” of an entire end-to-end program segment’s loudness to ideally match a target number, and more particularly, not meaningfully exceed it. Practically, in order to give an operator a “live” indication of how well he is achieving this end, a medium-term rolling average over a period of 3s called short-term reading (S) is displayed to him, either in conjunction with or supplanting the long-term segment average figure.



Actual screens are in color.

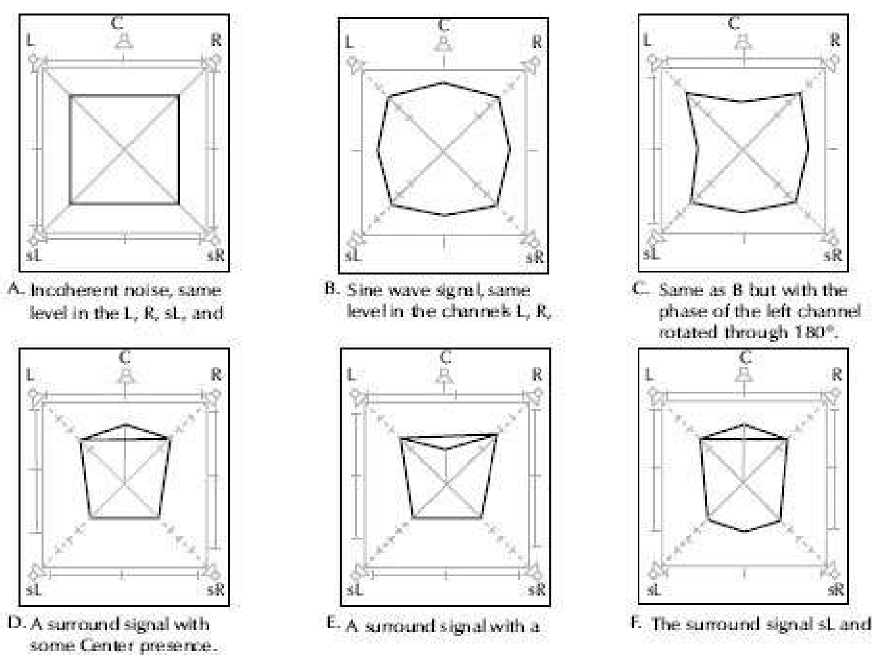


Figure 30-17. Surround Sound Analyzer screen views. Courtesy RTW GmbH & Co.KG, Cologne.

As can be seen in Fig. 30-18, a console implementation, the Big Number in LU, which by virtue of its long integration time changes infrequently, is often supported by other measurements. BS1770 suggests over-sampled peak sensing in order to minimize reconstruction-error indications; the resulting familiarly behaving instant-attack/slow-release peak indicator is in the leftmost vertical bargraph; an overage is flash-indicated above the column. The second bargraph has quasi-syllabic-rate dynamics (similar to a classic VU meter) while the third column is at a selectable medium-length integration (typically between 3 and 30 s) for a "live" loudness indication, regardless of what the Big Number is showing.

Figure 30-18. BS1770 console implementation. Courtesy Wheatstone Corporation.

The combination of all affords the tools to remain within the electrical bounds of the system, be highly aware of the long-term loudness, and achieve this goal consistently. To this end also, the horizontal bargraph beneath The Big Number provides a corner-of-the-eye windowed higher/lower indication as well.

Other full-featured and embellished implementations include that in Fig. 30-19, in which a useful loudness history is diagrammatically displayed clock-fashion; additionally, the component source levels are shown, since this is intended as a complete metering solution, rather than just for loudness.

Figure 30-19. A novel BS1 770 stand-alone and DAW plug-in implementation. Courtesy TC Electronic.

30.11.2 Constructing the Data

Each incoming audio stream (one for mono, a pair for stereo, or five for surround; LFE or sub-low channels being ignored) is filtered according to Fig. 30-20 ("K"-weighting) providing spherical boundary (head shape) correction at the high-mids, and by a 70Hz high-pass filter which nods to the ear's relative insensitivity to low frequencies. Each stream is then mean-squared (so detected) and summed together with additional weighting given to any surrounds' contributions. This combined sample is averaged in successive 100ms segments; the syllabic-rate (400ms) metric is a rolling average of the four most recent 100ms segments ("gated block"); similarly longer integration times can be built up of suitable numbered rolling averages of blocks; ultimately the scheme can average blocks continually until told to stop for total program loudness, Fig. 30-21. The displayed average reading is said to be in LKFS (Loudness, K-weighted, referred to Full-Scale) although LU is the term commonly applied to the end result, Fig. 30-22.

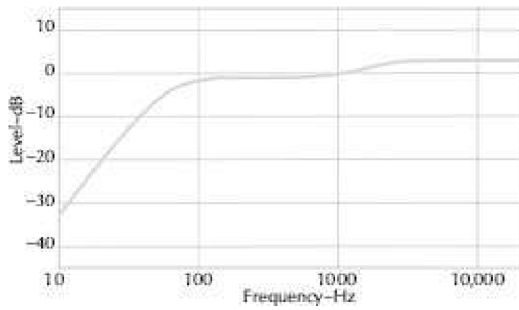


Figure 30-20. K-filter shape as specified in ITU (including preemphasis.

To be precise, LKFS is the term used in the ITU BS.1770 standard where LUFS is the unit term in the EBU R128 standard. There is no difference in the value just the wording for this absolute measurement, −24 LKFS equals −24 LUFS. Relative measurements are defined in all loudness standards with LU as the unit.

30.11.3 Making Consistent Readings

The frequency-response-shaped and mean-square detection scheme has shown strong correlation with perceived loudness in extensive listening tests. The long averaging time means the meter is insensitive to short-term impulses, such as gunshots, allowing deliberate and desired dynamics to prevail that shorter-term measurement and control would preclude. To a similar end, the measurement is self-gating such that deliberate near-silences (moody pauses in dialogue, say, or not) do not cause the meter to start to "droop"—falsely under-read. Essentially, if the present "block" value is 10dB or more below the LU value (8dB for EBU) long-term averaging is halted and LU remains frozen until valid levels re-appear.

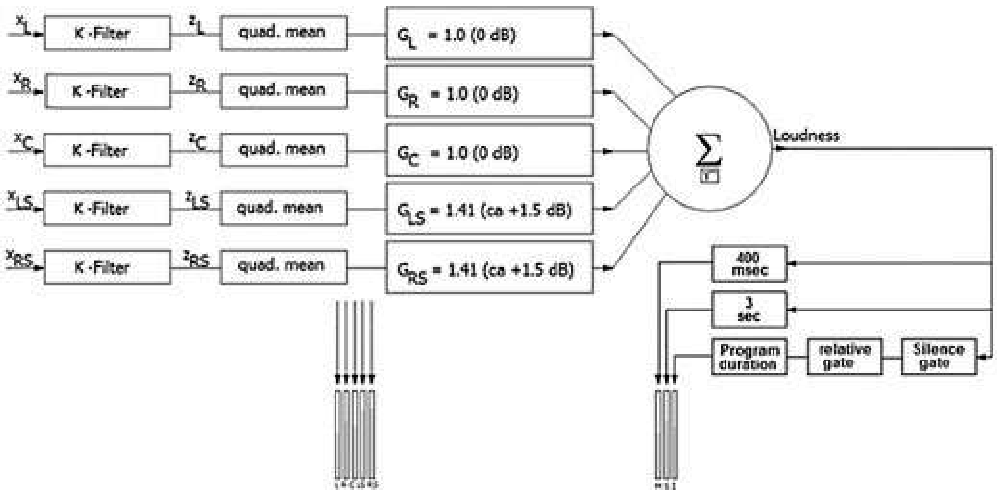


Figure 30-21. Simplified block diagram of a RTW TM (TouchMonitor). Courtesy RTW GmbH & Co.KG, Cologne.

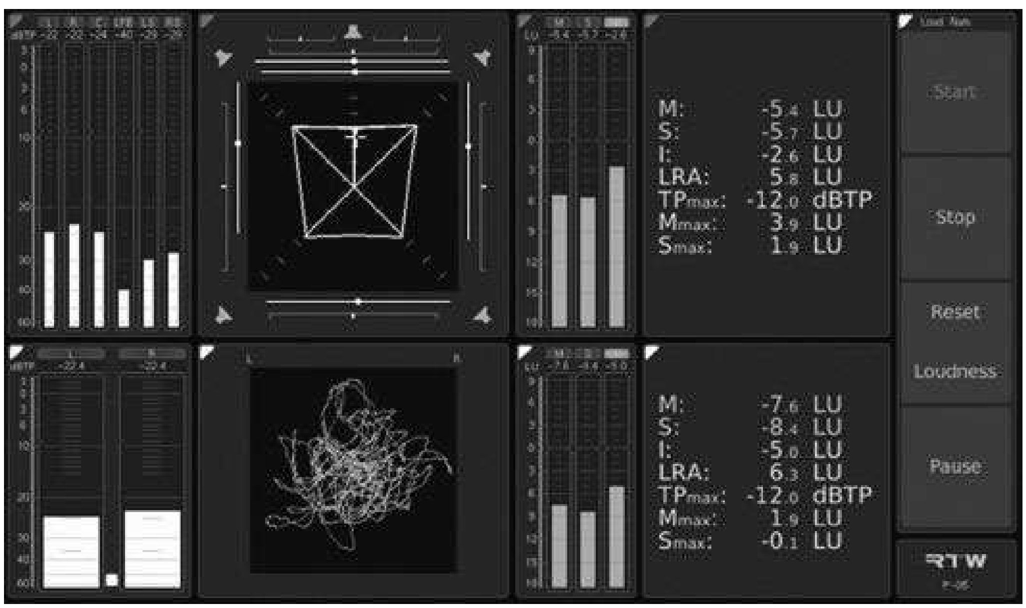


Figure 30-22. Screenshot of a RTW TM9 TouchMonitor displaying a 5.1 and a stereo signal simultaneously for level, loudness, truepeak, surround, and stereo image. Courtesy RTW GmbH & Co.KG, Cologne.

30.11.4 Controls

In many implementations there are some user-control buttons, typically for resetting, engaging, and pausing the long-term measurement, the maximum momentary value (Mmax), the maximum short-term value (Smax) or the loudness range reading (LRA). The default desired target loudness value is −23 dB LUFS in Europe or −24dB LUKS in the USA.

Annex 1 of ITU BS.1770-3 specifies the objective multichannel loudness measurement algorithm. The algorithm consists of four stages:

1. “K” frequency weighting.
2. Mean square calculation for each channel.
3. Channel-weighted summation (surround channels have larger weights, and the LFE channel is excluded).
4. Gating of 400ms blocks (overlapping by 75%), where two thresholds are used:

- The first at −70 LKFS.
- The second at −10dB relative to the level measured after application of the first threshold.

The use of a relative gate guarantees ubiquitous audio source material use. It works with speech anchored material as well as with non-speech audio.

Bibliography

H. M. Tremaine, *Audio Cyclopedia*, Indianapolis, IN: Howard W. Sams & Co., Inc., 1969.

“VU Meter and Peak Program Meter—Peaceful Coexistence,” *J. Audio Eng. Soc.*, Preprint 1474 (G-5), Stephen F. Temmer, Gotham Audio Corporation.

“The Peak Program Meter and the VU Meter in Broadcasting,” *J. Audio Eng. Soc.*, Preprint

1691 (I-8), Hans Schmid, ABC.
“A Different Approach to the Old Problem of Audio Level Monitoring,” *J. Audio Eng. Soc.*, Preprint 2567 (B-1), R. W. Burden Associates, Canoga Park, CA, and M. Dorrough Electronics, Woodland Hills, CA, March 1988.
“The VU Meter,” *J. Audio Eng. Soc.*, Preprint 1794 (J-3), Michael Dorrough, Dorrough Electronics.
Communications with Mike Kahsnitz, RTW GmbH & Co.KG, Cologne.

Part 6

Recording and Playback

Chapter 31

Analog Disk Playback

by George Alexandrovich and Glen Ballou

31.1 Introduction

In the past 100+ years approximately 30 billion phonograph records have been produced and sold. Music of the most famous composers and performers, orchestras, and bands, and sounds of events have been immortalized in intricate excursions of the analog record groove. Millions and perhaps billions of discs are still in the hand of the audiophiles, archives, musical libraries, DJs and radio stations.

The contents of all of these records can never be completely rerecorded onto the compact discs or another medium, so it is important that we can preserve, restore, and reproduce analog recordings.

The information contained in this section is directed toward the new generation of engineers and technicians so they may understand the reproduction techniques that led to digital technology. As we witness the decline in popularity of analog LP discs, remember that many developing countries around the world are still very much dependent on analog technology and in some cases what we consider the old 78 rpm format is the only source of prerecorded music and entertainment available to them.

Early recorded sounds had a high-frequency cutoff of 2–3kHz. It took over 100 years to reach the sophistication of today's recording technology only to take a couple of steps backward in sound realism by approximating the waveforms at the high frequencies and limiting them to 20kHz with brick wall filters. Theoretically digital recording is fine, but the human ear deserves a higher sampling frequency. Perhaps only a select few can really hear the difference, but then how can we argue with them? In other fields, such as television, the trend is toward high-definition TV, in VCRs and camcorders there is a SVHS system, and yet tube-type audio amplifiers are still sold at premium prices because of many so-called golden ear audiophiles don't want to give up the tube sound. The same is with LP records. For the average listener, CDs are great as long as they don't hear pops and clicks and cannot break the stylus or the tonearm.

This chapter will discuss playback equipment. To understand the production of records/discs, refer to the *Handbook for Sound Engineers–The New Audio Cyclopedia*, First or Second Edition.

31.2 Disc/Record Dimensions

The analog record has been standardized to 7inch, 10inch, and 12inch discs and 33⅓ and 45 revolutions per minute (rpm).

Excerpts from the EIA standard for producing analog disc records are:

31.2.1 Record Diameter

The diameter of records are:

12 inch LP disc, 33⅓ rpm	11.875 ±0.031in (301.6 ±0.8mm)
10 inch disc, 33⅓ rpm	9.875 ±0.031in (250.8 ±0.8mm)
7 inch disc, 45rpm disc	6.875 ±0.031in (174.6 ±0.8mm)

The recorded surface shall start with at least one turn of unmodulated groove.

31.2.2 Maximum Outer Diameter

The maximum outer diameter of a recorded surface shall be:

12 inch LP disc, 33⅓ rpm	11.500 in (292.1mm)
10 inch disc, 33⅓ rpm	9.500in (241.3mm)
7 inch disc, 45rpm	6.625in (168.3mm)

31.2.3 Groove Dimensions

The groove dimensions shall be:

Minimum top width (monophonic only)	0.0022in (0.56mm)
Maximum bottom radius	0.00025in (0.006mm)
Included angle	90°±5°

On stereophonic records, the instantaneous groove width should be not less than 0.001in (0.025mm). The average groove width should preferably be not less than 0.0014in (0.035mm).

31.2.4 Stereophonic Groove

The *stereophonic groove* shall carry two channels of information. The two channels shall be recorded in such a manner that they can be reproduced by movement of a reproducing stylus tip in two directions at 90° to each other and at 45° to a radial line through the stylus tip and the center of the record. The reproducing stylus tip motion shall be tangential to, or lie in a plane through, the stylus tip and the record center, preferably inclined at an angle of 20±5°

clockwise to the normal to the record surface through the stylus tip, as viewed from the record center. In practice, angles of between 0° and 25° may be encountered.

31.2.5 Channel Orientation

The groove shall be recorded for reproduction with the right-hand loudspeaker(s), as viewed from the audience, actuated by movement of the groove wall, which is farther away from the center of the record.

31.2.6 Channel Phasing

The phasing of the two recorded signals shall be suitable for reproduction on equipment so connected that movement of the reproducing stylus tip parallel to the record surface (as with a monophonic record) produces in-phase signals across the output terminals of the phono cartridge.

31.2.7 Channel Levels

The levels of the two recorded signals should be such that peak excursions of the groove should not exceed 100µm or 0.004in in lateral plane and 50µm or 0.002in in vertical plane.

31.2.8 Speed of Rotation

Records shall be recorded for reproduction at one of the following speeds:

50Hz Electric Supplies	60Hz Electric Supplies
45.11rpm±0.5%	45.00rpm±0.5%

33⅓rpm±0.5%	33⅓rpm ±0.5%

(Note: 16¾rpm and 78rpm speeds omitted.)

31.2.9 Lead-In Groove Pitch

The lead-in groove pitch shall be 16 ±2 lines/inch (l/in).

31.2.10 Lead-Out Groove

The pitch of the lead-out groove shall be 2–6 lines/in (l/in). The top width of the lead-out groove shall increase to a minimum of 0.003in (0.076mm) when the pitch exceeds 0.25 in (6.4mm).

31.2.11 Finishing Groove

The diameter of the finishing groove shall be:

12inch and 10inch discs	4.187 ±0.31in (106.4 ±0.8mm)
7inch discs	3.875 ±0.078in (98.4 ±2mm)

31.3 Signal Equalization in Disc Recording

To overcome the limitations found in the basic disc-cutting and reproducing process, special equalization of the signals before and after the recording was developed. When all signals that appear in the program bus are analyzed, the amplitude is the highest at low frequencies and the lowest at high frequencies. The relationship between the frequency of the signal and its amplitude where amplitude is inversely proportional to frequency is called a *constant* *velocity characteristic*, Fig. 31-1.

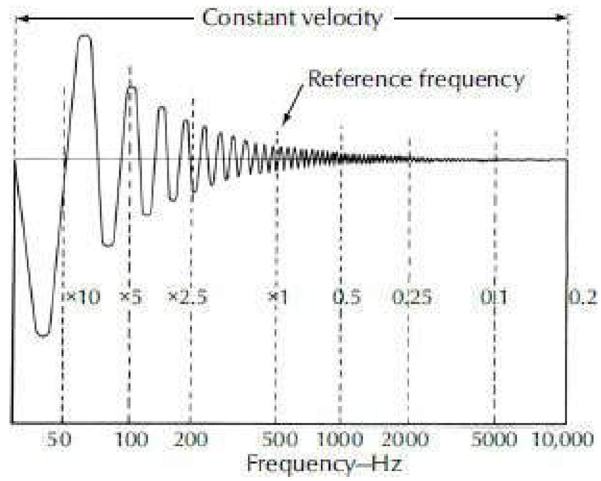


Figure 31-1. Constant velocity characteristics.

If the signals are recorded without equalization as they arrive, the low-frequency excursions would take all the space. The high frequencies would be of such a low amplitude that during the playback, high-frequency signals could be very close to the noise level of the system. The *SNR* then would be extremely small.

To eliminate this problem, equalization was introduced to the record and playback circuits. The equalization used for cutting was called *preemphasis*, and equalization used in playback equipment, *postemphasis*. There are three equalization curves; the RIAA, NAB, and DIN equalization curves. The first curve was used by the Record Industry Association of America (RIAA) and the second, which is almost identical to the first curve, by the National Association of Broadcasters (NAB). The DIN (Deutsche Industrie Norm) standard used in European countries calls for additional equalization at the extreme low end during playback to improve the *SNR* and stability of the system due to mechanical disturbances, i.e., turntable rumble, which can affect the overall performance of the system.

The NAB (RIAA) curve used presently in the playback equipment is shown in Fig. 31-2. The numerical values for the characteristic are shown in Table 31-1. For recording, the inverse curve is used. It means that if the playback signal is boosted +19.3dB, the same signal should be recorded at the level of −19.3dB so that the overall result will be 0dB deviation from the ideal flat response curve.

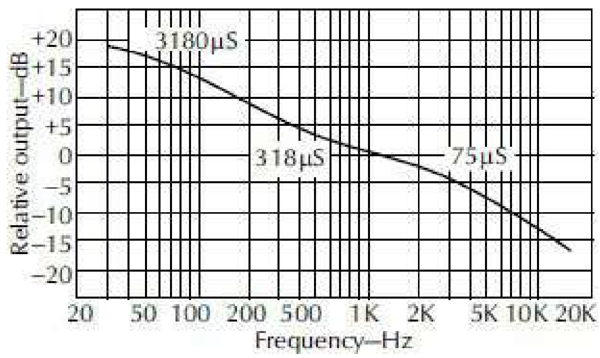


Figure 31-2. NAB (RIAA) standard reproducing characteristic.

Equalization is used to record the sound at the most advantageous levels for the best results as far as distortion and noise are concerned and to reproduce it so that the original balance between the frequencies can be restored.

The RIAA curve covers the range from 20Hz to 20kHz. The DIN curve, as shown in Fig. 31-3, extends the control over playback down to 2Hz where the equalization returns back to 0dB. As can be seen from the graphs, the curves have complex shapes; equalizer circuits use capacitors and resistors, and their values determine the amount of signal equalization that can be expressed as a function of a time constant in microseconds as derived from the equation

$$T = CR \tag{31-1}$$

where,
T is a time constant,
C is capacitance in F,
R is the total effective resistance of the supply network in Ω.

Table 31-1. Preferred Frequencies and Calculated Recording Characteristics

Frequency (Hz)	Recording Characteristics (dB)	Frequency (Hz)	Recording Characteristics (dB)
20.0	−19.3	800.0	−0.8
25.0	−19.0	1000.0	0.0
31.5	−18.5	1250.0	+0.7
40.0	−17.8	1600.0	+1.6
50.0	−16.9	2000.0	+2.6
63.0	−15.8	2500.0	+3.7
80.0	−14.5	3150.0	+5.0
100.0	−13.1	4000.0	+6.6
125.0	−11.6	5000.0	+8.2
160.0	−9.8	6300.0	+10.0
200.0	−8.2	8000.0	+11.9
250.0	−6.7	10,000.0	+13.7
315.0	−5.2	12,500.0	+15.6
400.0	−3.8	16,000.0	+17.7
500.0	−2.6	20,000.0	+19.6
630.0	−1.6		

This is part of the equation to determine the attenuation at various frequencies:

$$attenuation_{dB} = 10\log(1 + \omega^2 T^2) \qquad (31\text{-}2)$$

where,

ω is 2πf

T is *CR* of Eq. 31-1.

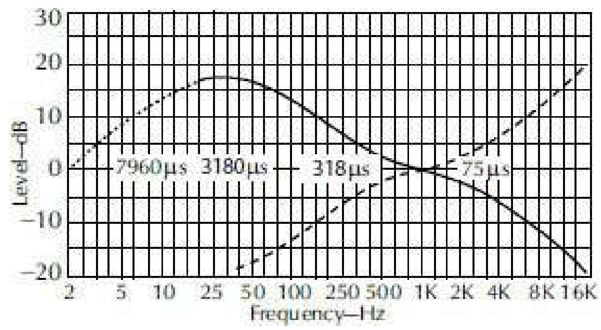


Figure 31-3. DIN recording and playback characteristics.

The RIAA curve consists of three time constants; 75μs to roll off the high frequencies, 318μs to produce the slope below 1kHz with a knee at 500Hz, and a 3180μs time constant to flatten the low end of the curve.

Because the recording space on the record disc is limited, records are cut with constant amplitude characteristics of the signals in the upper half of the frequency range. When reproduced by the pickup, these signals are equalized to a constant velocity characteristic. In playing back these preemphasized disc recordings, different equalization has to be used for different types of cartridges. For instance, dynamic cartridges, which include moving-magnet, moving-iron, moving-coil, and variable-reluctance pickups, are constant velocity devices; therefore, they respond to the speed of the stylus movement. The faster the stylus is deflected, the higher the output voltage. Ceramic or crystal cartridges are pressure-sensitive devices, and they respond to the force applied to the stylus. They are called *constant amplitude devices*, and when records with constant velocity recording are played with ceramic cartridges, no additional equalization is required. The combined characteristics of both the recording and the cartridge complement each other, returning the signals to their original form. Only a minimal amount of signal grooming may be necessary to compensate for the effects of capacitive loading and nonlinearity of the cartridge.

31.4 Turntables

To play a record, the turntable or device to rotate the disc at the required speed is needed. The construction and execution of the requirement may differ greatly between the models and the designs of different manufacturers. The history of evolution of the record drive mechanisms takes us from the days of hand-cranked cylinder machines, through the age of spring-wound phonographs with mechanical governors for speed control, and into the age of electrically driven machines with electronic control. Today the accuracy of turntable speed is measured in small fractions of 1% in deviation from the desired speed.

31.4.1 Drive Systems

Turntables are driven by electric motors. The method by which the power from the motor is transferred to the turntable platter classifies the drive mechanism. The turntable platters can be belt driven, puck or idler driven, or driven direct.

The first category, the *belt-driven type*, encompasses all models that have motors mounted to the side of the platter with the belt stretched over the motor pulley and outer rim of the platter, Fig. 31-4A. Some platter designs have an additional internal rim to hide and to protect the belt.

Many turntables have synchronous motors or motors with some type of speed control mechanism, such as a centrifugal switch that disconnects the power to the motor when the speed exceeds the preset value. The later types of motors are usually low-voltage, battery-driven motors used in portable equipment. Also, in portable turntables there is electrical feedback to control the speed of the low-voltage motor.

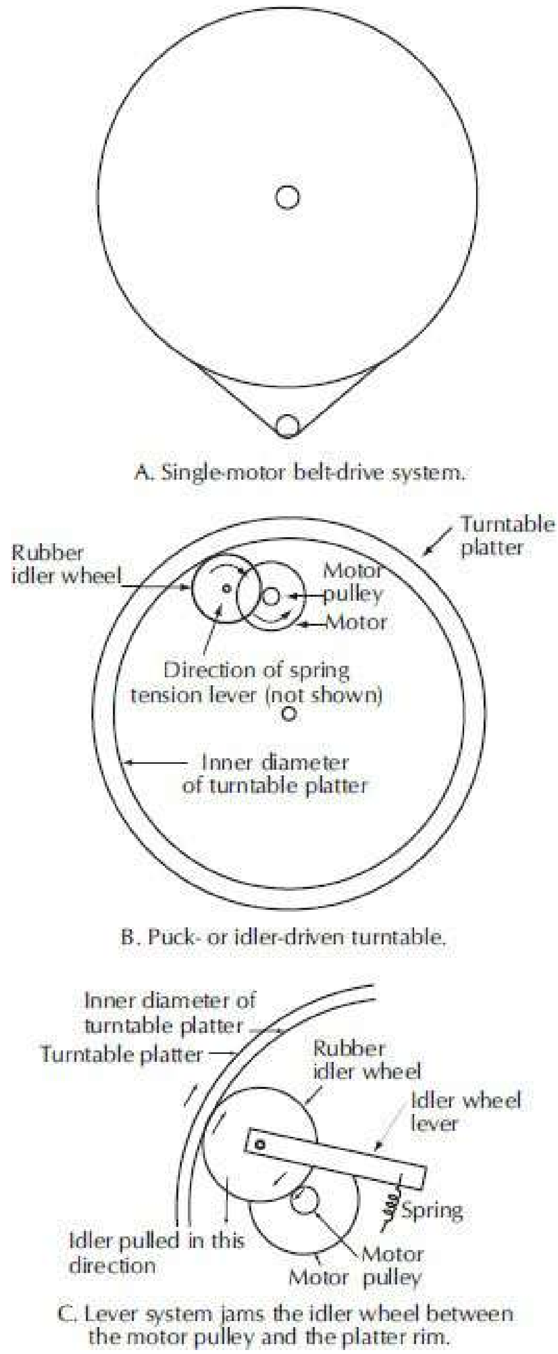


Figure 31-4. Various types of drive mechanisms.

Another version of the same idea uses a low-voltage ac motor driven by a self-contained crystal-controlled oscillator allowing variation of the speed of the platter and achievement of great speed precision. The only source of speed variation can come from belt slippage or a defective belt. Belt-driven turntables are normally the quietest turntables. The speed selection of the belt-driven turntable can be accomplished either by changing the speed of the motor or by having the stepped pulley on the motor and by shifting the belt from one pulley onto another.

The second type of turntable is a *puck-driven* or *idler-driven turntable*, Fig. 31-4B. The coupling between the platter and the motor shaft is achieved through the intermediate idler wheel or puck, which has the outer edge covered with neoprene rubber or polyure-thane for positive drive and to isolate the motor vibration from the platter. The idler wheel rotates on the shaft that is attached to a sliding bracket. When one side of the idler pulley (or puck) is in contact with the inner side of the rim of the platter and on the other side with the motor shaft, the idler wheel will transmit the motor rotation to the turntable platter. The mechanism is designed so when the motor is turned off the idler wheel retracts away from the motor shaft to protect the rubber ridge from forming a flat spot.

The advantage of the rim drive is that it provides positive torque to the platter, and if the motor is strong enough, it can bring the turntable to the desired speed almost instantly. The mechanism is simple, and it is the most reliable type of drive. Unfortunately it is also the noisiest because of the positive coupling between the motor and the platter idler or puck that transmits a certain amount of the motor vibrations to the platter and consequently to the record, as shown in Fig. 31-4C.

The third kind of turntable drive is the *direct drive* where the motor drives the shaft of the platter directly. There are also variations of the design. Some turntable designs are very sophisticated, using the platter itself as a rotor of the motor and drive is provided by the self-contained, quartz-controlled oscillator. The motion is extremely accurate and the speed of rotation may be displayed on the digital display, which is part of the control panel. There is also a weak point in this seemingly perfect drive. Because of the slow speed at which the turntable rotates, and because the motor has a finite number of poles, there is a slight cogging action in the platter motion, which may manifest itself with increased loads. This handicap is only related to turntable platters with fairly small mass and small moments of inertia. If the platter is heavy, it will overcome this problem.

The ideal turntable should have the following properties:

- It will start fast without hesitation.
- It will rotate with exact speed without variations.
- There will be no motor noises or vibrations heard while the system is in operation, they will not be transmitted to the platter.
- The turntable should be adequately shock mounted and isolated from the surface on which it sits to prevent the transmission of rumble and vibrations from the room. These loud sounds can actually shake the platter and the tonearm.
- The platter should be treated against ringing either by using a turntable mat with damping properties or by undercoating the platter.
- The turntable must be easy to maintain and to repair.

Not many turntables meet all these criteria; therefore, in order to know how to evaluate the unit, it is important to know how they work.

Speed of Rotation. Before evaluating the entire system, there are tests that can be performed on the turntable alone. The first one is speed of rotation. There are many ways of checking the speed of rotation, but the simplest one is by using the stroboscopic disc.

Starting Time. *Starting time* is the time it takes for the platter to reach its operating speed from a complete stop. This time period is important to know for professionals who have to begin playing the song or selection at the exact moment. To check the starting time requires either a stop watch or timing device and a strobe disc or the test record. As soon as the lines on the strobe disc appear stationary, the turntable has reached its operating speed. In playing the record test tone, the pitch changes as the correct speed is attained. Starting time may vary anywhere from a fraction of a second to two or more seconds, depending on the construction of the turntable. Turntables used by disc jockeys have to start as fast as possible without overshoot, which means that the speed should not, even for a moment, exceed the desired speed. If this overshoot occurs as the program material is already being transmitted, the variations of the speed will be most objectionable.

Acoustical Noise. The third test concerns the *acoustical noise* the motor and the turntable are producing. Normally, this test can be easily performed in a quiet listening room when everything is turned off and only the turntable is energized. If the turntable noise is clearly heard and it overshadows the normal room noise, turntable drive is below an acceptable performance level. A second part of the same test is conducted when the turntable is turned off and the system is adjusted to a normal listening level. When the record with the quiet groove is placed on the turntable, a slight hiss can be heard when putting your ear to the loudspeakers. When the record with the quiet groove is placed on the turntable and the stylus is placed into the groove, listening to the increase in noise will show the extent to which the turntable transmits the building rumble. If the power to the turntable is turned on, the noise contributed by the motor drive can be measured. During this test, slightly tapping the base of the turntable can determine if the shock mounting is adequate and whether or not loud music will add coloration to the signal being reproduced. In summary, what is required from the good turntable is that it reproduces only what is recorded on the disc and is insensitive to all other sources of vibration.

31.4.2 Turntable Design in the 21st Century

One of the most important features of turntable design is the ability to keep noise and rumble created by motors and bearings from being picked up by the cartridge stylus. Many inexpensive turntables have a direct drive between the motor and the platter and inexpensive bearings, allowing motor noise and vibration to be transmitted to the platter and then to the cartridge stylus. Remember, it doesn't make any difference to the signal whether it comes from the stylus moving versus the disc or the disc moving versus the stylus.

VPI turntables, Fig. 31-5 use inverted bearings instead of conventional bearings. In this design the bearing assembly is in the platter rather than in a bearing well below the platter. The spindle and ball are attached to the chassis and the bearing well is inverted and placed in the platter itself. With this design the drive belt pulls through the center of the bearing assembly rather than many inches away from the center of the assembly, reducing teeter-totter effects to near zero for better stability.

All motor assemblies are completely separated from the turntable platter and tonearm, so there is no mechanical connection between the motor and the chassis except through the belt. This gives much lower noise levels due to isolation from the source of noise.

The VPI HR-X turntable uses a dual motor flywheel assembly to drive the platter. Two synchronous motors, driven by a perfect sine-wave ac power supply, drive a 14lb flywheel spinning at 300rpm, which in turn drives the platter. In this configuration the platter is driven by a non-electromotive source as opposed to other tables that are driven by the motor or combination of motors. Running the platter with no motor or multiple motors produces a velvety black background and perfect speed stability.

Figure 31-5. VPI HR-X high-quality noiseless turntable. Courtesy VPI Industries, Inc.

31.5 Tonearms

Tonearms can be classified into two categories: *pivoted* and *tangential tracking*, Fig. 31-6A and 31-6B.

Contemporary tonearms are designed to cope with a variety of problems. However, rarely can one find a tonearm with nearly perfect geometry and correct design to establish correct performance. Most tonearms have built in antiskating devices, adjustable counterweights to accommodate a variety of cartridge weights and tracking forces, vertical height adjustment to set the tonearm parallel to the record, and a variety of features to facilitate installation and operation of the device. All tonearms are at best a compromise. Very few tonearms are dynamically balanced, and most rely on dynamic unbalance to produce vertical tracking force. The dynamically balanced tonearm is the tonearm that is capable of playing a record with the turntable tipped at almost any angle without changing the tracking force and tracking ability.

Tonearm Geometry. Tonearms are designed to retrace the modulation of the groove in the same way as it was recorded. Design of the tonearm takes into consideration the diameter of the records or the turntable, and the distance between the center of the platter and the pivot point of the tonearm. Older tonearms suffered from a tangent error because the cartridge was aligned properly at only one point on the record. Today's pivoted tonearms have a built-in offset angle at which the cartridge is positioned so it is always perpendicular within a couple of degrees to the radius of the disc. This reduces distortion in the lateral plane and improves tracking. There are many protractors available today using different approaches to help position the cartridge as accurately as possible in the tonearm to minimize tracking error.

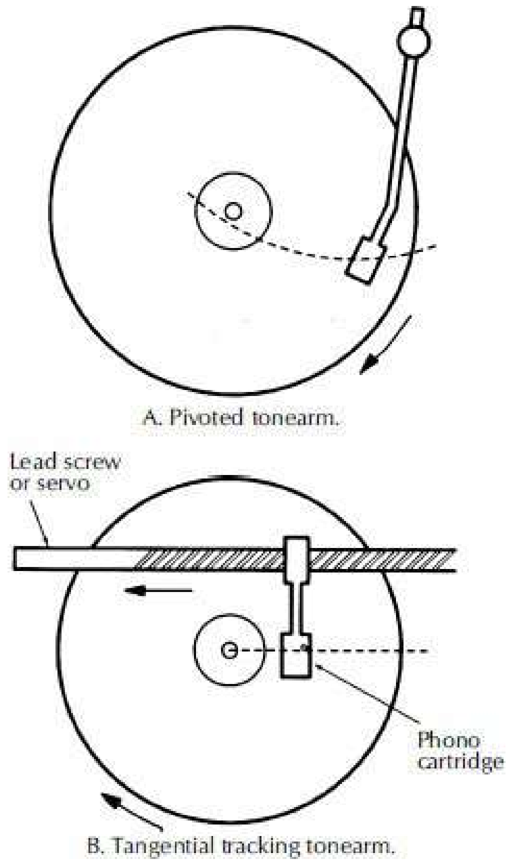


Figure 31-6. Tonearm classification.

When a disc is being cut, the cutting head is carried across the face of the recording disc following the radius. However, when in playback, the pickup is at the right angle to the radius of the disc only at two points, because the pickup arm is pivoted in such a manner that it swings across the face of the disc in an arc, as shown in Fig. 31-7.

Generally, the manufacturer of the arm supplies a template and mounting instructions for a particular arm. In the absence of such information, the pickup arm is mounted in such a manner that the tangent error is at a minimum. One method of mounting the arm is shown in Fig. 31-8. Regardless of where the pivoted arm is placed, a tangent error cannot be eliminated entirely.

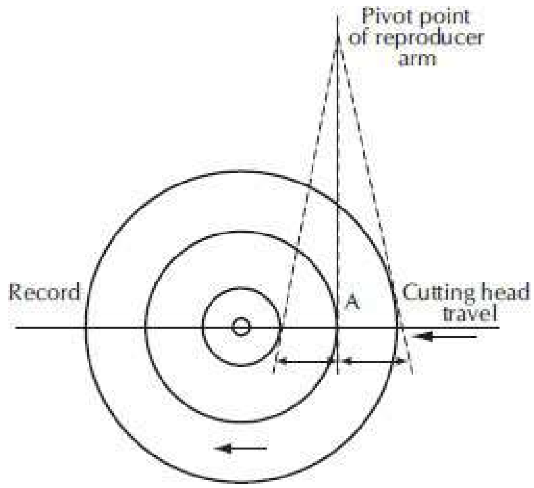


Figure 31-7. Tangent error in a reproducing arm. The error is zero at point A only.

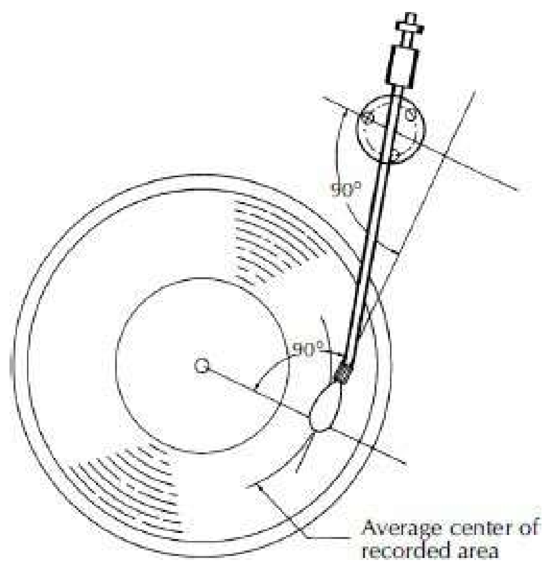


Figure 31-8. Typical mounting for an offset pickup arm.

The error can be made so small, however, that it can be neglected. In offsetting the tonearm by bending it into an S or J shape, Fig. 31-9, it is possible to position the cartridge so that at two points on the record the error shall be zero. The deviation from this ideal groove-cartridge interface will be only 2–3° in the horizontal plane.

Offsetting the tonearm introduces the skating force that pulls the tonearm toward the center of the record. In tonearms without the offset angle the skating force is zero at one point and increases as the tonearm moves away from this position. The zero tangent error point in this tonearm coincides with the zero skating force position, point A in Fig. 31-7.

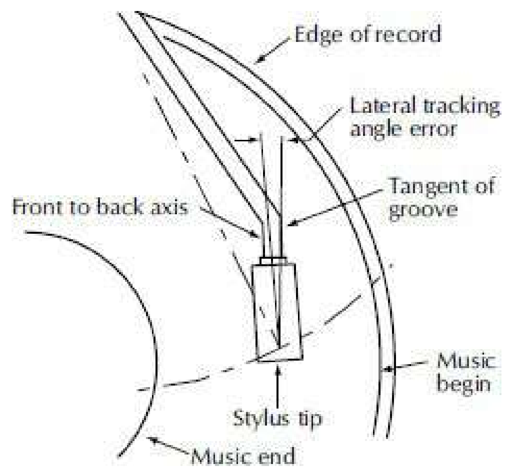


Figure 31-9. Geometry of a modern tonearm.

Theoretically, the pivoted tonearm without the offset angle and without any tangent error has to be infinitely long. The tonearm designed by the Rabinoff brothers revived the principle of tangential tracking used by Edison and found application in many turntable systems. In this system the tonearm motion has been achieved using servomechanisms and utilizing various types of arm position sensors. These tangential tracking turntables practically eliminated the tracking error and are quite popular with many hi-fi enthusiasts. There are also drawbacks to this design. Usually, such tonearms cannot be moved as fast as pivoted counterparts, and this may become a handicap in operations when speed of positioning the tonearm is of essence. The advantage of tangential tonearms is that they are shorter, lighter, and can be made more rigid to prevent many tonearm resonances found in some pivoted tonearms. But the mechanical complexity of tangential tracking tonearms requires the use of modern technology including special integrated circuits and sensors.

Effective Tonearm Length. Fig. 31-10 defines the turntable platter and spindle location in relationship to the effective tonearm length, which is the distance between the stylus tip and the tonearm pivot.

Modern tonearms have a built-in stop preventing them from moving farther than the locking groove so only three dimensions are of importance: effective tonearm length, vertical pivot-to-spindle distance, and the offset angle.

The accuracy of the cartridge tracking and mounting depends on the effective length of the tonearm. If the effective length of the tonearm is 7.87 in and it is properly mounted (7.04 in away from the turntable spindle), the cartridge will track to within $+2\frac{1}{4}°$ and $-1\frac{1}{2}°$ providing the cartridge is mounted at an offset angle of 27.8°. If the tonearm is longer, the lateral tracking error gets smaller so that the tonearm with the effective length of 10in will have a maximum tracking error of less than 1° at the smaller disc radius and a 1.7° error at the maximum radius.

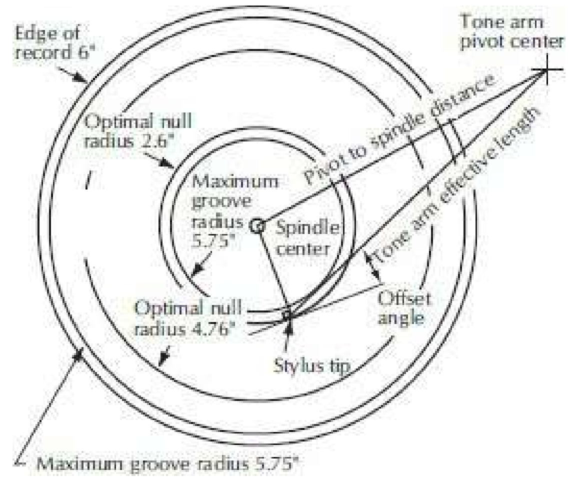
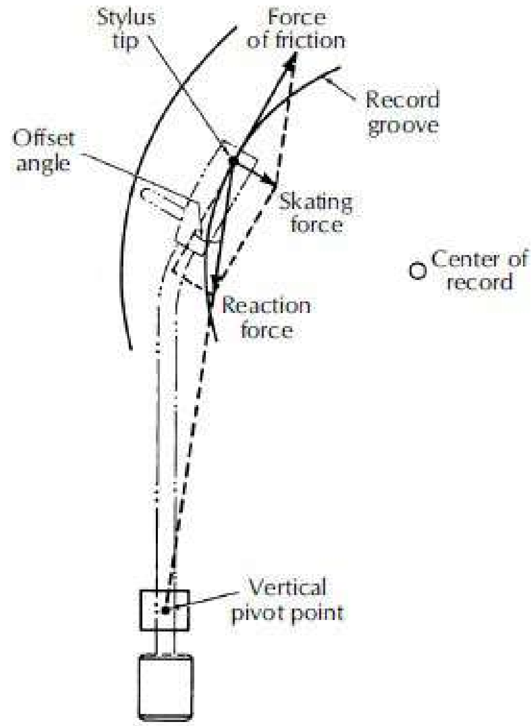


Figure 31-10. Relationship of the lateral components of a tonearm.

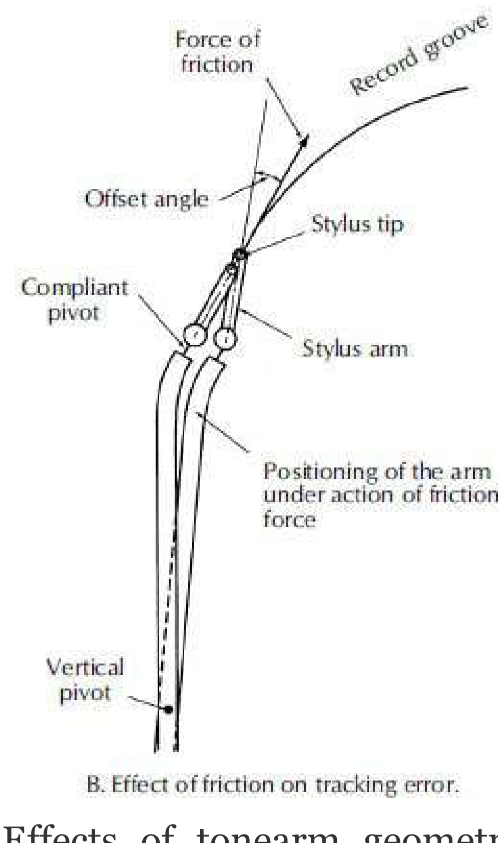
Since the linear speed of the outer grooves is higher and the wavelengths are longer, tracking angle errors have lesser effect on the signal quality so tracking errors should be minimized at the inner grooves for consistent quality of playback signal at all radii.

Skating Force. *Skating force* is a force that can upset the best aligned tonearm and cause considerable tracking error. The skating force is the result of tonearm geometry and the friction between the stylus and the record groove. Because of the offset angle and the overhang, one vector of this force pulls the stylus in a direction away from the pivot point of the tonearm and the second vector pulls the tonearm toward the center of the turntable, Fig. 31-11A. If this skating force is not compensated for, the stylus will be deflected toward the outside of the disc at the angle much greater than the error angle encountered in tracking the groove at different radii, Fig. 31-11B.

The skating force compensation consists of applying a force to the tonearm that is equal to but opposite in direction to the skating force, Fig. 31-12. For all practical purposes, the skating force is constant for all radii of the music groove if the tracking error is small and the tonearm alignment is correct. There are slight variations of the skating force due to heavy modulation and groove wall plastic deformation caused by the sharpness of the stylus, but the largest deviation in skating force is due to the variations in record material. From the study of various materials, it was established that the softest materials produce more friction and larger skating force. Lacquer masters produce up to 25% more friction (i.e., skating force) than vinyl records. 45rpm discs, have approximately 30% less friction than vinyl, requiring less antiskating compensation than vinyl LPs.



A. Position of the arm on the record and the forces acting on it.



B. Effect of friction on tracking error.

Figure 31-11. Effects of tonearm geometry. Courtesy George

Alexandrovich.

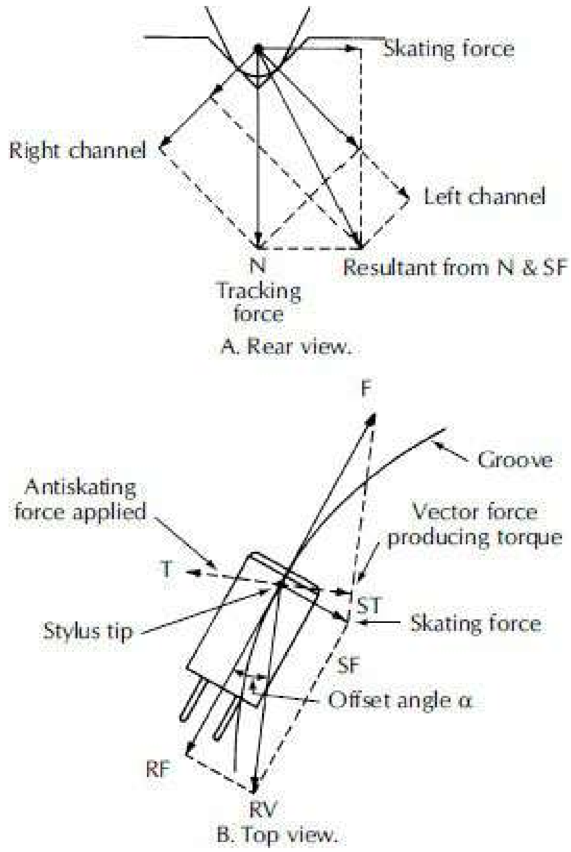


Figure 31-12. Skating and antiforces in a record groove. Courtesy George Alexandrovich.

There are many different ways to generate the antiskating force. It is incorrect to assume that increasing the drag on the horizontal motion of the tonearm will compensate for skating. Skating force is independent of groove spiraling speed; drag is not. Also, because of the variable pitch common to all present-day recordings, the speed with which the tonearm moves across the record varies and at times may even be zero. Because of this variation, the mechanism that generates the antiskating force should be able to generate a uniform force at all times, regardless of the motion of the tonearm. Antiskating force can be generated by using springs, magnets, weights with pulleys, electrical devices, and mechanical linkages and weights, Fig. 31-13. Any method to apply the clockwise bias in a horizontal plane to the tonearm to counteract the skating force produces positive results; however, compensation may not be accurate for all types of systems.

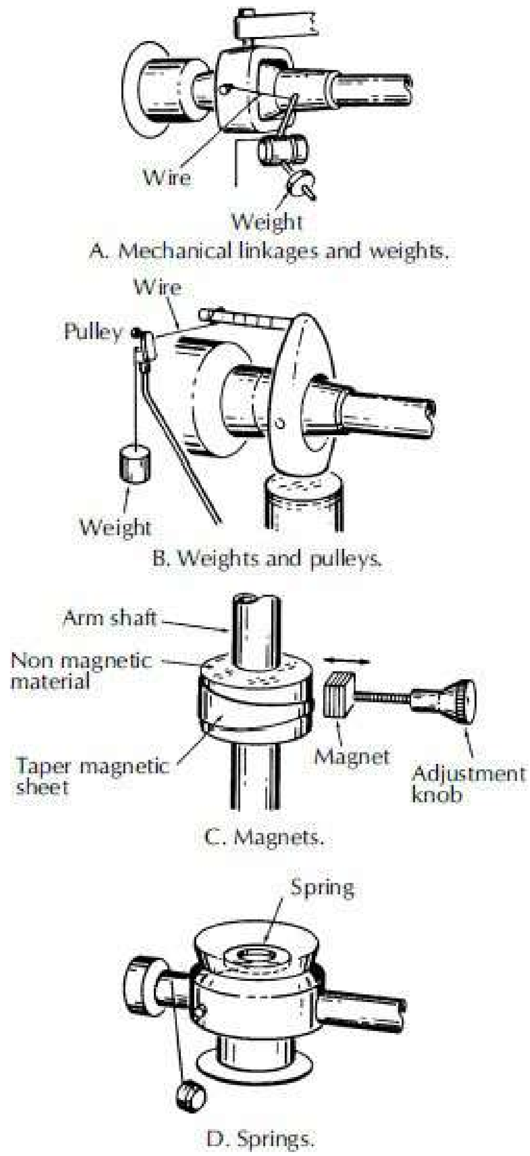


Figure 31-13. Different methods of generating antiskating force.

The effectiveness of the antiskating force mechanism depends to a high degree on the dynamic behavior of the tonearm. If the tonearm is not dynamically balanced (and most of them are not), any tilt of the turntable may result in a change of skating force, endangering the tracking ability of the pickup. Dynamic balancing of the tonearm implies that the pivot point of the tonearm is also the center of mass. In most modern tonearms this center of weight is shifted toward the cartridge end in order to produce tracking force, Fig. 31-14. In a dynamically balanced tonearm, tracking force is produced by using either a spring or a permanent or electromagnet (solenoid). A properly dynamically balanced tonearm could play a record with the turntable being in any position and is completely insensitive to jarring of the turntable or floor vibrations.

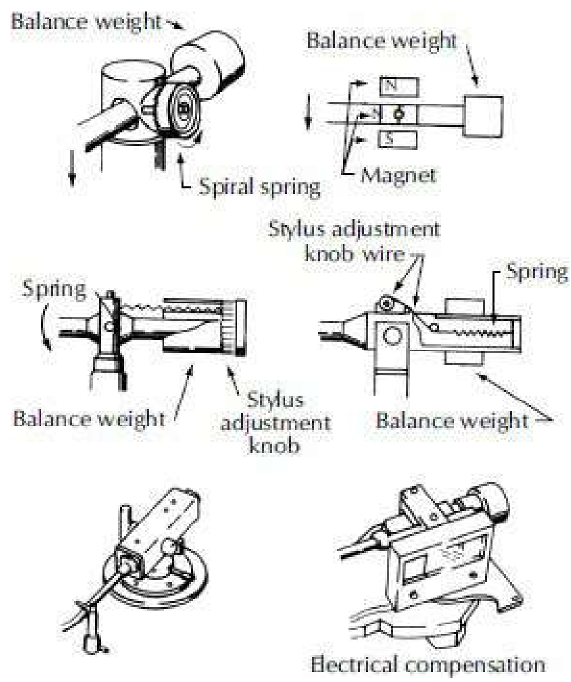


Figure 31-14. Examples of how dynamic balance of the tonearm can be achieved.

Vertical Tracking Angle. An important adjustment of the tonearm is in positioning the cartridge over the surface of the disc.

Cartridges are mounted in tonearms so that the mounting surface of the cartridge is parallel to the record surface, Fig. 31-15A. Sometimes tilting the cartridge fore or aft results in lower tracking distortion. Some cartridges are designed to produce the lowest distortion when playing vertical modulation that was recorded at the vertical cutting angle of 25°, Fig. 31-15B. At the same time most of today's records are cut with the vertical angle of 10°–15°. So in order to reduce the distortion during playback, matching the two angles by moving or tilting the cartridge backward a few degrees may help reduce tracking distortion.

Tonearm Resonance Damping. A Shure Brothers, Inc. study revealed that the warp frequencies of LP records lie in the region from one revolution (0.5Hz) peaking at 3Hz and tapering down at 7–8Hz. Because the audible range of frequencies starts at around 20Hz, tonearm resonance placed between the warp frequency region and the audible region allows minimum distortion of the signal due to tonearm bounce. Improvements have been made in the tonearms by applying vertical damping to the tonearm. The vertical tonearm motion control was attacked by Discwasher, Inc., with a special damping mechanism named Disctracker, which attached to the cartridge. Shure Brothers introduced a stabilizer brush that attached to the cartridge similar to the brushes invented and used by Pickering and Stanton since 1971, except that the Shure stabilizer brush had its pivots filled with damping fluid. These devices helped to stabilize the tonearm as the brush cleaned the record groove.

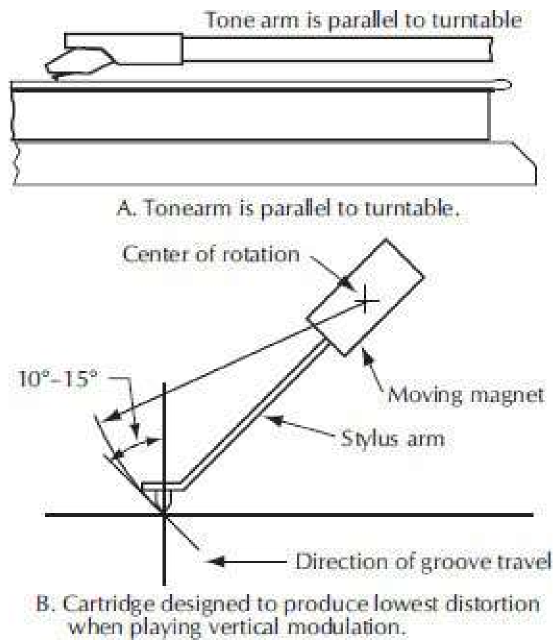


Figure 31-15. Schematic representation of the moving system of a pickup, illustrating the vertical tracking angle.

The other approach was to adjust the effective mass of the tonearm by pivoting only the front part of the tonearm and selecting a cartridge with compliance that would match the mass of this portion of the arm, Fig. 31-16. Dynavector tonearm is an example of such design. Another variation is the design by Sony that employs electronic control of the tonearm motion. Instead of relying on weights, springs, or magnets, the Sony tonearm uses linear dc electromotors driven, operated, and controlled by electrical signals. Unfortunately, not all functions of the tonearm are controlled automatically and are subject to misadjustment.

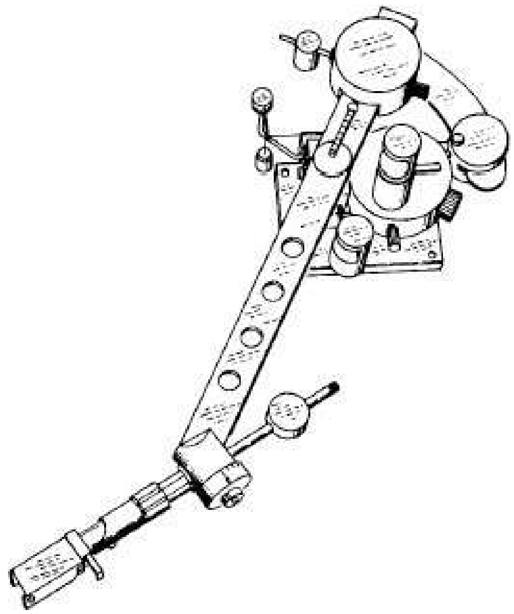


Figure 31-16. Dynavector tonearm with pivoted front portion for lower dynamic tonearm mass. Courtesy Onlife Research, Inc.

31.6 Phono Pickup/Transducers and Styli

In order to reproduce signals recorded on the phonograph record, a *transducer* (phonograph pickup, phono cartridge, or needle) converts the groove modulation into the electrical signals. Unlike microphones, loudspeakers, and other types of devices or transducers that convert one form of energy into another, the phonograph pickup has to perform more than one function. The phonograph pickup or cartridge, so called since the invention of the removable stylus or needle assembly, has to convert modulation of the record groove into the electrical signals, and at the same time support the tone-arm at the proper height above the record surface, all the while moving the tonearm across the surface of the record.

The phonograph cartridge is an electromechanical device designed to track or follow the excursions of the record groove and to convert this motion, with the help of a tracking mechanism-stylus assembly, into electrical signals.

Cartridges are classified by the principle by which they convert mechanical motion into the electric current or signal, *electrodynamic* and *piezoelectric*. There are also pickups designed to operate using strain gauges, variable capacity, and light as sensors.

Electrodynamic-Type Cartridges. Electrodynamic-type cartridges are subdivided into three categories: *moving magnet, moving coil*, and *induced magnet* or *moving-iron* type. The electrodynamic principle consists of using a magnetic field that, when it intersects the coil windings, generates electric current. The construction of the cartridge classifies the type. If the magnet is attached to the stylus tube or cantilever and the coils are stationary, it is called a moving-magnet cartridge. If the magnet is made stationary and the coils move in the magnetic field, it is a moving-coil cartridge; and if the magnet and the coils are made stationary and there is a slug of soft magnetic iron moving in place of a magnet while being magnetized by the stationary magnet, it is called a moving-iron or induced-magnet cartridge.

Variable-Reluctance Cartridges. Since the introduction of the original variable-reluctance pickup, Fig. 31-17, many different versions of its design have appeared. The magnetic structure consists of two pole pieces A, with a small permanent magnet B between them. The coil C is mounted with a soft rubber insert D.

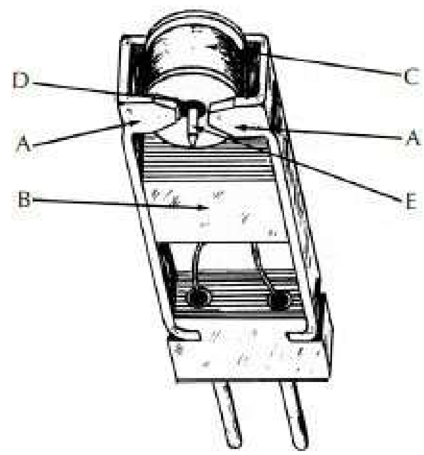


Figure 31-17. Variable-reluctance magnetic pickup.

The stylus, which is also the armature, is held in the exact center of the magnetic structure by the rubber insert. When the stylus is actuated, its movement causes a voltage to be generated in the coil. Because of its construction, the frequency response extends beyond the normal audio-frequency band. Output voltage is on the order of 100mV at 1kHz, with an output impedance of 500Ω. The recommended stylus pressure is 15–20g. The stylus weighs 31mg and is removable. Although the recommended pressure is 15–20g, the pressure could be as low as 7g. The frequency response is ±2dB, 20Hz–20kHz.

Moving-Coil Cartridges. Modern moving-coil cartridges are represented by a variety of designs. All of them have coils that move, but not all of them are entitled to be called moving-coil type. Designs that depend for their functioning on the motion of the soft iron core rather than on the motion of the coil itself should not be classified as a pure moving-coil device. There the motion of the coil is coincidental. Fig. 31-18 shows the cross sections of moving-coil stylus assemblies as they move during the playing of the record. The magnetic flux is directed by the iron core or armature of the coil. If the coil is made stationary and the core is vibrated, the signal will still be generated. This fact prevents it from being classified as a pure moving-coil device.

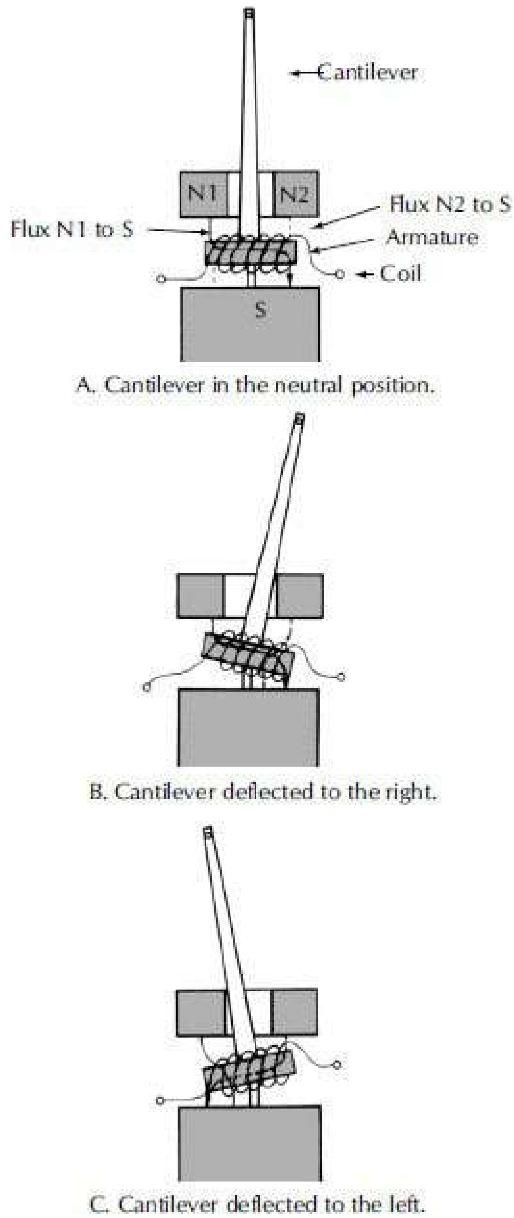


Figure 31-18. Bottom view of the moving-coil cartridge generator assembly.

The advantage of this design is extremely low output impedance, making the cartridge insensitive to capacitive loading and allowing the use of very long cables without altering the frequency response of the device. On the negative side, the output of the cartridge is very low, measuring in the tenths of a millivolt requiring an extra 20–30dB amplification to bring the electrical signal to the required level as referenced to an established sensitivity of 1mV/1cm/s of recorded velocity. A step-up transformer or an extra stage of amplification usually introduces additional noise and the effect on capacitive loading. Other drawbacks of such design are the weight of the cartridge and the need to use a heavier tracking force.

One of the debatable points about MC cartridges is the sound they produce. MC cartridges have a very fast response to transients because of the very low inductance and the impedance of the coils and the very rigid cantilever, which has to be strong in order to move a relatively heavy coil assembly. Another factor in this type of design is the construction of the coil assembly, which may have a number of turns in the coil unsupported and free to vibrate, producing random signals at higher frequencies. Also lead-in and lead-out wires may not be secured properly and can vibrate in the magnetic field producing random coloration of the signal. Lead dressing, coil impregnation, and gluing techniques control the purity of the sound produced by this design.

Step-up transformers require winding ratios of 1:10 or more. The transformer’s high-impedance secondary winding is reflected back into the primary, and any loading of the secondary in excess of the specified value affects the signal output level and electrical damping of the coils. Theoretically, shorted coils produce maximum damping, while an unterminated winding of the transformer’s secondary will emphasize electrical resonances and unchecked mechanical motion. It is important to locate the step-up transformer near the preamplifier input to minimize the capacitive load of the shielded wires between the transformer and the input stage of the amplifier. Because the levels handled by this input transformer are extremely low, good transformer shielding is necessary.

In lieu of the step-up transformer, a prepreamplifier may be used. Additional preamplification, obtained from active gain circuits, requires super low-noise circuits in order to preserve an acceptable *SNR*. There have been many such pre-preamplifiers designed using the most exotic devices and circuits, operating with batteries or special ac power supplies with maximum filtering and voltage regulation and using magnetic shielding.

Moving-Magnet Cartridges. The most popular highperformance stereo cartridges are the moving-magnet type. Moving-magnet cartridges offer one of the most sensible ways to design the stereo cartridge with a replaceable stylus. This cartridge has low dynamic tip mass, high compliance, and fairly high output. By using the most powerful rare earth magnets and using the most modern manufacturing methods, the frequency response is extended from almost direct current to well past the threshold of hearing, Fig. 31-19.

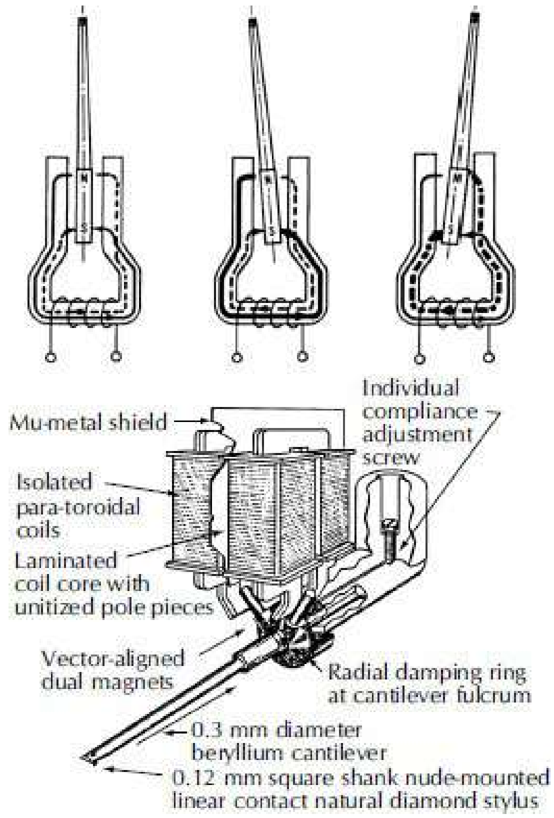


Figure 31-19. Basic principle of the moving-magnet pickup.

Induced-Magnet Cartridges. Fig. 31-20A is an example of an induced-magnet or variable-reluctance pickup manufactured by Bang and Olufsen of Denmark. It consists of a small armature in the form of a cross, made of Mumetal, which swings between four pole pins. A stylus bar constructed of aluminum tubing 0.002 in (0.05mm) thick is attached to the Mumetal armature cross at one end. The stylus is secured to the other end of the tube. Four pole pins with four coils are placed at each end of the cross. With a 45° motion to the right, a reverse voltage induction takes place. Such action permits the coils to be connected push-pull, thus reducing harmonic distortion induced by the nonlinearity of the magnetic field. In addition, the coils provide an effective hum-bucking circuit.

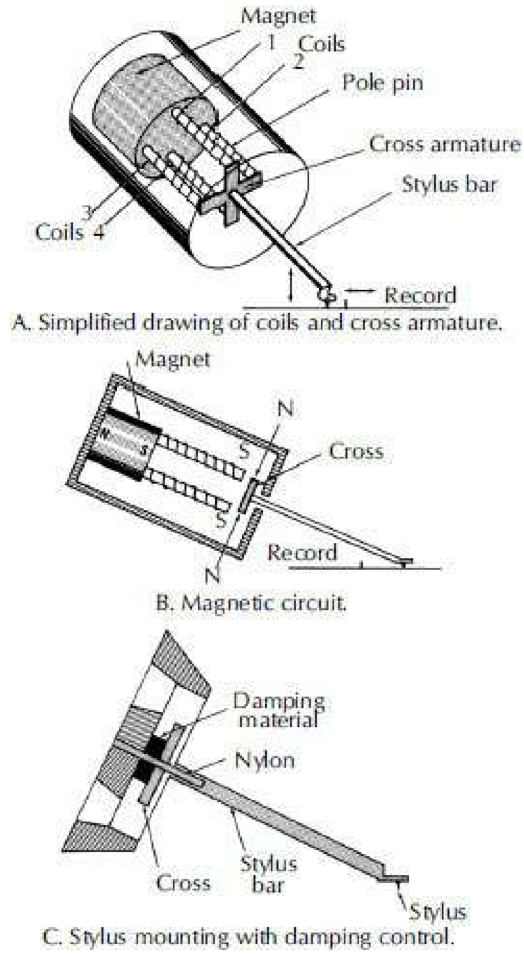


Figure 31-20. Induced-magnet cartridge construction.

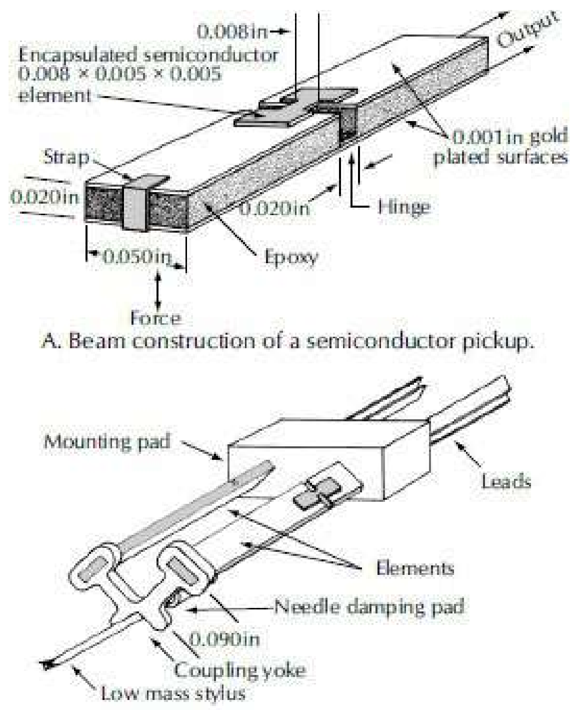
Crosstalk between the left and right channels is minimized, since such components are bucked out. Modulating one channel 45°, the cross arms on the orthogonal channel rotate without changing the spacing; so there is no induced voltage in this channel, assuming the positioning of the unit, with respect to the groove, is correct.

A cross-sectional view of the magnetic circuit is shown in Fig. 31-20B and is similar to the magnetic structure of a loudspeaker employing a center magnet. Thus, a closed magnetic circuit, which prevents leakage of the magnetic field, is provided and being nonmagnetic, it cannot be attracted to the steel turntable plate. It also provides an effective shield for the coils. The stylus bar pivots on a nylon thread, bonded to a plastic support. The armature cross bears on a resilient disc, Fig. 31-20C, which controls compliance and supplies damping for the moving system. The rotational point of the system is at the junction of the armature cross and the nylon thread support. The output voltage is 7mV for each channel for a 5cm/s cut. The stylus has an angle of 15° at 2g of tracking force and may be operated at a pressure of 1–3g. Compliance is 15×10^{-6}cm/dyn for both directions of motion. Frequency response is 20Hz–20kHz ±2.5dB.

Semiconductor Pickup Cartridge. A *semiconductor pickup cartridge* operates on the principle of the strain gauge. The pickup mechanism employs two small, highly doped silicon semiconductor elements 0.008in × 0.005in whose resistance varies as a function of the stylus deflection, Fig. 31-21. The elements are mounted on laminated beams of lightweight epoxy with gold-plated surfaces. A notch in the beam under the assembly acts as a hinge for stress concentration. In this structure, two beams are used, each driven by an elastic yoke, coupled to the stylus. Aside from the compliance of the yoke and mounting pads, a mechanical advantage of over 40:1 can be attained in the beam and stylus lever. This mechanical transformer provides high compliance and reduces the mass of the elements reflected to the stylus. The stylus is elliptical in shape and set at an angle of 15°.

Since the semiconductor elements are sensitive modulating devices and not generators as in the conventional pickup, very little energy is required for their operation. The compliance at 1kHz is approximately 25×10^{-6}cm/dyn and the frequency response is from 20Hz–50kHz. A power supply, two single-stage preamplifiers, and one inverter stage are required. As the elements are deflected by the stylus action, the resistance of the semiconductors, about 800Ω, changes slightly, causing a varying dc voltage across the output. This dc signal is ac coupled to the preamplifiers in the power supply, providing an output voltage of 0.4V for each side. The cartridge employs mechanical equalization that, in combination with the RC equalizer at the output of each preamplifier, results in an RIAA reproducing characteristic.

Piezoelectricity. Piezoelectricity is pressure electricity. The voltage generated by the crystals in piezoelectric cartridges is proportional to the amplitude of the stylus displacement. The output voltage of the average piezoelectric pickup is considerably higher than for other type pickups. Piezoelectric pickups are treated electrically as a capacitive-reactance device since the impedance rises with a decrease of frequency. Simple RC networks are used with this type of pickup to obtain a frequency response corresponding to the RIAA standard. Records recorded using a constant-amplitude characteristic may be reproduced without equalization.



A. Beam construction of a semiconductor pickup.

B. Cartridge construction of a semiconductor stereophonic pickup.

Figure 31-21. Stereophonic semiconductor pickup.

In the ceramic stereophonic pickup, Fig. 31-22, the moving system consists of two piezoelectric crystal slabs of lead-zirconium titanate or similar material. This particular material offers good mechanical and electrical properties with high sensitivity and high capacitance. The ends of the slabs are mounted rigidly in a mounting block, and the front end is connected by a yoke made of injected molded plastic. This coupling is critical because the electrical performance and the mechanical impedance seen at the stylus point by the record groove depends on it. The coupling system is defined as that portion of the mechanism that lies between the stylus tip and the ceramic slabs.

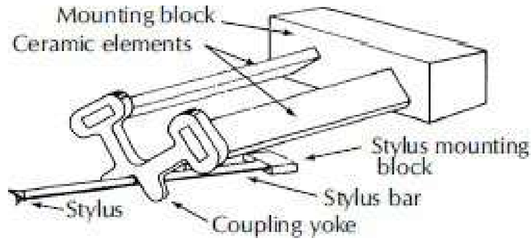


Figure 31-22. Simplified drawing showing the construction of a ceramic stereophonic pickup.

The stylus bar is made from heat-treated, thin-walled aluminum alloy tubing, with one end flattened to hold the stylus at the desired angle. The other end of the stylus bar is held in place by the stylus mounting block. The coupling yoke is connected at a point about midway on the stylus bar. This point is chosen because it affords the most desirable electrical performance and substantially reduces the mechanical impedance of the yoke and ceramic elements as seen by the stylus tip.

Better designs have four output terminals, two for each channel to ensure the complete isolation of one side from the other. Damping in the form of a viscous material is used to control the frequency characteristics. These pickups are of the constant-amplitude type with the output voltage 10mV for a peak velocity of 5cm/s. Ceramic pickups are not affected by magnetic or electrostatic fields.

RC equalizer networks for both crystal and ceramic pickups are shown in Fig. 31-23. The networks are connected between the output of the piezoelectric pickup and the input of the preamplifier. The characteristics of these networks are such that they correspond to the standard RIAA curve. A pickup with a compliance of 15×10^{-6}cm/dyn or greater can have a response ±2dB.

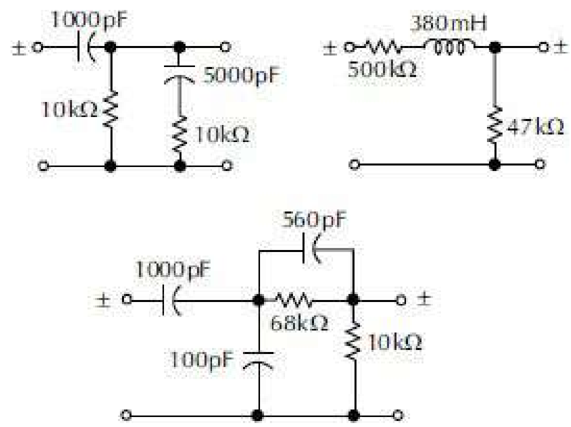


Figure 31-23. Networks for equalizing ceramic cartridges.

The internal impedance of the average crystal pickup is approximately 100kΩ, with a capacitance of 0.001 to 0.0015µF.

31.6.1 Cartridge Styli

Stereo Disc Groove. The playback stylus is the first link between the information stored in the record groove and the playback system. The quality of the reproduced sound is influenced by the precision with which the stylus follows the groove modulation.

In stereophonic recordings with 45°/45° modulation, the two channels are isolated from each other because modulation of each channel is at 90° to the other, Fig. 31-24.

To minimize the effects of vertical excursions at low frequencies, the phase of both channels is adjusted so low-frequency signals are in phase in order to produce lateral modulation. The phase relationship of the two channels determines the location of the sound image between the two loudspeakers, and in some cases the phase is a deciding factor as to whether there is going to be a signal reproduced at all.

Stylus Tip. The function of the playback stylus of the cartridge is to follow all deflections of the groove. Since the stylus is attached to the end of the cantilever, any motion of the stylus tip is transmitted to the other end of the tube or shank, where the electrical signals are generated by a moving magnet, a moving coil, or a crystal. The stylus has rounded off edges that are polished for smooth tracking. Ideally, the playback stylus should be centered in the groove, and its centerline should match that of the cutting stylus. There are always minute imperfections in the alignment of the stylus and of the groove. Therefore the shape of the playback stylus is made to compensate and allow some misalignment of the stylus in the record groove.

The stylus touches the groove walls at two points. The contact area is curved and is a part of the tip radius so that if the stylus is slightly tilted due to misalignment of the cartridge or the tonearm, tracking will not be affected.

Spherical Stylus. There are several types of styli today. The simplest and the oldest one is the *spherical tip*. The spherical stylus is a tiny diamond or sapphire cylinder with one end ground to a cone shape with its tip polished to an accurate sphere. The included angle of the cone is about 55°, and the tip radius is about 0.0007in or 0.7 mil. Because grooves can be as narrow as 0.001in, the stylus tip has to be equal to or smaller than the groove in order to track it. The standard tip radius dimensions for today's spherical styli range from 0.0005–0.0007in (12.7–17.7µm).

Elliptical Stylus. The second type is the *elliptical stylus*. From the front it looks like a spherical stylus; however, there are two flats polished in the front and the back of the stylus. The side radius of the elliptical tip is much slimmer than that of the spherical stylus.

The intersections of the two flats are polished to form small radii called the *tracing radii*, which measure about 0.0002in (5μm). These small side radii are actually in contact with the modulation of the groove and, because they are small, they follow the high-frequency excursions of the groove more easily.

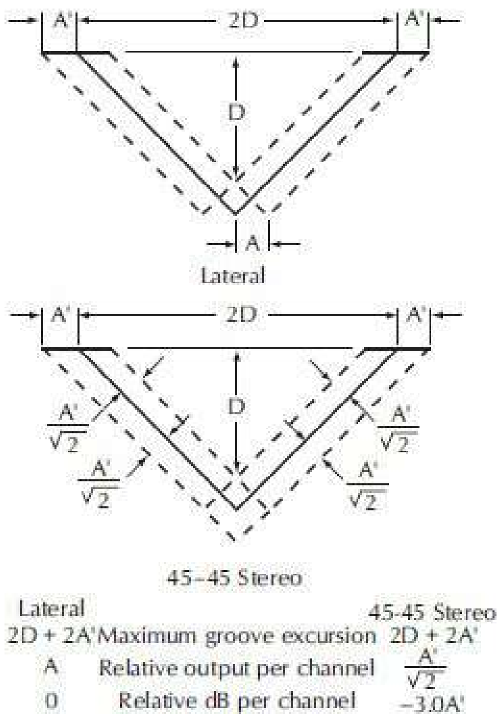


Figure 31-24. Comparison of 45°/45° stereophonic groove with standard lateral groove.

Stylus Characteristics. All playback styli are designed to contact only the walls of the groove; therefore, the stylus tip has to ride without touching the bottom of the groove. Since the diamond gets slimmer as it wears down, the tip gets closer and closer to the bottom of the groove. When it starts touching it, noise increases because debris has accumulated on the bottom of the groove and is scooped up by the stylus. This is a clue to change the stylus in order to reduce the noise and to preserve the record from being destroyed by the sharp edges of the worn diamond.

Currently, almost all styli manufactured are made out of diamond. The quality and the price of the stylus depends on whether it is made out of a solid piece of diamond or a small chip bonded onto another material that acts as an extension or pedestal for the diamond tip. The technology of manufacturing diamonds has advanced significantly so that chip bonding and encasing can be favorably compared to solid or nude diamond tips. In view of the fact that the area of contact is only 0.2 millionths of a square inch ($0.2 \times 10^{-6} in^2$) and as long as this area is made out of a diamond, the overall performance of the stylus will not be affected. All this is true providing that the mass of the bonded stylus assembly is not higher than that of a conventional diamond and not larger than the nude stone.

The vertical tracking force applied to the stylus is divided between the two walls. Each wall is experiencing force that is equal to the total vertical force times the cosine of 45° or 0.707, Fig. 31-25. For instance, if the vertical tracking force (VTF) is 1g, each groove wall will experience a force of 0.7g.

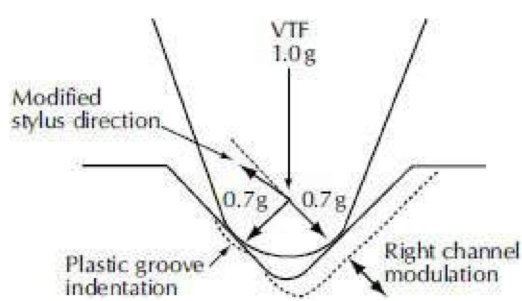


Figure 31-25. Stylus motion and forces acting upon it in a stereo groove.

A very small area of contact exists between the stylus tip and the groove so the pressure against the groove wall can rise up to many thousands of pounds per square inch. For instance, if each wall receives 0.7g of force applied through the contact area equal to two ten millionths of an inch (0.2×10^{-6}), the pressure is 7726lb/in. With such high pressures and force of friction between the stylus and the vinyl, the outer skin layer of the record material melts as the tip slides over the plastic and then refreezes almost as fast as it melted. Since the melting temperature of the vinyl is about 480°F, the same temperature exists in the contact area.

Stylus Cantilever. The stylus is attached to some type of coupler or cantilever that connects it to the generating element of the cartridge, which could be a magnet, a piece of iron, a coil, or a ceramic element. Because of a very wide range of frequencies this stylus assembly has to transmit, the construction material and shape of the cantilever are very important and has to be very light and rigid. Over the century of existence of mechanical sound recording, styli were made out of cactus needles, whale bones, and all kinds of metal, gems and stones, plastic, and wood. The final choice is centered around an aluminum alloy thin-wall tube. It is fairly strong, light, noncorrosive, nonmagnetic, electrically conductive, and easy to manufacture.

The average diameter of the aluminum cantilever tube is 0.03in (0.76mm), and the length may vary from ¼–½ in (6–12mm). A few exotic cartridges have cantilevers made out of solid ruby or even diamond and some from boron or beryllium copper alloy. Although ruby and diamond are extremely rigid materials, because of manufacturing difficulties and high weight/length ratio, they are made very short. This, in turn, brings the pivot point much closer to the stylus tip that moves in a much smaller arc when reproducing groove modulation. Since the grooves are modulated by the cutting stylus that has its pivot quite a distance away and is moving in an arc of much larger radius, the larger the difference between the motions of the cutting and of the playback styli, the larger the distortion.

On the other hand very long playback cantilevers are unable to produce sufficient motion of the generating element that results in a very low electrical output.

Compliance. The amount of force required to move the playback stylus depends on several factors; the first is the compliance of the stylus, and the second is mass.

Compliance of the cantilever or the stylus is the ability of the stylus assembly to react to the groove modulation. It is measured in cm/dyn or μm/mN (metric) and gives the amount of stylus tip deflection for the given force. Compliance is measured statically and dynamically.

Static compliance is the amount of deflection of the cantilever when a constant force is applied to the stylus tip. Dynamic compliance is a measure of tip deflection as it is reproducing the frequency of known amplitude at which the measurement is being made.

Vertical Resonance. The second variable in the equation is the tonearm/cartridge vertical resonance. Tonearms and cartridges resonate between 5Hz and 15Hz; the most desirable range is between 8Hz and 12Hz. Resonance below 8Hz will produce instability of the tonearm and will result in poor tracking of moderately warped records.

Stereo cartridges have fairly uniform compliance in all planes of stylus motion. Cartridges with higher compliance work best with light tonearms, and heavy tonearms should be set up with cartridges having low compliance. If the stylus compliance is low, the tracking force applied to the stylus should be higher than for a high-compliance stylus.

31.6.2 Cartridge Voltage Output

The output voltage of the cartridge depends on its design and the type of generator system used. Ceramic or crystal cartridges produce the highest voltage. Next are the moving-magnet cartridges and then the induced-magnet pickups; the last group is the moving-coil cartridges. The moving coil cartridge produces higher power output than other types so they can work with step-up transformers to increase the output voltage 10–20 times or 20–26dB. On the other hand, some high output voltage ceramic cartridges are connected to the loss pads and response-shaping networks to reduce the voltage down to the average output level of the moving-magnet cartridges. Today most of the preamplifiers are designed to accept moving-magnet cartridges.

31.6.3 Electrical Loading

With various output levels and different source impedances, cartridges respond differently to electrical loads. For instance, crystal or ceramic cartridges are the most susceptible to capacitive loading. The entire frequency response is dependent on the loading of the cartridge. In the moving-magnet cartridge, only the highest portion of the frequency range is affected by the capacitive loading. The moving-coil cartridge is almost completely immune to the loading effects. Once it is connected to the step-up transformers, the secondary of the transformer becomes very sensitive to loading, and excess capacity can play havoc with transformer resonance and the impedance of the secondary transformer winding. Therefore, cartridge manufacturers specify the recommended resistive and capacitive loads.

The most common resistive load is 47kΩ (50kΩ for Europe), paralleled by 200–400pF of capacitance for the moving-magnet cartridges, depending on the manufacturer and on the cartridge model. The capacitive loading for the cartridge includes capacitance of all interconnecting cables and tonearm wiring to ground (or between the conductors), and capacitance added by the connectors and switches. Finally the internal wiring of the preamplifier circuit and preamplifier input circuit capacitance, which varies widely depending on the circuit design, adds capacitive loading to the cartridge, Fig. 31-26. In many cases the total capacitance that appeared as a capacitive load for the cartridge exceeded 1000pF, which resulted in an electrical resonance peak around 7–8kHz followed by premature response roll-off at frequencies above this point.

31.7 Phono Preamplifiers

Phonograph cartridges require a special type of amplification to reproduce the recorded sound the way it existed during the recording session. The electrical signals from the cartridge, measuring only a few millivolts rms have to be amplified into signals of many volts. This has to be accomplished with:

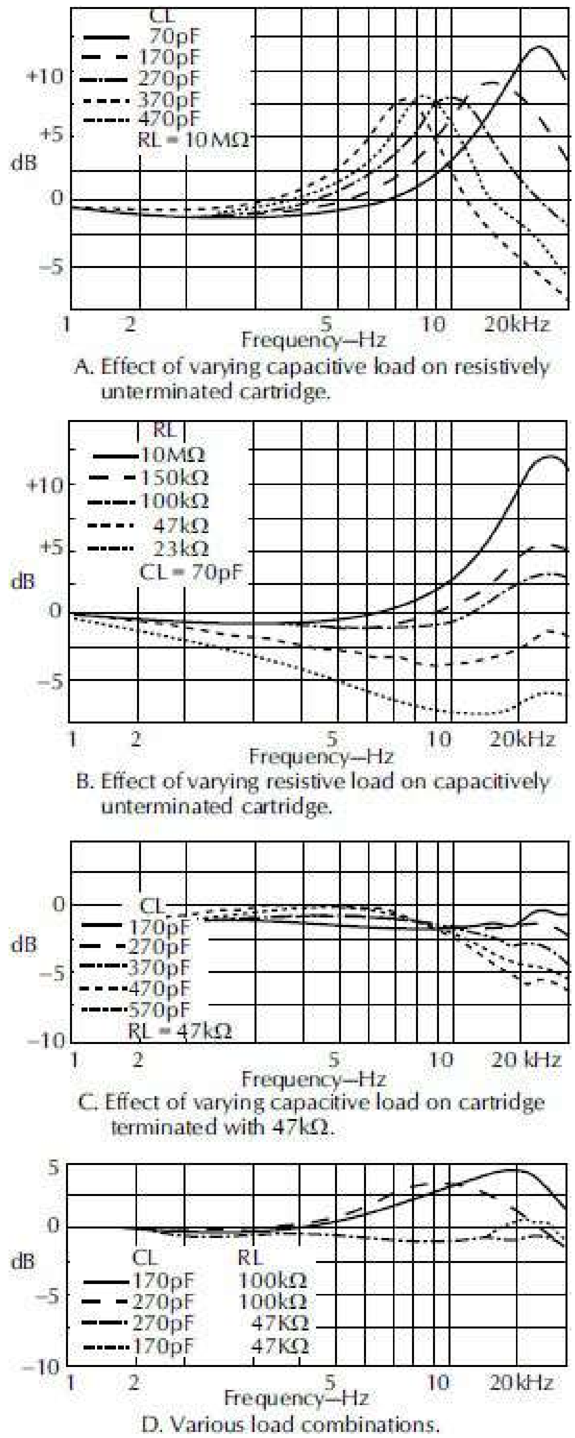


Figure 31-26. Effect of cartridge loading on frequency response. Courtesy Stanton Magnetics, Inc.

- Minimum distortion.
- Flat frequency response.

- Excellent *SNR*.

The phono preamplifier has to amplify a cartridge signal without:

- Changing its phase.
- Adding more than a small percentage of harmonic and intermodulation distortion.
- Adding to the noise content of the original signal from the cartridge.
- Needing enough reserve power to handle any unusually high transient signals.

The average required voltage amplification is 40–50 dB and is dependent on the output of the cartridge. The dynamic cartridge produces 4–5mV of output for the average recording signal. A preamplifier gain of 45dB will boost the signal output to nearly 1V, the level required to drive most power amplifiers to full output. The noise contribution of the cartridge and of the recording medium requires the preamplifier noise level to be at least 70dB below the average input signal of 10mV.

The frequency response of the circuit should follow the RIAA characteristics, with the low frequencies boosted about 20dB, and the high frequencies attenuated by the same amount, with respect to 1kHz, which implies that the preamplifier with 40dB of gain at 1kHz will have as much as 60dB of gain at 20Hz and only 20dB of gain at 20kHz.

It is not unusual for a cartridge producing an output of several millivolts for the average modulation to produce 100mV voltage peaks. Cartridges are designed to produce an output voltage of around 1mV for each centimeter per second of recorded velocity or for the average recorded level of 5cm/s, the cartridge output is 5mV. Some preamplifier circuits when overloaded by fast spikes can recover in a matter of microseconds and resume their normal operation while others are incapable of recovering fast and once overloaded stay in this unbalanced state long enough to produce audible distortion of lower-level signals that may follow. Direct coupled stages, which don't employ large capacitors and inductors, have much higher slew rates and consequently react much faster and with less distortion to audio signals.

The average moving-coil cartridge produces from 0.1–0.6mV output with the source impedance of a few ohms and an inductance of a few millihenries, so 20-30dB of additional voltage gain is required from the pre-preamplifier. Because the output level of the cartridge is so low, an extra demand for low-noise performance is placed on the circuit. To maintain the same *SNR* as in high-output moving-magnet cartridges, the pre-preamplifier (or head-amplifier) circuit should have 20dB lower noise than the preamplifier for the moving-magnet cartridges. One of the ways to achieve this lower noise is by using a step-up transformer. The power supply for the low-level amplifiers requires excellent regulation and extremely low ripple voltage.

The preamplifier input for the moving-magnet cartridges requires a 47kΩ input resistance and a low, preferably adjustable, capacitive load. The proper termination of the moving-magnet cartridge is very important for the correct performance of the transducer. Moving-magnet cartridges have a resistive and inductive nature, so designers specify the capacitive load. If the specified capacitive load is higher than the total capacity of the circuit, the preamplifier should have a provision to add capacitance to the cartridge termination as required. If the total capacitance is larger than needed, cables can be made shorter or replaced with ones having lower capacitance.

31.8 Laser Turntable System

The Laser Turntable (LT), manufactured by the ELP Corporation, Japan, Fig. 31-27, features a contact-free optical pickup system that allows records to be played thousands of times without damage to the record. The Laser Turntable operates as follows: Two tracking laser beams are directed to the left and to the right shoulders of the groove of the record. Only the part of the beams that reach the groove are reflected to two PSD (Position Sensitive Detector) optical semiconductors. The part of the beams that fall on the land area of the record are deflected and not picked up by the PSD devices. The signals are sent to a microprocessor via analog to digital converters, then to servos to maintain the reader head position directly above the groove.

Two additional laser beams are directed at the left groove wall and the right groove wall just below the tracking beams. Modulation on the individual grooves is reflected to scanner mirrors and onto left and right photo optical sensors. The variations of the modulated light cause the audio sensors to develop an electrical representation of the mechanical modulation of the grooves. The entire sound reproduction chain is analog.

The distance from the surface of the record to the traveling pickup head is kept constant by using a separate laser beam. This is very similar to every CD player that uses a “focus” laser to move the laser that reads digital bits to the proper spacing between the reader and the disc. Since phono records vary in thickness, this feature assures precision alignment from the pickup head to the record. The servos are fast and responsive allowing the LT to accommodate even warped records. The new audiophile 180 gram (thick) records are also reproduced.

Figure 31-27. ELP Laser Turntable. Courtesy ELP Corporation.

The same audio information on records is engraved from the shoulder to the bottom of a record groove. The laser reads audio information that is 10 microns below the shoulder, Fig. 31-28, therefore, the laser picks up audio information which has not been touched or damaged by a pickup. It plays the virgin audio information on the groove without digitization.

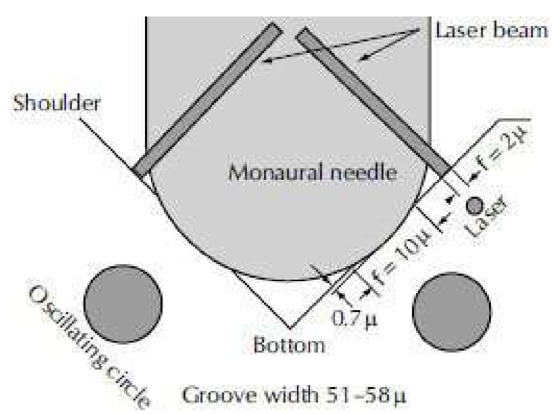


Figure 31-28. The laser beam picks up the signal closer to the

shoulder where the standard needle does not touch.

The incident area of the laser beam on the groove is one-fourth the contact area of the best stereo needle and 26 times smaller than a mono needle, Fig. 31-29. The laser beam travels to the wall of the groove and back. The reflection angle is transferred to the audio signal. Therefore, the LT maintains analog sound through the entire process, without any digitization. As a result, the LT cannot differentiate between an audio signal or dirt on the record, so the vinyl record must be absolutely clean and free of debris.

The laser beams must reflect from an opaque surface in order to be read. Clear or colored records are transparent, or translucent, and will not reflect light to the sensors. Other types of records that may have difficulty include:

- Vertical cut records like the early Edison "Diamond Cut" series. The modulation is up and down rather than lateral.
- Rounded groove shoulder.
- A groove with a rounded bottom produces distortion.
- No Acoustic Feedback or Sound Alteration.

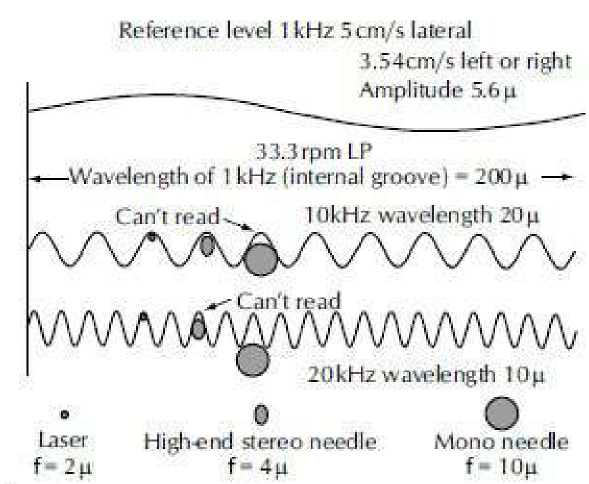


Figure 31-29. The incident area of the laser beam is 1/4 the size of

a stereo needle and 26 times smaller than a monaural needle.

Feedback is typically caused by sound from the loudspeakers (or from elsewhere) reaching the turntable and mechanically picking up the vibrations, to be amplified again. There is no needle singing. The LP is safely in a drawer and the laser reads only the undulations of the groove, therefore there is no need for elaborate vibration isolation pads. The LT will not hear outside noises such as footsteps on the floor, door slamming, or other vibrations in the area.

Operating the Laser Turntable. When a record is placed in the drawer and the play button pressed, the record tray closes and the LT scans the disc to identify the various bands or cuts. The bands are displayed on the front panel Record Profile LCD display. A single vertical line above the bands indicator shows the position of the laser pickup head. The vertical indicator travels across the record as it is played, showing its exact position on the record, and which band is playing.

On the initial scan the laser head moves from the inside (spindle) to the outside track while marking the bands. The machine then moves into the first band and measures the distance from the head to the record surface. After a few seconds the record will begin playing from the beginning.

The LP can repeat the same record up to five times, repeat a cut, listen to a segment of the cut, play a single groove segment repeatedly, or play selected bands in any order.

When a record starts to play, the message window displays the rpm of the platter. When a record is playing, the display shows the elapsed running time, the elapsed time of the current cut, the remaining time of the side, and total time of the side.

31.9 Record Care Suggestions

One of the most effective ways to keep the sound from the record free of noise and unwanted pops and clicks is to keep the groove and the stylus clean. The causes for dirty records are obvious; accumulation of airborne dust, finger grease, cigarette smoke, and anything that can be attracted by the static charges that exist on the surface of the vinyl disc. The dirt around the playback stylus is mainly due to raking the groove. Dust particles, as they settle down on the record surface, are attracted by the stylus, especially if it has a static charge on it. Better cartridges have their styli electrically grounded to bleed any static potential from the cantilever assembly to ground.

31.9.1 Brushes

One method to keep room dust out of the record groove is to have the cartridge work with the dust-collecting brush. In sliding over the surface of the vinyl record, the electrically insulated brush produces a static charge of its own that attracts and holds the dust particles from the surrounding area. The stylus cantilever, which is metallic and electrically neutral because of grounding, stays clean and free to vibrate and track the modulation of the groove.

31.9.2 Record-Cleaning Machines

The groove modulations in vinyl LPs are so small, on the order of the wavelength of light, that any compound, be it liquid or solid, will cause distortion in the reproduction of those grooves. The diamond stylus can be equated to a rock, and the vinyl record to Jell-O. Picture a rock running through Jell-O at a high velocity.

Anything that changes the way this rock moves through the Jell-O will cause changes in the recorded sound.

In the groove is a conglomerate of fungus, mold, dirt, ash, pollution, mold release compounds, various cleaning fluids and preservatives, etc. All these substances affect the way the stylus reads the groove and will affect the sound. A good vacuum cleaning machine will allow you to scrub the record with cleaning solution and then vacuum the record surface clean of the fluid carrying the contaminates away with it. A record cleaned on a good vacuum cleaning machine is microscopically clean and will sound it.

One of the great shocks in audio is the first time you hear a record you know very well cleaned by a vacuum cleaning machine. The sound is cleaner, clearer, crisper, with the sound of the hall or acoustic space very easy to hear. A clean record will not wear out. It is not the stylus that ruins the records' it's the stylus going through grunge and pressing the grunge into the vinyl grove that kills the sound of records.

Vacuum record cleaning machines all work the same way; a record is placed on a turntable, the turntable turns the record while the machine or the operator scrubs the record, the vacuum nozzle then sucks the contaminated fluid off the disc. VPI's HW-16.5 (in production for almost 30 years), Fig. 31-30, is an inexpensive record cleaner.

Figure 31-30. VPI HW-16.5 basic record cleaner machine. Courtesy VPI Industries.

It is strongly advised that before using any cleaning device the instructions be followed precisely and some experimentation be done on a few records before the entire library is cleaned or covered with a preservative coating. A word of caution: if too much record preservative is used, it will do more harm than good. Not only does the excess of material not lower the surface noise, but it contaminates the stylus tip to the extent that it is no longer able to stay in the groove. Accumulation of the cleaning or antistatic substance on the stylus tip also increases its dynamic tip mass, interfering with tracking of high-frequency modulation. Consequently, cleaning the cartridge stylus becomes as important if not more important than cleaning records.

Records should not be washed unless necessary. Dry clean them first with a soft brush or lint free velvet cloth. If the record must be washed, use distilled water; never use hot water or water containing dissolved minerals. Record labels should be protected by placing a piece of thin plastic over the labels. Use a soft camel hair brush or piece of moistened velvet with a couple of drops of liquid detergent or shampoo applied to clean the grooves in a circular motion. Rinse thoroughly with distilled water, and then wipe with a clean lint free cloth. The record can be blow dried with a hand dryer set to the cool position (never hot). Most of the dirt in the groove is dust attracted by the static charges that exist on the record surface. Washing or rinsing the record surface dissipates these electrical charges, allowing the dust to float away.

Warning: Old 78 rpm records should never be washed with solutions containing alcohol or other chemicals that dissolve shellac, the major binding ingredient in the record material. Vinyl LP records are much more forgiving and can be cleaned with alcohol solvents. The safest and most effective cleaning solvents are simple household liquid soaps that can do the job well if certain precautions are followed.

Turntable mats are the greatest contributors of dust contamination because turntables are left to stand open for prolonged periods of time, accumulating dust on the mat. When clean records are placed on the mat, the underside of the disc picks up most of the dust off the mat. It is important to clean, even wash the mat.

Vinyl records (and CDs) are sensitive to heat. When the record is pressed under very high pressure, vinyl is flattened into a thin plastic disc that is forced to cool down under pressure until the vinyl is no longer pliable. Then the disc is cooled down further to room temperature. The forces applied to the plastic during stamping remain in the record. If the record is exposed to elevated temperatures again, the forces retained within the material will be released and the disc will warp. Once this happens, the disc is destroyed. Leaving the disc in a closed car or on a window sill on a sunny day will accomplish this.

31.9.3 Record Storage

The worst enemies of records are dust, heat, and mildew. To protect records from contamination they should be kept covered in their sleeves. Sleeves should be static free if possible. Records should be stored either vertically or horizontally (freshly pressed LPs are stacked one on top of each other to prevent warpage). If stacking horizontally, sizes should not be intermixed, and the stacks should be neat and not too high. If stored vertically the records should not be loose and should not be leaning; this will introduce warpage. Record cleaners or preservatives should not be applied prior to storage because there is a good chance of mildew forming on the records if they are stored damp.

Chapter 32

Magnetic Recording and Playback

by Doug Jones

32.1 Introduction

We are now in the fifth edition of the *Handbook for Sound Engineers*. Past editions, especially editions three and four, contained lengthy and very detailed chapters on the topic of magnetic recording and playback. These chapters contained thorough discussions of all the history and theory of magnetic recording and playback. This is a very mature technology and in the years since the release of edition four, it is safe to say that there have been no breakthroughs in magnetic recording that warrant inclusion in these pages. Indeed at the time of this writing a cursory search on the internet reveals that there is only one company left building a reel to reel tape machine, a 1/4 inch 2 track. The days of the analog multi-track are sadly gone. Of course they are still being used but tape will continue to become scarcer and therefore more expensive. Hard disc drives are still being used of course, and technically they are magnetic media. However the theory is well presented elsewhere and since there are no user serviceable parts or adjustments to computer hard drives, little is to be gained by re-presenting this material. We decided that this chapter should focus on the practical aspects of keeping an analog tape recorder alive and well and properly adjusted. Those readers who wish to read about the history and theory are referred to previous editions of the *Handbook for Sound Engineers*.

32.2 Tape Transports

The mechanism responsible for moving the tape past the heads in a constant and repeatable manner is called the transport. The role of the transport is as follows:

1. To drive the tape at a repeatable, and preferably constant speed over the surface of the transducer heads.
2. To maintain a fixed mechanical alignment of the tape as it crosses the heads.
3. To provide contact pressure between the tape and head by either tensioning the tape or pushing the tape against the head.
4. To provide the necessary motions of the tape required for functions such as rewind, search, and editing.

Virtually all tape recorders utilize the layout shown in Fig. 32-1. The reels of tape are mounted on the shafts of two motors that provide the high-speed spooling and the play-mode tape tensioning. The tape moves from the supply reel on the left to the takeup reel on the right. As the tape leaves the supply reel, it is steered by guides to pass over the erase, record, and playback heads. Following the heads is a constant-speed tape drive consisting of a rotating shaft called a capstan and a pinch roller to press the tape against the surface of the capstan. The tape then passes to the takeup reel.

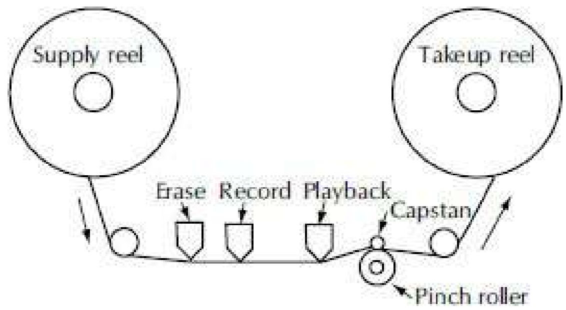


Figure 32-1. Classic tape transport layout.

This layout was used on virtually every tape recorder ever built except for the infamous Ampex 400 built sometime in the early 1950s that placed the capstan/pinch roller assembly to the left of the heads.

The typical degree of precision that was available when professional tape recorders were at their zenith includes a tape speed variation of a few hundredths of a percent, mechanical alignments of less than one-thousandth of an inch (0.001in) and three-thousandths of a degree (0.003°), and tension variations of a few percent. Even these seemingly small variations create readily observable errors in recordings. Proper transport adjustments are critical to the proper operation of a tape machine.

A very simple speed control technique was to clamp the tape to a surface that is moving at the desired tape speed, such as the outer periphery of a rotating drum. The tape is thus forced to move at exactly the same correct speed. Various implementations use drums that range from over 2 inches in diameter to tiny shafts less than 0.1

inch in diameter. In general, the larger the drum, the more accurate the tape speed control. The very small spindles are usually employed at the very slow tape speed found with compact cassettes and consumer videocassettes.

The rotating drum is called a capstan, named after a device used on sailing ships to pull in cables and hawsers, and the clamping device is called a *pinch roller*. The simplest capstan is the shaft at the end of a motor. The diameter of the shaft is chosen so that the shaft's circumference will move at the desired linear tape velocity when the motor is spinning at operating speed.

Any rotational speed disturbances in the capstan will show up as linear speed variation in the recording tape. This means that the capstan must spin at an absolutely constant speed. The simplest constant speed device is a hysteresis synchronous motor. *Synchronous* indicates that the motor runs at a speed that is locked to the frequency of the voltage driving the motor, and as long as the power company supplies the correct frequency (50Hz or 60Hz depending on the country) the machine will run on speed.

Most modern tape machines rely on servo motors for the capstan resulting in a much higher speed accuracy. A servo-controlled motor utilizes a speed measuring device on the capstan in the form of a high-resolution optical or magnetic tachometer. This tachometer may provide as many as 1200 speed samples per revolution of the motor, a rate high enough to detect not only overall average speed, but even very small speed transients due to imperfections in other components in the tape path. By comparing the speed sensed by the tachometer to a high-accuracy reference derived from a crystal oscillator, any variations or errors in speed are immediately detected. The control circuits use this error to generate corrections in the voltage driving the motor to cancel the speed error. The overall accuracy of this closed-loop system is primarily dependent on the accuracy of the tachometer and the reference clock.

Variable-speed operation for a servo system is much simpler than for the hysteresis synchronous motor. A simple variable-frequency oscillator can be substituted for the fixed reference to provide infinitely variable speeds!

The *de facto* standard for professional machines is that an external VSO frequency of 9600Hz from accessories will drive a servo at nominal speed. This 9600Hz signal can be substituted for the crystal's countdown signal at an appropriate point in the countdown chain before the final speed-determining dividers. The VSO signal is thus able to control the machine at any of the machine's running speeds.

If the tachometer is accurately mounted, if the tachometer samples occur frequently enough to provide precise sensing, if the control circuit sends the correction signal to the motor quickly so that errors are sensed as they start, and if the motor can respond swiftly to corrections in its control voltage, then the motor will turn at a constant speed. The string of "ifs" in the previous sentence is a clue to the complexity of this servo design. The results, however, of a good design are very impressive, with professional recorders being able to suppress mechanically induced speed variations to below 0.05%rms at 15in/s (38cm/s) on a routine basis.

32.2.1 A Word of Caution Regarding Pinch Rollers

The standard roller rubber is neoprene, a fairly stable rubber compound that can resist ozone and smog. Many newer compounds, especially various urethanes, have also been tried with some success. Sometimes the new roller will give excellent results when new, but then it will glaze over and lose its adhesion to the tape. In other cases the roller's elastomer will turn into a gummy ooze with the consistency of taffy.

The urethane is affected by temperature and humidity conditions, and by any solvents used to clean the tape path. Always check the cleaning pad after you clean the pinch roller. If the pad has just tape residue, you are providing proper cleaning. If, on the other hand, you see a residue that looks suspiciously like the surface of the roller, you may be dissolving your pinch roller!

32.2.2 Tape Tensioning

Magnetic recording tape, like all elastic media, must be stretched slightly to produce tension within the tape. For normal recording applications, the tape is stretched approximately 0.1% to achieve a typical tension of 4oz per ¼in of tape width. Since this small amount of stretch is less than one tenth the level of stress required to permanently deform the tape, no permanent deformation results.

Four separate and often conflicting functions are performed by tape tension on a tape recorder:

1. Tape tension holds the moving tape firmly against the record and playback heads to achieve good high-frequency performance.
2. Tension stiffens the tape on the tape guides so that the tape position will remain constant.
3. Tension controls the stacking of the layers of tape on the takeup reel.

4. On machines without pinch rollers, the tension holds the tape against the capstan to create enough drive traction for proper tape speed control.

The classic tape transport of Fig. 32-1 utilizes the supply reel spooling motor to generate tape tension over the heads in the Play mode. The supply motor is energized in the clockwise (rewind) direction with a reduced voltage, generating a constant torque from the motor. To convert motor torque to tape tension, divide the torque by the radius of the tape pack (the lever arm).

However, the radius of the pack on the supply reel decreases as the tape plays off. By the end of a 10½ inch NAB reel, the radius has dropped to half the starting value, causing the tape tension to double. (Some plastic 7 inch reels have outside to inside diameter ratios of more than 3:1.)

The tape tension is further altered to some degree by every component that comes into contact with the tape. When tape slides over any stationary guide or head surface, the tape tension changes slightly due to the friction between the tape and the stationary surface. (The bearing friction and viscous drag of rotating guides is usually negligible.) The relative contribution of friction tension to the total tape tension ranges from a low of 5% for transports with only rotating guides to over 50% for transports with numerous fixed guides and/or large tape deflection angles around fixed guides.

The amount of drag tension generated by a cylindrical post is shown in Fig. 32-2. The tension and friction build up as the tape moves around the guide. The true expression for the total drag is an exponential function, but for tape paths with only small amounts of wrap, we can approximate the tension change with the expression

$$\text{Tension change} = K \times \text{tape tension} \times \text{angle of wrap} \times \text{coefficient of friction} \quad (32\text{-}1)$$

Note that although the diameter of the guide does not appear in the tension expression, the pressure exerted by the guide against the tape surface increases as the diameter decreases. This increased pressure makes small guides wear faster and accumulate dirt more quickly. Since a speck of dirt trapped on the surface of a small guide would also be more prone to scratch the tape surface, small-radius fixed guides must be kept very clean.

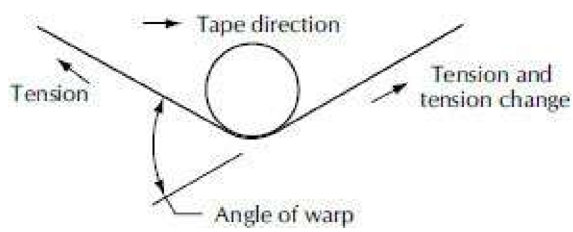


Figure 32-2. Tension increase due to guide friction.

The coefficient of friction depends not only on the type of tape, but also the condition of the roll of tape. Older tapes may lose the surface lubricants that allow the tape to slide freely across the stationary surfaces. This may result in a squealing sound as the tape runs through the recorder. Even worse, when debris from breakdown of the urethane binders in the tape collects on the tape guides, the tape may be dragged to a dead stop. These problems occur commonly when dealing with older archived tapes.

Some transport designs are more sensitive than others to changes in tape tension resulting from tape problems. Tape paths with either high amounts of drag tension or no pinch rollers may require a readjustment of the tape tension to maintain acceptable performance and avoid tape slippage. There are two primary ways that tension is controlled in transports; capstan derived tensioning and spooling motor derived tensioning. Once again, previous editions of *Handbook for Sound Engineers* has covered this topic in great detail. In general, there are few user adjustments that can be made to the capstan tensioning system. These are generally inherent to the physical design of the transport. However the spooling motor tensioning system may require periodic adjustment to keep it in proper alignment. The reader is referred to the manual for a specific transport, as there can be considerable difference between manufacturers.

The solution is to have slightly different surface velocities on the two capstans. If we need a 0.1% stretch of the tape to give us the desired 4 ounces of tape tension, then the outgoing capstan must have a surface velocity 0.1% higher than the incoming capstan. This can be achieved by using two hysteresis synchronous motors with slightly different capstan diameters. The Gauss high-speed tape duplicators used this technique with great success. A very similar technique is to use a nonstretching plastic belt to couple the drive motor to both capstans. If both capstan shafts are identical, but the pulley on the outgoing capstan is 0.1% smaller than the other pulley, the desired speed differential will be realized.

32.2.3 Tape Guiding

For proper recording and playback of a magnetic recording to occur, the tape must move over the heads in a very precise path. This tape path should be the natural path that the tape would follow without any external vertical constraints. The purpose of the guiding system is not only to protect the tape and to overcome the slight reel-to-reel variations in tape such as twists and bends due to tape-manufacturing tolerances but not to force the tape to perform any unnatural acts. Any such use of brute force will lead to tape damage, excessive guide wear, and/or instabilities and jumping of the tape.

The tape guiding system deals with five aspects of the tape motion—height, azimuth, zenith, wrap, and rack—with primary concern for the motion of the tape at the heads. Each aspect is in turn composed of two components: fixed errors due to misadjustment and dynamic errors due to tolerances and tape variations.

32.2.3.1 Tape Height

Height must be controlled so that the recorded tracks on the tape will pass directly over the pickup areas of the head. The required degree of height accuracy increases as the tracks become narrower.

For a tape guide to position the tape accurately, the tape must fit snugly into the guide, but the guide must not squeeze the tape edges. The typical manufacturing tolerances of 2mils to 4mils (50µm to 100µm) on tape width and 1–3mils (25–75µm) on tape guide width result in a loose fit for many rolls of tape.

Sources of height error also include fixed errors in head and guide height and core placement tolerances within the head. A good alignment should contain no more than 1mil (25µm) combined error for the head and guides, but this degree of accuracy requires the use of optical measurement devices that are not commonly available in a recording studio. Typical maintenance shop practices will yield errors in the range of 2–3mils (50–75µm). When this alignment error is added to a typical core placement error of 1mil (25µm) and a tape guide clearance error of 2mils (50µm), the signal loss or variation can easily exceed 1dB on a 24 track recorder.

A relatively simple method of reducing the sensitivity to height errors is to use different widths for the record and reproduce head core widths. Using either a wide playback head on a narrow recording or a narrow playback head on a wide recording will reduce or eliminate the losses due to height variation. Differing track widths, however, give rise to a common operator error. Setting the normal and sync reproduce levels from a full-track alignment tape, which has signal recorded across the entire width of the tape, will produce a level error on the wider of the two heads. The amount of error, which depends on the ratio of the core widths of the two heads, must be subtracted from the actual meter reading of the wider core to determine the true flux level. For example, a recorder with 37mil (0.93mm) record cores and 43mil (1.08mm) reproduce cores would be set to read 0VU in sync playback and +1.3VU in normal playback from a full-track alignment tape.

32.2.3.2 Head Azimuth

Not only must the tape passing across the head be at the correct height, but also the recorded signal on the tape must be parallel to the pickup gap in the reproduce head. Any angular error is referred to as *azimuth error*. For a typical professional recorder with guides spaced 6in (15.2cm) apart, the worst case combination of guide and tape sizes could produce a maximum dynamic guiding error of ± 5 mils (125 μm) at each guide, yielding an azimuth error of ±0.1 or ± 6min. This error would generate a signal fluctuation of 3dB for a 250mil (6.35 mm). Overlapping heads or tracks offer no azimuth loss improvement.

For multitrack recorders, the time and phase relationship between audio channels that are recorded on separate tracks may be more critical than the level of short-wavelength signals. Azimuth errors contribute to differential timing errors between tracks, since the azimuth tilting causes one track to be reproduced slightly later than the other. As the distance between tracks becomes large, such as for 1 inch and 2 inch (2.5cm and 5 cm) formats, the timing error becomes critical. A typical method to measure this timing error is to record the same high-frequency signal on two tracks, and then measure the phase difference between tracks. Table 32-1 shows the amount of worst-case phase difference and timing difference at a 1 mil (25µm) wavelength introduced by a 0.5 dB head azimuth error for the outer pair of tracks.

The magnitude of both the height loss and the azimuth loss could be greatly reduced if the widths of the tape guides and tape matched perfectly. One method to achieve this objective is to use adapting guides with spring-loaded movable flanges so that the guide adjusts itself to the tape width. Some digital audio recorders with numerous very narrow tracks utilize spring-loaded guides to maintain close repeatability of the tape path.

Table 32-1. Errors Due to 0.5dB Azimuth Error (1mil Wavelength)

Format	Phase Error	Timing Error
¼ inch stereo	151°	0.28ms
1 inch 8 track	867° (2.4 rotations)	0.16ms
2 inch 24 track	3500° (9.7 rotations)	0.65ms

32.2.3.3 Tape Guides

Tape guides come in many shapes, sizes, and basic types, as shown in Fig. 32-3. Each guide contains flanges that press against the edges of the tape to steer it. In all cases except the edge-only guide, the tape wraps around the guide to generate stiffness so that the steering force exerted by the flange can move the entire width of the tape and not just buckle the edge. Typically, at least 10° of wrap is required for adequate stiffness.

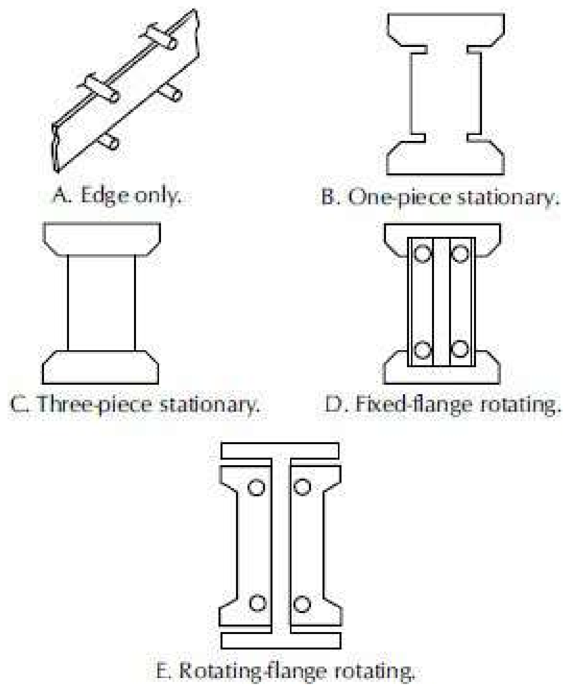


Figure 32-3. Five styles of tape guides.

Rotating guides are generally less effective than stationary guides. Since the tape is in firm contact with the spinning surface of a rotating guide, rather than in sliding contact as with the stationary guide, the force required to slide the tape up or down is determined by the tape tension and the coefficient of *static friction*. The tension component is identical for the stationary guide, but in this case the coefficient of *sliding friction*, which is typically half the static value, is used.

Although both stationary and rotating guides are commonly used in tape transports, rotating guides are slightly more prone to damage the edge of the tape. Guides with large rotating flanges can produce ruffles on the edge of the tape if the tape edge contacts the outer radius of the moving flange. Most guide designs taper the flange to minimize this hazard, but a small flat area at the bottom of the taper is still required if the guide is used for precise tape positioning.

The edge-only guide is very limited in effectiveness since any appreciable force on the edge of the tape may cause the tape to twist rather than move up or down.

32.3 Magnetic Heads

So far we have discussed the parts of the transport which guide and control the motion of the tape. Now we will take a look at the transducer which converts the magnetic signal to an audio signal; the tape heads. Many of the non-linearities that can occur in the record/playback process are a result of the construction and alignment of the magnetic heads.

32.3.1 Geometric Characteristics

Most of the characteristics of magnetic heads are controlled by the geometry of the head and the magnetic tape. Since wavelength on tape is determined by the recorded frequency in hertz and the relative tape-to-head speed, there can be many combinations of frequency and speed that will result in the same effects in a head. For example, the wavelength of a 15Hz tone on a mastering recorder at 15in/s will have the same wavelength as a 240kHz signal on a high-speed tape duplicator running at 240in/s. The geometric considerations for both applications are identical, despite the 16:1 difference in tape speed.

Gap Length Loss

Each of the tiny magnetic particles on the surface of the tape produces a magnetic force or flux in the space surrounding the particle. This invisible magnetic effect, called a *magnetic field*, will interact with other nearby magnetic particles. To measure the strength of this field, a flux concentrator in the form of a reproduce head is scanned along the tape. The resulting electrical output from the head is dependent on the flux pattern recorded on the tape.

The reproduce head must be able to collect flux selectively from a very small span of tape. For example, flux patterns on a compact cassette may be as small as 100 millionths of an inch (100×10^{-6}in or 2.5μm) in wavelength. To achieve this fine resolution, a small gap must be created in a ring of magnetic material, as shown in Fig. 32-4A.

The length of the gap ranges from two ten-thousandths of an inch (2×10^{-4}in or 5μm) for studio mastering recorders down to less than 30 millionths of an inch (30×10^{-6}in or 0.75μm)—the wavelength of red light—for cassette and high-density digital recorders. Since no slicing technique is available to cut accurate gaps that short, the core is usually fabricated as two pole pieces that are fastened together with a shim spacer of the desired dimension inserted in the gap. Fig. 32-4B shows a typical studio head core drawn full size, with the critical gap area at the pole tips and adjacent tape magnified in Fig. 32-4C.

The operation of the gap, which serves as a sensing aperture, can be analyzed in terms of a flux pickup focused at the surface of the tape. The amount of flux picked up by the core, and thus made available to generate an output voltage in the winding, is determined by the *net* magnetic flux from pole tip to pole tip across the gap area. If the tape segment at the gap consists of a strong magnetization of only one polarity, the flux in the core will be maximized. If, on the other hand, the segment contains two strong portions of opposite polarity that cancel each other, the net flux in the core will be zero.

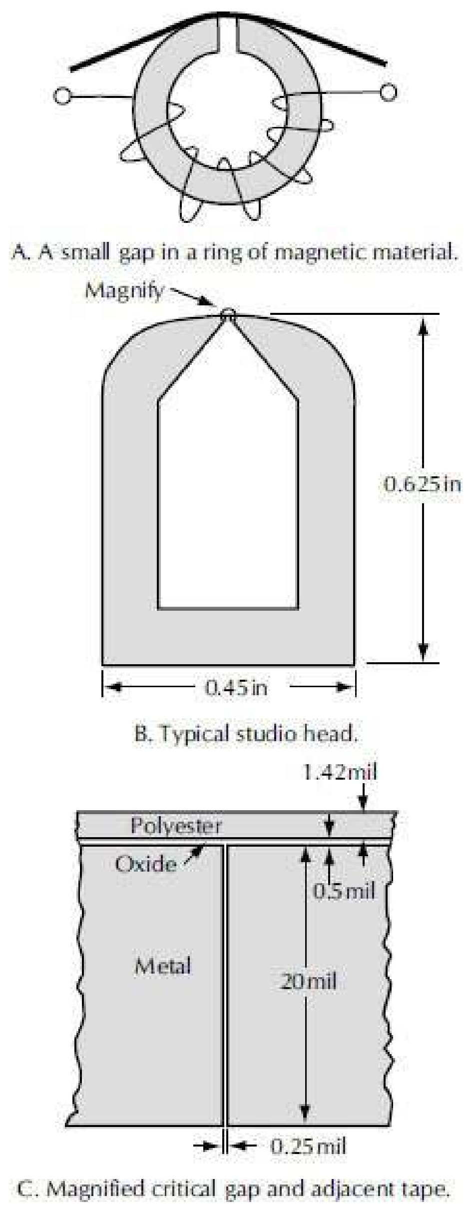


Figure 32-4. Ideal and practical magnetic heads.

The efficiency of the gap due to this averaging effect is illustrated in Fig. 32-5. The output of the head declines, slowly at first, and then quite rapidly to zero as the wavelength decreases to the length of the gap. As the gap length becomes longer than the wavelength, an output of opposite polarity appears. When the wavelength drops to half the gap length, another null will occur. This pattern of diminishing peaks of alternating polarity is repeated over and over, with nulls occurring at each wavelength that produces an odd or even number of complete cycles in the gap.

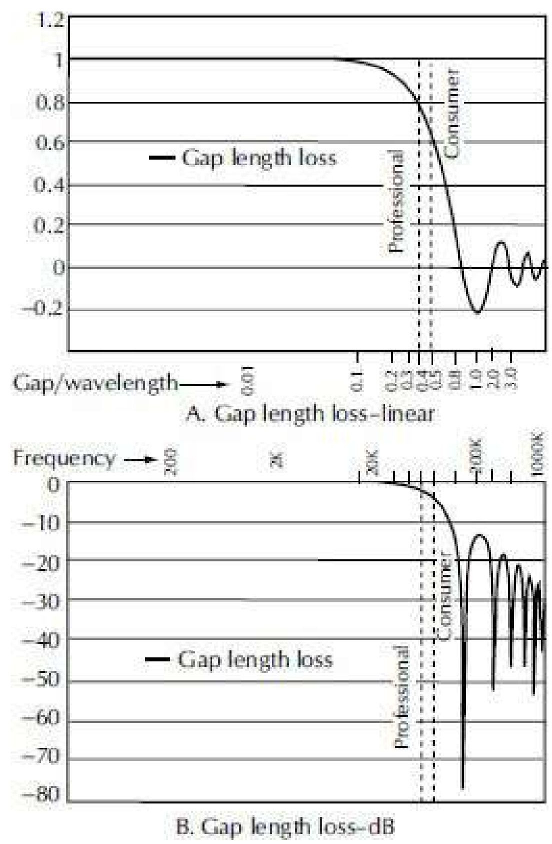


Figure 32-5. Loss due to gap length and the ratio of gap length to wavelength.

The example recorder at 15in/s has only 1dB of gap length loss at 40kHz and 3dB at 67kHz, certainly not a dominating loss. At 30in/s these losses become even more insignificant, with the −1dB and

−3dB frequencies doubling to 80kHz and 134kHz.

Audio recorders are seldom designed to operate beyond the dashed lines shown in Fig. 32-5. With this constraint, gap length loss for professional machines can be held below 1dB or 2dB by choosing an appropriate gap length for a given application and minimum wavelength. Mastering recorders operating at 15in/s and 30in/s and broadcast machines operating at 7.5in/s have playback gaps ranging from 100–200μin, compact cassette machines operating at 1⅞ in/s have gaps of 30–60μin.

Mastering recorders may also use the record head for playback in the *sync* mode. Since the record heads may have gaps ranging from 250μin to 1000μin, the sync response may suffer significant high-end loss. For example, a 1950s vintage recorder with a 1000μin record gap will reach its first null at 15kHz for a tape speed of 15in/s. As sync response became more important in the mid-1960s the recorder manufacturers tightened up the record gaps to 350μin or less to improve sync response.

If the gap length is inferred from the first measured null, this *effective gap length* may be 10% to 15% longer than the mechanical gap determined by the shim. Various proposed explanations include magnetic degradation of the inner surfaces of the pole tips due to manufacturing stresses and pole tip saturation. When in doubt, add 10% to the optically measured length or shift the response points down to 91% (1/1.1) of the theoretical values. For the ATR100 example, the −1dB and recorded at the level −3dB points would shift to 36kHz and 61kHz.

Use of an excessively short gap will cause an additional loss in overall head sensitivity due to shunted flux that jumps the gap rather than traveling through the core, as shown in Fig. 32-6. For this reason, the reproduce head gap length is usually chosen to give the largest acceptable loss at the shortest expected wavelength.

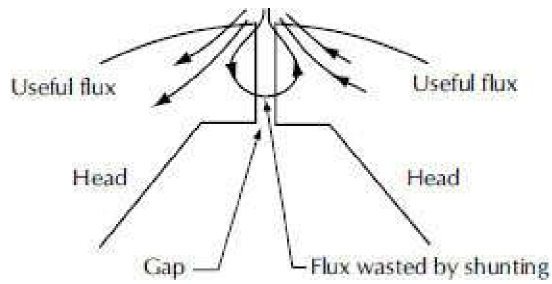


Figure 32-6. Gap shunting loss.

32.3.2 Spacing Losses and Thickness

The recording process magnetically aligns groups of the tiny randomly oriented magnetic particles so that they act as if they were a single larger particle. We could visualize these groups as little bar magnets that have dimensions determined by the tape and signal. The track width defines the vertical direction and the tape coating thickness sets the depth. The length is determined by the wavelength of the recorded signal. To simplify the example, assume that a 1.5kHz square wave is recorded at a tape speed of 15 in/s, yielding a wavelength of 10mils or 0.010in. The recorded image is similar to a series of bar magnets each 5mils long with alternating polarity.

Actually, gap length loss and shunting loss are only a part of what determines the performance of an audio recorder. The most critical parameter is the relative thickness of the magnetic coating on the tape. The ratio of tape thickness to the shortest wavelength to be recorded has a profound effect on the frequency response, maximum output, noise, and signal-level fluctuations.

The magnetic particles at the surface of the tape are very tightly coupled to the core of the head, producing a maximum amount of playback flux in the core. Particles that are buried below the surface of the tape, however, produce a weaker flux in the core. The amount of flux that is lost depends on the spacing distance and the wavelength—just as a small font size is more difficult to read at a distance than a larger font. An approximate expression for this spacing loss is

$$Spacing\ loss_{dB} = 55 \times \frac{distance}{wavelength} \qquad (32\text{-}2)$$

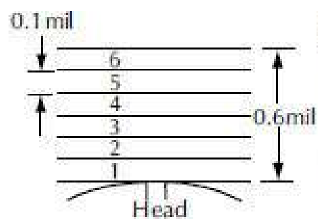
One example of the use of this spacing loss formula is to determine the playback signal loss due to a piece of dirt on the surface of a reproduce head. Assuming a typical recording studio tape speed of 15in/s (38cm/s), a dirt speck only 0.0001in (2.5μm) high will produce losses at the following frequencies of

$$\begin{aligned} 150\text{ dB spacing loss} &= 55 \times \frac{0.0001}{\frac{15}{150}} \\ &= 0.055\text{ dB} \\ 1500\text{ Hz spacing loss} &= 0.55\text{ dB} \\ 15\text{ kHz spacing loss} &= 5.5\text{ dB} \end{aligned}$$

Note that this seemingly insignificant dirt particle has produced a serious loss in high frequencies.

Spacing loss due to dirt is not the major problem created by the "nearsightedness" of the gap since proper head cleaning will keep spacing distances to less than 10^{-5} in, which is (0.25μm), producing virtually no error at studio tape speeds. The problem is eight times more severe for cassette speeds of 1⅞ in/s (4.8mm/s).

The major spacing problem arises within the tape itself since the magnetic coating thickness spaces most of the particles away from the head with other particles. Consider the tape to be composed of several independent layers of oxide, as shown in Fig. 32-7. The average spacing loss for each layer, calculated using the midpoint of each layer to determine the spacing distance, is tabulated for the example with a typical 0.6mil (15µm) coating thickness.



Layer	Spacing	15kHz loss
1	0.05	−2.75 dB
2	0.15	−8.25 dB
3	0.25	−13.75 dB
4	0.35	−19.25 dB
5	0.45	−24.75 dB
6	0.55	−30.25 dB
Total output		+ 3.63 dB

Figure 32-7. Tape thickness loss.

The contributions of layers 2 through 6 fall off so rapidly due to spacing loss that their *combined* contribution is only equal to layer 1 by itself at this wavelength. Indeed, shaving off layer 6, which constitutes 17% of the coating thickness, would produce a loss of only 2% or 0.18dB in output at this wavelength.

This coating thickness loss can be expressed as

$$\text{Coating thickness loss}_{dB} = 20\log\frac{x}{1-e^{-x}} \qquad (32\text{-}3)$$

where,
x is $2\pi \times$ thickness/wavelength.

Although this expression yields a drop of 6dB per octave, as shown in Fig. 32-8, this curve is not the same shape as the response of a low-pass filter made from a resistor and capacitor. The response in Fig. 32-8 is down 4dB at the intersection of the asymptotes rather than the typical 3dB for a single pole RC filter.

This difference in shapes means that a simple RC boost circuit will not properly correct for the thickness loss. Depending on the choice of RC boost frequencies, the difference in shape will produce an error of 0.5–1.0dB in the midband response.

32.3.2.1 Equalization Boosts

This thickness loss of Fig. 32-8, must be corrected by applying compensating boosts in either the record or reproduce circuitry. Although this loss is a playback deficiency, the choice of whether to correct the loss during record or playback is somewhat arbitrary. The amount of record boost is limited by the magnetic saturation characteristics of the tape; playback boost is limited by the high-frequency noise characteristics of the tape and the reproduce head and associated circuitry.

The minimum amount of boosting required to achieve flat response can be considered to be a necessary equalization. The industry has developed a set of internationally recognized standards to promote compatibility of tapes. Each standard deals with the necessary and discretionary equalizations to define the exact characteristics of the recorded tape. Using the tape flux characteristics as a standard implicitly specifies the partitioning of equalizations between the recording and reproducing functions. Table 32-2 lists the commonly encountered standards.

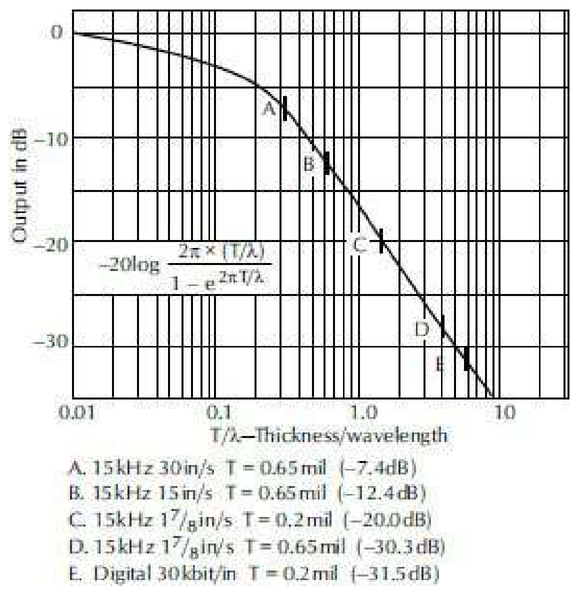


Figure 32-8. Loss due to ratio of coating thickness to wavelength.

Unlike the absolute nature of the reproduce characteristics, the record characteristics of the recorder must have enough flexibility to accommodate a number of different tape sensitivities and frequency characteristics. Once the reproduce section has been calibrated to the standard with a standard alignment tape, all further adjustments are to produce a recorded tape on the machine that accurately matches the standard tape.

The amount of thickness loss can always be reduced by utilizing thinner coatings, but any decrease in thickness also causes an equal drop in low- and midfrequency output and *SNR*. To preserve the existing standards, the tendency has been to adjust the coating thickness of new tapes to emulate the high-frequency losses of the older tape types while trying to achieve maximum low-frequency output. This somewhat self-defeating strategy has been overcome in recent thin-coat high-energy tapes that retain the low-frequency output capability of older tapes, but utilize new equalization curves optimized for the new tape thickness.

Table 32-2. Common Tape Record-Playback Equipment Equalization Standards

Standard	Type Tape	Transition Frequencies and Time Constants				
		1⅞ ips	3¾ ips	7½ ips	15 ips	30 ips
IEC	Fe_2O_3	100/1326Hz	50/1768Hz	0/2274Hz	0/4547Hz	
		1590/120µs	3180/90µs	∞/70µs	∞/35µs	
	Metal	100/2274Hz				
		1590/70µs				
NAB		50/1768Hz	50/1768Hz	50/3180Hz	50/3180Hz	
		3180/90µs	3180/90µs	3180/50µs	3180/50µs	
AES						0/9095Hz ∞/17.5µs

32.3.2.2 Contour Effect

At very low frequencies, the wavelength of the recorded signal may become as long as the magnetic core of the playback head. These long wavelengths enter the core at the gap and at the sides and rear of the core. The resulting flux in the core will consist of the desired flux from the gap plus additions and/or subtractions of the fringing flux leaking into the core at the sides and back. The voltage output of the head, which is dependent on the net flux coupled into the windings, will undulate at low frequencies as the wavelengths create varying levels of constructive and destructive interference due to the fringing flux.

The response curve in Fig. 32-9 illustrates the nature of the undulations commonly called *head bumps* for a typical mastering recorder at 15in/s (38cm/s) and 30in/s (76 cm/s) using a reproduce head that has a 0.5 in (12mm) core face. Two well-defined head bumps are usually evident for such mastering heads. The bumps shift up an octave in frequency for each doubling of tape speed, creating an even more severe problem at 30in/s (76cm/s).

Heads with either very small cores or only a small window in the head shielding at the gap area can produce numerous ripples in the low-frequency response. Such heads should be avoided unless the tape speed is slow enough to avoid serious problems within the normal band of audio frequencies.

The exact shape of the head bumps is determined by the size and shape of the reproduce core, surrounding shielding material, and angle of wrap of the tape. Since the user cannot adjust these parameters during the normal alignment procedure, the bumps can only be modified by adding an outboard equalizer, which cancels the bumps with an inverse response curve.

Improvements in the control of head bumps has reduced the magnitude of the bumps in present-day mastering recorders to less than 1dB peak-to-peak at 15in/s (38 cm/s) and 1.5dB peak-to-peak at 30 in/s (76cm/s). Beware that this level of error will be introduced each time the tape is rerecorded during mixdown and subsequent protection copying. The total error can easily reach 5dB or more for a typical sequence of operations.

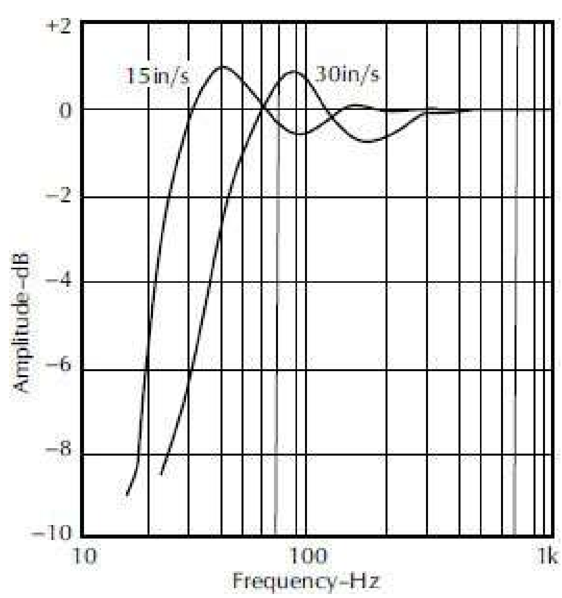


Figure 32-9. Contour effect. Courtesy Sony Corporation of

America.

32.3.2.3 Crosstalk

Multitrack recorders will exhibit playback signal leakage or crosstalk between adjacent tracks at long wavelengths. The unused area or guard bands between the cores of the head, which are nearly equal in width to the recorded track, usually provide enough of a physical gap to prevent flux from spilling from one track to the next. At long wavelengths, however, the fringing flux will jump the guard band, producing low-frequency crosstalk.

The crosstalk component due to fringing will initially decrease as the frequency is increased, but at midband the decrease will eventually bottom out. The remaining residual level of crosstalk is not due to fringing, but it is a direct transformer-like coupling of leakage flux between the adjacent cores in either the record or reproduce head. A layer of magnetic shielding material is typically placed between the cores of the head as a crosstalk shield to reduce this flux leakage.

32.3.3 Noise

The useful range of signal levels that pass through the tape recorder is limited by the maximum signal at which all the magnetic tape particles become completely magnetized or saturated and also by the amount of noise that remains when the input signal is removed. Noise in tape recorders has many sources; the electronics, the tape, and the heads themselves all contribute to the residual noise.

The distortion content of the signal from a tape recorder rises so dramatically near tape saturation that the normal operating range must be limited to less-than-maximum levels. For the purpose of specifying and comparing tape recorders, the distortion-free maximum operating level is typically considered to be the output signal level at which the *THD*, which is dominated by third harmonic and other odd components, reaches 3%. The ratio of the level for 3% THD at medium wavelength to the residual noise is defined as the *SNR* of the recorder.

Track Width

The second factor is the loss in *SNR* in narrow-track consumer tape formats due to the dissimilar ways that random noise sources and coherent signals increase. The noise due to the tape, heads, and electronics is a random combination of many small independent noise bursts. If two equal and independent random noise sources of this type are added together, the noise *power* is doubled, producing an increase of 3dB on a voltmeter.

Coherent sources, on the other hand, are merely duplicates of the same waveform. If two identical sources are added together, the value at each point on the output waveform is exactly twice the value of either of the input waveforms. In this case the output *voltage* is doubled, or a 6dB increase.

Consider the case of two tracks of a tape recorder that have recorded the same signal. If the output signals of the two tracks are added, the noise will add randomly and the signals will add coherently. The combined tracks have 6dB more signal and 3dB more noise, yielding a net *SNR* improvement of 3dB. Using a single track of double the original track width would produce the same result if the noise sources were statistically independent in nature.

The tape noise will follow the 3dB per doubling rate if the reproduce amplifier noise is less than the tape noise. The reproduce amplifier noise typically remains nearly constant regardless of track width of the head. The apparent noise will vary, however, as the gain of the amplifier is adjusted to compensate for changes in the head output due to increased or decreased track width. When tracks are made narrower, the amplifier noise that functions as a coherent source will eventually dominate the tape noise, creating a signal-to-noise loss of 6dB per halving of tape width.

32.3.4 Record Heads

The magnetic core and gap of a reproduce head obey the principle of reciprocity, which states that the roles of an excitation source and sensor can be interchanged. For a head used in the reproduce mode, external flux at the gap produces a voltage across the head winding. If, instead, a voltage is applied to the head winding, a concentrated external flux field will be generated at the gap and can be used to record a signal on a piece of moving tape.

The shape and strength of the magnetic field at the gap is the basis for the operation of a recording head. The flux generated in the core by the current in the winding must jump across the gap to complete a closed magnetic path. The gap, which is a very poor magnetic path compared to the core, produces an obstruction that forces the flux to spread sideways, as shown in Fig. 32-10.

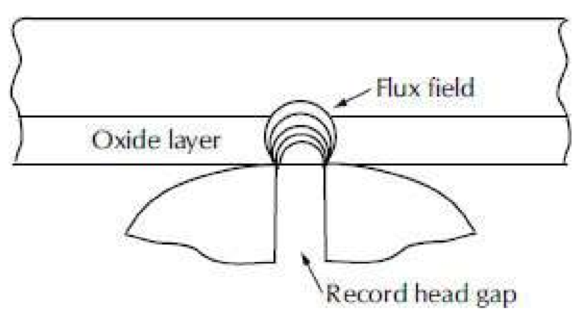


Figure 32-10. Record head flux field.

An analogous situation occurs with a crowd of people moving down a hallway. If the hallway widens for a crossing corridor or small lobby, the crowd will broaden out into the open area and then narrow down again to reenter the continuation of the hallway. The broadening will increase if the pressure within the hallway should increase due to an emergency such as a fire. The stress is greatest at the transitions between the wide and narrow spaces since this is where people are squeezing to try to change the shape of the flow.

A magnetic tape passing over a record head gap experiences a similar buildup and decline in the magnetic recording field as it moves across the gap. To produce a permanent recording on the tape, the flux must first rise to a level sufficient to overcome the magnetic memory force of the tape, which normally keeps the magnetic particles on the tape from changing state spontaneously. In the central zone of complete excitation, the tape particles will follow any change in the input signal driving the head. As the tape particles exit the strong central zone, a well-defined point will be reached at which the driving flux drops below the memory force, leaving a fixed magnetic image impressed on the tape. This transition region in which the image freezes at the trailing edge of the gap is called the *trapping plane*.

The shape of the trapping plane depends primarily on the gap size and the thickness and magnetic characteristics of the tape. Since trapping planes that are narrow and vertical will produce short-wavelength recordings that are more easily reproduced, several techniques have been developed to sharpen the transition zone, as shown in Fig. 32-11.

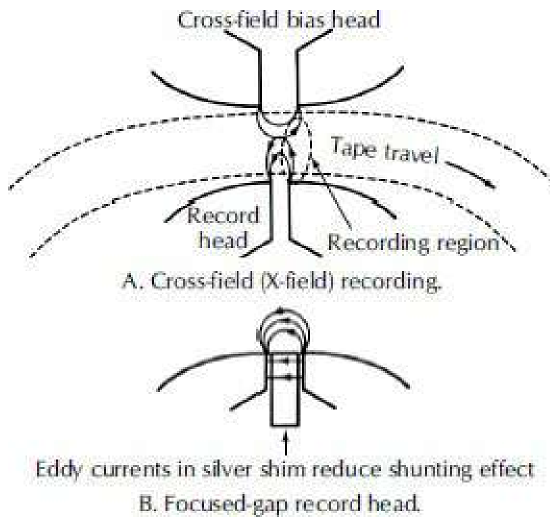


Figure 32-11. Focused-gap and cross-field (X-field) head.

The focused-gap technique in Fig. 32-11B uses a highly conductive gap shim made of silver to serve as a barrier to flux jumping straight across the gap. Eddy currents in the shim force the flux away from the shim, squirting the flux deeper into the tape. The reduction in shunted flux raises the efficiency of the head by requiring less drive power.

The conductive shim is only effective at high frequencies at which large eddy currents are generated in the shim. As a result, focused gap recorders utilize bias frequencies that are approximately ten times higher than conventional systems.

In practical use the silver shim proved to be a major problem because the soft silver would smear onto the trailing pole piece of the head and short the head's laminations together.

A second technique, which yields similar results, is the *crossed field* or *X-field*, Fig. 32-11A. This method typically places a second bias-only head on the back side of the tape to create a shaped bias flux field jumping from one head to the other.

Bias

The magnetization of the tape particles is not easily changed due to the memory force or hysteresis of the particles. In fact, the particles have a form of inertia that must be overcome if a linear transfer is to be achieved. See section 32.3.10 for methods of setting bias levels.

If a rapidly varying signal of sufficient amplitude to just begin magnetizing the particles is added to the audio flux signal, the magnetic particles will more readily conform to changes in the audio waveform. A high-frequency biasing signal produces a hysteresis-free or anhysteretic recording.

Fig. 32-12 shows a typical waveform of the current in a low impedance Ampex record head that is recording 10kHz at a level of 250nW/m. The bias component of $7mA_{p\text{-}p}$ is approximately ten times larger than the 10kHz component at 650µA. (The voltage waveform across the record head would be totally dominated by the bias component due to the 6dB/octave rise in head impedance with increasing frequency, in this case 35V of bias versus 500mV of 10kHz or 70:1.)

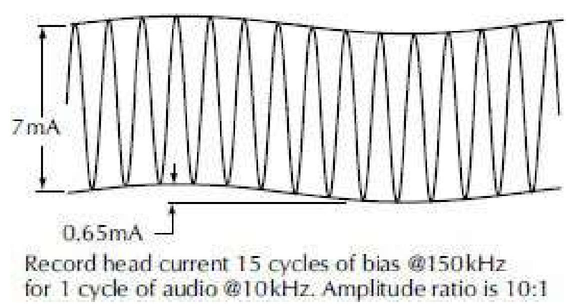


Figure 32-12. Record head current.

The audio and bias signals must be added together in a linear manner without generating any of the sidebands that are present in either amplitude or frequency modulation techniques. The short-wavelength bias signal can therefore be easily filtered out during playback by the gap and thickness losses so that only the audio signal remains. (The high level of bias signal transformer crosstalk that is present during sync/overdub operation requires a sharp notch filter in the playback preamplifier to remove the bias signal.)

Typical bias frequencies range from 100kHz for slow-speed recorders to over 10MHz for high-speed tape duplicators. Although high bias frequencies are desirable to permit easy filtering and thorough tape excitation, a practical upper limit for mastering recorders is reached at 500kHz due to a combination of increased eddy current and hysteresis losses in the core and the increase in bias drive voltage required due to the inductance of the head.

Head losses can be reduced by using a very small core to reduce hysteresis losses and by choosing either thin laminations or a ferrite material to reduce eddy current losses. If, however, the record head will also be used for the reproduce function during sync/overdub, a small core will cause serious long-wavelength contour effects. The compromise *hammerhead* design shown in Fig. 32-13 improves the playback performance of the small core by adding extensions to the face of the core. The tips function only to play back low-frequency signals for which core losses are insignificant.

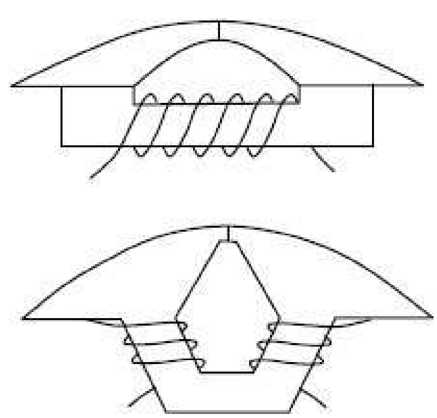


Figure 32-13. Hammerhead cores.

The bias voltage required to drive a record head doubles each time the bias frequency is doubled due to the inductance of the record head. To keep the required bias voltage within the range of common integrated circuits, the inductance can be lowered either by reducing the number of turns in the winding or by lengthening the gap. Reducing the number of turns once again degrades the sync/overdub performance by reducing the playback voltage generated by the head. Heads with very low inductance typically require a step-up transformer to achieve adequate playback *SNR*s, but the transformer will also contribute some additional small amounts of distortion, noise, and frequency response anomalies.

Lengthening the record head gap will reduce shunting and give better bias penetration into the tape, but the short-wavelength sync/overdub response will suffer greatly.

A more straightforward approach to optimize the record head for both recording and playback is to use separate flux paths or windings for each of the functions. One simple method of switching windings and flux paths is to use parallel paths that can be selectively blocked. As shown in Fig. 32-14, when the high-inductance playback winding is shorted, flux will be blocked from the playback shunt magnetic leg of the core, effectively eliminating this path and thereby forcing all of the flux from the low-impedance bias winding to the front of the head. During reproduce, when the bias winding is shorted, the flux picked up from the tape will pass only through the reproduce winding. Although the cost of this dual-winding head is significantly higher than for a conventional single-path design, each coil can be optimized for its intended function without the need for compromise, yielding playback-to-record inductance ratios of up to 1000:1.

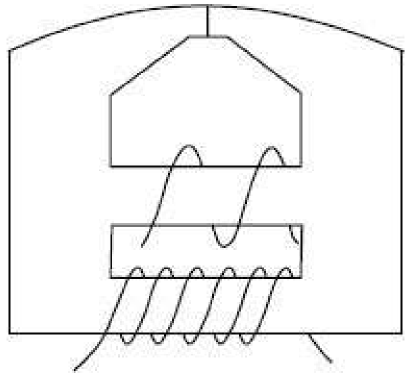


Figure 32-14. Dual winding record head.

32.3.5 Erase Heads

A major advantage of magnetic tape recording is the ability to erase easily and reuse the magnetic tape. Although physical wear may eventually degrade the performance of the tape, the magnetic properties of the tape never wear out.

Erasure of the tape can be accomplished by remagnetizing the tape with either a very strong static field or a very strong alternating field. For audio applications the alternating field, which produces a completely random flux pattern that is very quiet, is used exclusively.

32.3.6 Head Degaussing (Demagnetizing)

Early tape recorders used permanent magnets rather than an ac high-frequency signal to bias the tape so that small signals could be recorded without high distortion. These fixed magnetic fields produced a very high background noise level that severely limited the *SNR* of the taped recording. The introduction of ac bias upgraded the tape recorder from a voice-grade recording instrument to a true high fidelity recorder for music.

Modern recorders all use ac bias, but occasionally the background noise on a tape will be well above the normal level. The culprit is usually a permanently magnetized head, guide, or capstan that is acting like one of the old biasing magnets. The problem is most commonly created by touching a magnetized tool such as a screwdriver or razor blade to a component in the tape path. On rare occasions a faulty electronic circuit will create a dc current in one of the heads, leaving a residual magnetic field. (Loud clicks or thumps may be symptoms of dc currents.)

Since there are no commonly available instruments which can detect the very small magnetic fields which will result in noise, the best strategy is to frequently demagnetize all magnetic components in the tape path with a head degausser, Fig. 32-15.

The head degausser in Fig. 32-15 is an electromagnet with an extended core. The extension probe conducts an alternating magnetic flux generated in the coil to the tip of the probe. The probe is passed close to the magnetic components on the tape deck so that the alternating flux can flood the components. The actual demagnetizing occurs as the probe is slowly withdrawn from the component, creating the gradually decreasing alternating magnetic field mentioned previously in the discussion of bulk degaussers and erase heads.

Caution! Before using a head degausser, always verify that the tip of the probe is covered by a soft material that will not scratch the face of the magnetic heads. If necessary, wrap the tip with vinyl electrical tape or a similar tape.

Degauss the heads and other steel tape-guiding parts with a commercial-grade head degausser as follows:

1. Although a typical head degausser will not disturb a recorded tape that is more than a few inches from the degausser, always remove all tapes from the vicinity of the transport prior to energizing the degausser.

Figure 32-15. A head degausser.

2. Hold the degausser at least 1ft from the tape transport when applying power to the degausser. The degausser will produce a large voltage in the playback and record heads, which will probably not damage the respective electronics but will certainly "peg" any analog meters in the circuit. Turn off the power to the recorder before using the degausser.
3. Move the degausser slowly and smoothly from bottom to top along the gap line of each head, moving at a rate of approximately 0.125 in/s (3mm/s). At the top of the head, smoothly withdraw the degausser 6in (15cm) and then move

smoothly to the next item to be degaussed.

4. To be safe, move the degausser at least 3ft (1m) away from the transport before disconnecting the power from the degausser.
5. Multiple degaussing passes on a component do not improve the quality of the results. A single slow, smooth pass is adequate.

The rapid collapse of the magnetic degaussing field at turnoff can easily undo all of the benefits of degaussing if the degausser has not been pulled away sufficiently. (For this reason, avoid degaussers that have momentary power switches that might be accidentally released in the middle of the degaussing routine.)

32.4 Tape Components

Modern magnetic tape consists of a powder of very small magnetic particles, which has been glued to one surface of a plastic substrate or base film. The backside of the substrate is coated with a very thin layer of carbon particles to improve winding characteristics and to reduce the buildup of static electricity.

32.4.1 Base Films

Although several base film materials were used in the past, including paper and acetate film, virtually all tape manufactured today uses polyester film (polyethylene terephthalate) such as Dupont’s Mylar ™. Polyester is not only extremely strong and tear-resistant, but it is also relatively stable with respect to changes in temperature and humidity.

Depending on the intended application, the nominal base film thickness ranges from 1.4mils (35µm) for heavy-duty professional tapes down to a scant 0.25mil (6.25µm) for a C-120 cassette. To achieve reliable performance with these very thin films, the film must be not only very thin but also uniform in thickness from end to end and from edge to edge.

To enhance the strength of the thin base films used for cassettes, the polyester is prestretched or Tensilized. Although Tensilized tapes are more resistant to stretching than normal tapes, residual stresses that result from the Tensilizing process can produce physical distortion of the tape. For thin, narrow tapes these distortions are satisfactorily flattened out at the record and playback heads. The thicker, wider tapes used for professional formats, which would manifest severe contact problems due to these distortions, are considered to be strong enough without Tensilizing to provide adequate performance.

32.4.1.1 Binders

The glue or binder that holds the magnetic particles to the base film is a necessary evil that makes no active contribution to the magnetic performance of the tape. The use of new high-strength binders containing urethanes has improved both the durability and the recording characteristics of recent tapes.

The magnetic characteristics of the magnetic particles never wear out. The particles can be recorded and/or reproduced an unlimited number of times without any performance degradation.

The useful life of the tape is determined by three factors—the inherent strength of the tape, the amount of physical wear caused by the tape transport, and the performance required by the application. A typical test to measure the life of the tape would consist of many repetitive cycles on the intended transport while monitoring the gradual (hopefully) drop in playback level at the shortest wavelength of interest. When these losses exceed the application's requirements, the tape is worn out.

Some specialized audio transports designed for repetitive playback are capable of making over a quarter of a million passes on a tape. On the other hand, a poorly maintained studio recorder can destroy a master tape in ten passes or less! In general, if the abrasive forces exerted by the transport on the tape are well below the inherent strength of the binder, the tape will last virtually indefinitely. Any increase in the abrasive force due to dirty contact surfaces, excessive tape tension, or poorly designed tape guiding will accelerate the wear.

A very rapid catastrophic failure will occur once the abrasion force becomes sufficient to build up a small clump of debris on a contact surface. The friction between the debris and the tape surface is very high due to both the similarity of materials and the high pressure exerted by the tip of the clump as it pushes on the tape. The binder is overwhelmed, causing the clump to grow rapidly to the point at which the tape will show an obvious scratch or crease. If this situation should arise, the source of the problem should be corrected, and a copy of the damaged tape should be used for subsequent work.

From the magnetic performance standpoint, the combination of smoother magnetic particles and newer binders has enabled the tape manufacturers to use a smaller quantity of binder material to affix the magnetic particles. The ratio of useful particles to the magnetically inert binder rose from approximately 40% by volume for typical mastering tapes in 1970 to approximately 60% in 1980 with virtually no improvement since then. This improved magnetic density yields a higher maximum output for a given particle type and coating thickness.

32.4.1.2 Magnetic Particles

The ultimate performance of a tape recorder is determined not by the tape drive, heads, or electronics, but rather by the physical and magnetic characteristics of the magnetic particles of the tape. If basic performance parameters such as maximum output levels, noise, and distortion are truly determined only by the tape, the recorder is said to be tape limited. As a practical rule of thumb, if the noise and distortion products of the recorder are at least 10dB lower than the products produced by the tape, the overall performance of the machine and tape will be within 0.5dB of the theoretical levels of the tape alone.

Of primary importance in magnetic recording is the ability of each magnetic tape particle to assume and retain a magnetic pattern. These particles are chosen for their ability to maintain a magnetic field along one preferred direction or axis, permitting alignment of the particles for maximum performance. The amount of preferred orientation or anisotropy in the material depends on the nature and crystalline structure of the particles.

The shape of the particles determines the degree of physical alignment that can be achieved during the coating process. Smooth cylindrical or spherical particles that have no jagged edges or branches can be densely packed, yielding maximum output level.

The size of the particles is determined by the crystalline structure of each material. The residual noise of the tape decreases as the particles become smaller. Small particles with high anisotropy are therefore most desirable. Typical iron oxide magnetic particles for recording tape are cigar-shaped particles with a length-to-width ratio in the range of 4:1 to 8:1.

The newest recording products are abandoning particulate coatings in favor of thin layers that are plated or evaporated onto the surface of the plastic. These very thin layers of high coercivity materials are ideal for very short video wavelengths or very high digital bit densities. The new technology brings with it a whole new set of problems such as coating durability and how to include adequate lubrication in the metallic coating.

Coercivity. The *coercivity* is a measure of the magnetic force required to cause the tape particles to change magnetic polarity. High coercivity particles are more difficult to bias, record, and erase. On the beneficial side, they are also better able to resist external influences due to neighboring particles after recording, reducing the smearing of short-wavelength signals during storage.

Retentivity and Remanence. If the coercivity is considered to be the input drive, then the *retentivity* and *remanence* are the output of magnetism left in the tape. Retentivity measures the maximum output per unit volume of coating cross section; remanence (remanent flux), which is the output per ¼ inch of tape width, varies not only with retentivity, but also with coating thickness. Remanence specifications should be used to compare the maximum long-wavelength outputs of different tape types.

32.4.2 Magnetic Performance Curves

The input-output relationship for typical magnetic materials is very nonlinear. As shown in Fig. 32-16A, the magnetization characteristic curve can be broken into three zones. For low excitation levels, the initial output is very small and nonlinear. As the excitation increases, a fairly linear region is encountered, which produces low distortions. As the level continues to increase, the magnetic particles finally become fully magnetized or saturated. Further increase at the input yields no more magnetization in the material.

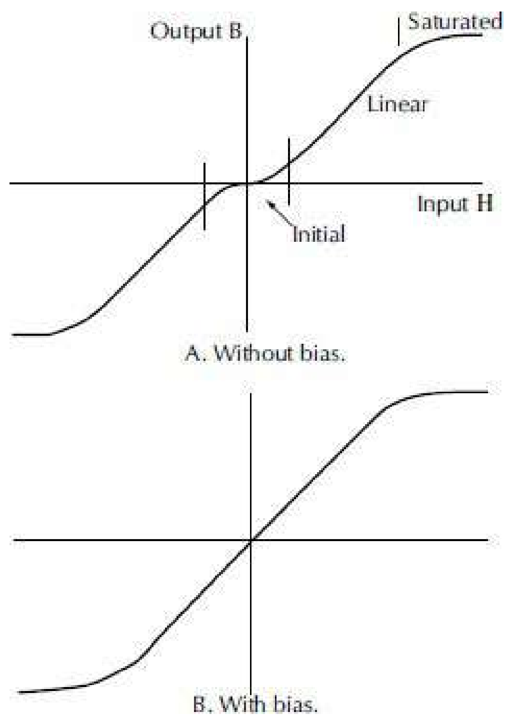


Figure 32-16. Tape transfer characteristics.

The nonlinear initial region must be avoided in audio recording if low distortion is to be achieved. The high-frequency bias signal provides enough excitation to jolt the magnetic particles into an active state. Optimizing the bias level yields the much more linear transfer characteristic of Fig. 32-16B.

Another representation of the magnetic characteristics is given by the B–H curves of the tape, as shown in Fig. 32-17. The curves show the amount of magnetic flux density created within the magnetic material by a cyclically varying intensity of applied magnetic excitation. Since the particles store part of the magnetic field, the path for increasing excitation differs from the decay path for decreasing excitation.

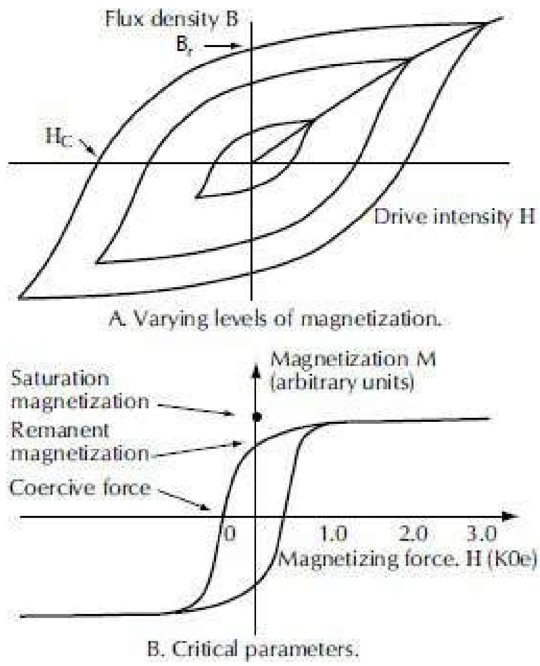


Figure 32-17. B-H curves.

32.4.3 Magnetic Tape Specifications

The performance of a magnetic tape involves many parameters such as maximum output level, distortion, noise, print through, and frequency response. As a result, the data sheet that characterizes this performance must include many operating characteristics. The user must be very careful, however, to determine the test conditions under which the data is derived, including record head gap length, tape speed, operating level, and equalization.

One form for presenting this data is shown in Table 32-3. The data entries are measured for one specific recommended bias setting. Some of the values, such as sensitivity at long and short wavelengths, are comparisons to the performance of a standardized reference tape. The notes contain important information defining the test conditions used to derive the data.

In contrast, the graphical data in Fig. 32-18 depicts how the various values change as the bias value is adjusted over a range of 16dB. All values are absolute values without any comparisons to a reference tape. The parameters of the recorder used for testing are shown above the graph.

The bias point recommended by the tape's manufacturer is the 0dB value on the bottom scale. This value is a compromise value determined by simultaneously evaluating the distortion, noise, and maximum output levels for each bias setting.

Table 32-3. Tabular Tape Specifications

	Unit	Typical Values	Test Notes
I. Electromagnetic Properties			
Recommended Bias Setting	dB	3.0	1
Sensitivity at 1 kHz (81 kHz)	dB	0.8	2
Sensitivity at 10 kHz (Sl 0 kHz)	dB	1.1	2
10 kHz Saturated Output (SAT, 0 kHz)	dB	18.5	13
Third Harmonic Distortion at Reference Level (THD)	%	0.06	3
Output Level at 3% Third Harmonic			
Distortion (1 kHz) (MOLL kHz)	dB	17.5	4
Weighted SNR			
a. related to reference level	dB	–58.0	5
b. related to output level at 3% third harmonic distortion	dB	–75.5	5
Modulation Noise Ratio	dB	–73.0	6
Print through	dB	–58.0	7
II. Magnetic Properties			
Coercivity (Hci)	Oe (kA/m)	350 (28)	8
Retentivity (Brs)	Gs (mT)	1500 (150)	8
III. Physical Properties			
Thickeners:			
Oxide Coating	mils	0.690	9
Backcoating	mils	0.040	9
Base	mils	1.400	9
Total	mils	2.130	9
Standard Widths:			
¼ inch	inches	0.246	
½ inch	inches	0.496	
2 inches	inches	1.996	
Width Tolerance:			
¼ inch	inches	+0.001	
½ inch	inches	+0.002	
2 inches	inches	–0.000	
Tensile:			
Yield Strength	lbs/qtr in	5.8	10
Breaking Strength	lbs/qtr in	11.6	11
Backcoating Resistivity	ohms/sq	5×10^4	12

	Unit	Typical Values	Test Notes
IV. Measuring Conditions			
Tape Speed	in/s	15	
Reference Level	nWb/m	320	
Record Head: Gap Length	mils	0.50	
Track Width	mils	70	
Reproduce Head: Gap Length	mils	0.25	
Track Width	mils	70	
Reproduce Equalization	Ps	50 +3180	
Record Equalization		none	

Test Notes

1. Recommended bias setting is determined by adjusting the bias current for maximum sensitivity at 10 kHz and then increasing the bias until the sensitivity changes by 3.0 dB. The adjustments made with a playback reference approximately 20 dB below reference level. The recommended bias setting corresponds to low third harmonic distortion and high output at 1 kHz.

2. Sensitivity is a measure of the output level compared to a standard reference tape A342D, when the recording is made at a constant input voltage approximately 20 dB below reference level and at the recommended bias setting.

3. Third harmonic distortion is the ratio between the level of the third order harmonic and the fundamental frequency (1kHz) expressed in percent when recorded at reference level and at the recommended bias setting.

4. Output level at 3% third harmonic distortion is a measure of the output level capabilities of a tape at 1kHz when recorded at 3% third harmonic distortion and at the recommended bias setting.

5. Weighted signal-to-noise ratio is defined as the ratio in dB between the 1kHz output at reference level or at 3% third harmonic distortion and the ASA weighted (NAB standard) noise level. The noise measurement is made with the recommended bias and without input signal.

6. Modulation noise ratio is defined as the difference in amplitude between a 1.0kHz signal level and its noise skirt at 800Hz with a bandwidth of 10Hz. The recording is made at reference level and the recommended bias.

7. Print through is the level of the accidental printing effect due to a signal recorded on an adjacent layer of tape. The printing signal is recorded at 1kHz at reference level and the tape is held at 70° F for 24 hours.

8. Coercivity is the magnetic field required to reduce the magnetization of a saturated magnetic specimen to zero. The coercivity is a direct measure of the bias current requirement of a tape. Retentivity is the maximum remnant magnetization possible in a magnetic material. The long wavelength saturated output is directly proportional to the retentivity. Coercivity and retentivity values are obtained from a 60Hz B-H loop tester with 1000 Oersted field calibrated to that maintained by the National Bureau of Standards.

9. Thickness measurements are made on Standard Gauges, 8000 Series, Smart Box.

10. Yield strength is defined as the force that produces 3% elongation of the samples. The measurement is made on an Instron tensile tester at a jaw separation of 5in and a cross-head speed of 2in/min.

11. Breaking strength is the ultimate tensile strength indicating the force at which the tape breaks and is measured on an Instron tensile tester at a jaw separation of 5in and a cross-head speed of 2in/min.

12. Backcoating resistivity relates to the tendency of magnetic tape to retain static charge. A resistivity value of 5×10^4 ohms per square is sufficiently low to prevent static buildup which might result in tape damage on high-speed bin loop duplicating systems or in normal use at low humidity conditions.

13. See bias curves.

Specifications are subject to change without notice.

32.4.4 Setting Bias Level

There are two common methods used for setting the bias level. One technique is to adjust the bias while recording a long wavelength such as 1kHz. The bias is increased until the recorded signal peaks. The bias level is then further increased until the recorded signal drops by 0.5dB.

A second technique is to use a short wavelength, typically 1.5mils, and adjust for a significant amount of overbias. The bias is increased until the recorded signal peaks. The bias is then further increased until the recorded signal decreases from peak by several dB.

How do these two techniques compare? Find the sensitivity curves S_1 and S_{10} near the center of the graph. These curves show how the 1kHz and 10kHz signals will change in level as the bias is increased. Note that the S_1 curve is very flat, changing only ¼dB from peak over a bias range of 5dB. In comparison, the S_{10} curve is falling at a rate of approximately 1dB/dB of bias increase.

The flat shape of the S_1 curve provides very little signal drop for a rather large bias change. A 0.1dB error in the signal level adjustment, perhaps due to a sticky meter, may change the 10kHz sensitivity by 2dB or 3dB. This error would require an additional record equalization boost of 2–3dB to correct the overall response.

In contrast, the rapid signal level change when using a 10kHz signal gives a much more precise adjustment and better uniformity from track to track. It is clear that both techniques are trying to achieve the same adjustment, but the short wavelength technique offers much better resolution.

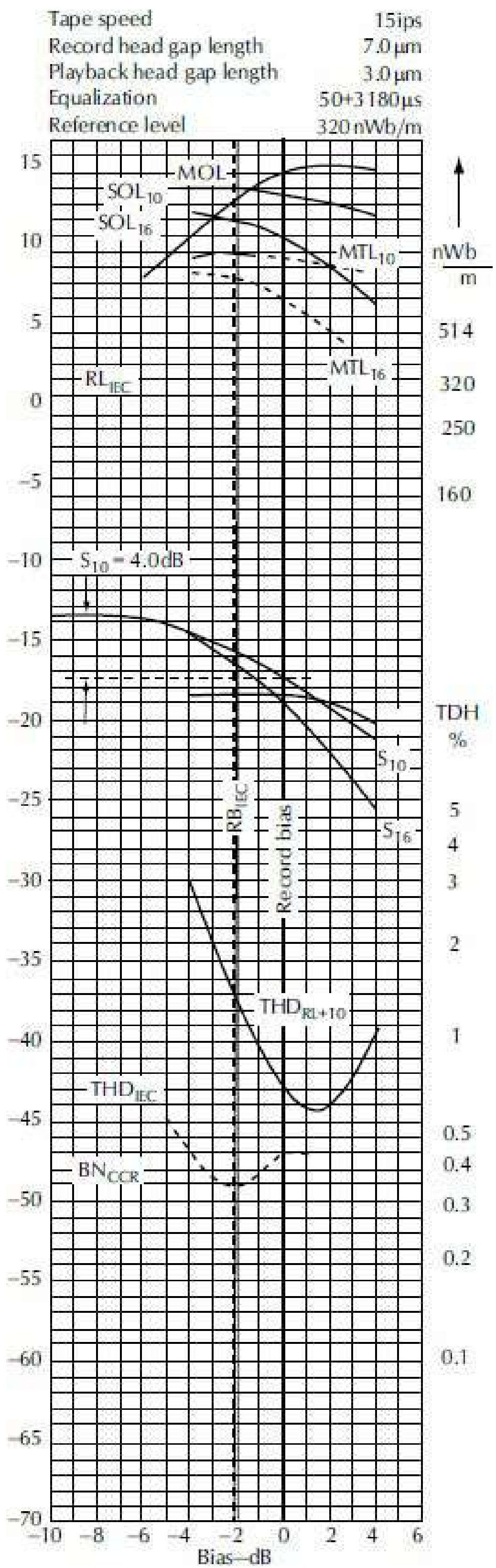


Figure 32-18. Tape performance data in graphical form.

This technique can be a trap for those who don't understand what is actually happening. The S_1 and S_{10} curves are really curves for the specific wavelengths of 15mils and 1.5mils, respectively. If the tape speed is doubled, these curves will now represent performance at 2kHz and 20kHz. The 15in/s overbias specifications for 10kHz must not be used at any other speed! For example, at 30in/s the S_{10} curve of the example tape has a downward slope of only 0.5dB/dB of bias. Why? Because the wavelength is 3mils, not the 1.5mils of the previous example at 15in/s. The manufacturer recommends only 1.5dB of overbias at 10kHz and 15in/s. It is important to use the same wavelength at all speeds by shifting the test frequency to 20kHz or 5kHz at 30in/s and 7.5in/s, respectively.

As mentioned previously, the test data is very dependent on the characteristics of the recorder used during the testing. In particular, the shape of the S_{10} curve varies greatly with changes in the record head gap length. Table 32-4 illustrates how this gap length affects the recommended amount of overbias.

Table 32-4. Short-Wavelength Dependence on Record Gap Length

Record Gap Length	10kHz Overbias @15in/s
1.0mil	1.0dB
0.5mil	2.5dB
0.25mil	3.0dB

32.4.5 Problems with Older Tapes

The archives of American tape recordings contain tapes that are up to 50 years old. Unfortunately, many of these tapes have problems that could easily damage or destroy their recordings. Some of these problems can be corrected, but others are irreparable.

32.4.5.1 Adhesion and Peeling Oxide

Adhesion is the binding force that firmly holds the oxide layer onto the plastic substrate. Two simple tests can be used to evaluate the strength of the adhesion—the Scotch tape test and the sharp edge test.

The Scotch tape test tries to rip the oxide from the plastic substrate by brute force. Start with a strip of Scotch Brand Magic Mending tape several inches long. Adhere about 3 inches of the sticky tape to the oxide surface of the recording tape. Rub the joint to assure complete binding of the tapes. The test is to peel back the sticky tape with a quick jerk parallel to the tape. If the sticky tape comes off cleanly, the adhesion is good. If the oxide layer delaminates and peels off with the sticky tape, the adhesion is poor.

The second adhesion test utilizes a blunt edge to create a very sharp bend in the tape. Find a sharp perpendicular edge on a desk or a piece of plastic ruler that has no rounding radius. Place the backside of the tape sample against the sharp edge. Pull on the ends of the tape to establish a firm tension in the tape. While maintaining the tension, drag the tape over the edge in a 90° bend. If the oxide does not loosen from the backing, the tape passes the test. A poorly adhered tape may suffer complete delamination of the oxide, with a solid band of oxide peeling off and shooting away from the backing.

32.4.5.2 Brittleness

The polyester base films and urethane binders of modern tapes remain flexible under all normal circumstances. The base films and oxide layers of earlier tapes, however, could become brittle. Plasticizers were included in the binders and the acetate backing to provide flexibility in the tape. Unfortunately, these plasticizers can harden with age, causing the tape to become brittle. Harsh storage conditions can accelerate the breakdown of the plasticizers.

Brittleness cannot be reversed. The only remedy is to use a tape transport that is extremely gentle on the tape. Choose a transport with dynamic braking rather than harsh mechanical brakes. A transport with constant tape tension can be set for the lowest practical tape tension. Some decks also feature gentle start capability that ramps the capstan up to speed to smoothly accelerate the tape rather than just slamming a pinch roller onto a running capstan.

32.4.5.3 Splice Failures

In the early days, standard Scotch Brand cellophane tape was the only splicing tape. Later on, splicing tapes such as Scotch #41 and #620 were developed with improved characteristics. Although these tapes were fine for day-to-day operations, they have not survived the test of 50 years of storage. For example, the adhesive of both cellophane tape and splicing tape can ooze out and stick to adjacent layers of tape. A common remedy was to apply talcum powder to the sticky oozed adhesive to avoid layer-to-layer adhesion.

With even more time the adhesive can dry out completely, causing the splice to fail. In this case the only remedy is to replace the splicing tape with new tape. The newest splicing tapes, such as blue #67, replace the original latex adhesives with synthetic adhesives that do not ooze or dry out.

The tape operator must be watchful for two problems created by bad splices. The first is layer-to-layer adhesion that can produce strong tugs that break older acetate tapes. Second is tape separation at a failed splice. DO NOT run the tape through the recorder at high speed if you suspect either problem may exist. If the tape tugs or separates at high speed, the loose end may be slapped around and broken off before you can stop the spinning reels.

If the tape is old, wind the tape slowly and carefully to examine each splice. Re-splice all splices if there is any hint that the splices may separate. Do not try to remove any of the old splicing tape adhesive that has dried out on the recording tape. Sticky adhesive residue that could bond to adjacent layers, on the other hand, must be removed.

32.4.5.4 Print Through

The energy required to activate a particle to switch magnetic states depends on the size of the particle, with the overall characteristics of a magnetic tape being determined by the average size and characteristics of many particles in the coating. A more detailed analysis of the particles would yield a distribution of sizes, as shown in Fig. 32-19. Although the majority of the particles cluster around the average value, a small portion of the particles are either much smaller or much larger than the average. The small particles give rise to spontaneous recording as print through; the large particles produce noise bursts.

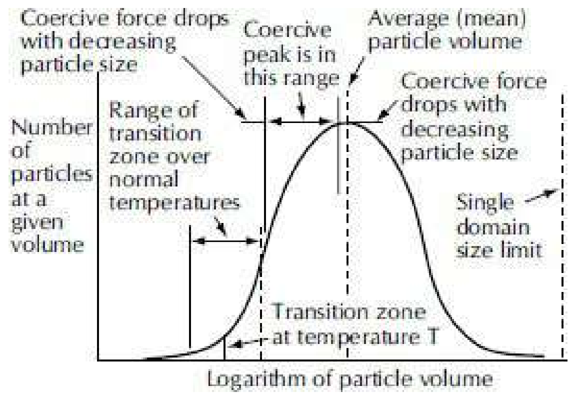


Figure 32-19. Particle size distribution.

The small particles require so little activation energy to assume a new magnetization state that even the thermal energy of the particles may provide enough bias to cause the particles to be recorded by the stray magnetic fields due to adjacent layers of recorded material. This spontaneous recording is most evident as pre-or post-echo at the beginning and end of a recording. The strength of the print through image depends on both the percentage of thermal idiots in the coating and the ratio of remanence to coercivity of the tape. The remanence measures the driving force of the signals trying to print through. The coercivity, on the other hand, is the stubbornness of the particles to resist this imprinting. The effective coercivity of the small particles is diminished because the domain size is sub-optimal, rendering the small particles more susceptible to printing.

The milling process used to provide thorough mixing of the particles, binder, and additives prior to coating is a rather abusive process that can create thermal idiots by fracturing some of the desirable large particles into smaller, low-coercivity particles. Insufficient milling, on the other hand, provides an uneven particle dispersion that creates noise on the tape. The tape manufacturer must strike a compromise that yields both low noise and low print through.

Print through of a signal produces both pre-echoes and post-echoes. The pre-echoes are more troublesome in music, however, since the pre-echoes frequently occur in the quiet passages just before the loud note. The post-echoes, on the other hand, are frequently masked by the diminishing tail of the musical note and the room reverberation.

Fortunately, the print-through process does not produce equal amounts of pre- and post-echo, but unfortunately the more undesirable pre-print echo is the stronger. The vector magnetization components that arise during the recording process cause the levels of print on the outer adjacent tape layer to be several decibels higher than on the inner adjacent tape layer, as shown in Fig. 32-20. The more troublesome pre-print echoes on musical selections can therefore be minimized by storing the tape tails out to move the quiet lead-in to the inner layer. This will also bury the louder outer layer print through echo in the decaying signal at the end of the music.

The use of nonmagnetic leader tape between selections is also helpful to eliminate pre-echo on selections that begin with a rapid attack. Be aware, however, that paper leader tape can contain a small amount of magnetic debris that will raise the noise level as the leader passes over the playback head.

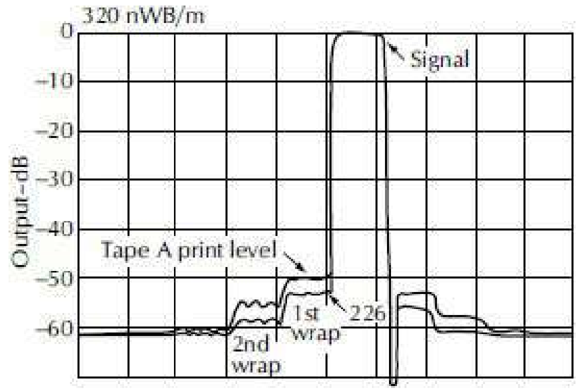


Figure 32-20. Pre- and post-echo print through. Courtesy 3M Co., Magnetic Audio/Video Products Div.

The user can take several steps that will minimize the amount of print through. First, the use of thicker base films increases the spacing between layers. Second, avoiding elevated temperatures and stray magnetic fields during use and storage will decrease the excitation of the thermal idiots. Third, exercising the tape by shuttling the tape from reel to reel several times will partially erase the printed particles. The flexing and rubbing of the tape produce enough activation energy to neutralize some of the printing. For this reason, never copy a stored master tape without exercise. The worst possible print through level exists on the very first pass of the tape. In some cases print through can drop as much as 4–6dB with five shuttle cycles.

32.4.5.5 Sticky Shed and Tape Baking

As mentioned earlier, the goal is to attach a maximum number of perfectly stacked and oriented magnetic particles onto the surface of the backing material. Anything that interferes with this goal by displacing some of the magnetic particles, such as additives for lubrication, fungicides, and static charge reduction, degrade the performance of the tape. Most important in this category is the very binder that holds the particles in place. Every bit of binder displaces some of the useful magnetic particles.

The best choice is to use a very strong binder that can do the job with the minimum amount of glue, allowing space for more magnetic particles. The winner is the highly crosslinked thermoset polymers in use today. Starting around 1970, these binders with a high urethane content gave a big boost to tape performance. Unfortunately, long term experience with these tapes now shows that the binder can break down during storage. The symptoms are a buildup of residue on the head and guides and a tendency for the tape to stick to these residue buildups, in some cases actually dragging the tape to a stop. The popular name for this problem is sticky shed. The problem is usually discussed in terms of binder breakdown, but there appears to be a second major problem related to lubricant oozing that is also present.

Urethane Binder Breakdown. The urethane binder contains long polymer chains that provide the high strength of the binder. Water in the surrounding air enters the tape and breaks the long chains through a process known as *hydrolysis*.

As a result of the chemical breakdown of the long polymer chains, the binder is weakened enough for the surface of the tape to begin to rub off onto the stationary guides and heads. Depending on the design of the transport, this residue can clog the heads in just a few seconds. Machines with rotating guides and low tape tensions take longer to clog, but the damage to the tape is still intolerable.

Fortunately, the hydrolysis is somewhat reversible. Tapes can be *baked* at a moderate temperature to reverse the hydrolysis and restore strength to the binder. Although this may seem like a bit of witchcraft, thousands of baked rolls of archived tapes have proven the technique.

The electric oven must provide a well-controlled temperature of about 120–140°F (50–60°C). Large dehydrators or fruit dryers are popular because of their size and limited temperature range. Only an electric oven should be used. The oven should be preheated and checked for temperature stability with a high-resolution thermometer such as a candy thermometer. The tape, wound onto a metal reel, is placed into the oven horizontally with generous space above, below, and around the reel for air circulation. The tape is baked for 15–20 hours, and then allowed to cool to room temperature undisturbed in the oven.

The baking process creates a low-humidity environment that draws some of the excess water from the tape binder. The short polymer chains may recombine with their neighbors to produce a better bond, but the breakdown process is not fully reversed.

Lubricant Oozing. The second failure mechanism also involves the binder, but in this case the culprit is the oxide. A change of oxide particles also changes the chemistry needed to make a liquid binder that can:

1. Hold all the magnetic particles in suspension.
2. Be smoothly coated onto a polyester backing material.
3. Have the volatile byproducts evaporated in the drying ovens.

In the early 1970s Phizer introduced a new high-output oxide with excellent signal characteristics, but the particle required a reformulated binder with a low pH in order to meet the above requirements for a usable dispersion. This particle was utilized by

3M in the 226 family of tapes (226, 227, 806, 807, 808, 809) and by Ampex in the 456 family.

The new binder formula included a component that served primarily as a lubricant. Unfortunately, however, this lubricant would migrate to the surface of the tape and concentrate into a sticky residue.

The baking operation described in the previous section warms the concentrated lubricant enough to allow the lubricant to flow and be reabsorbed into the depths of the coating.

Since both types of sticky shed problems are treated by baking, most people who bake tapes don't know for certain which problem they are treating, and if the sticky shed is eliminated, they probably don't care.

How long before a baked tape begins to again exhibit sticky shed? Results will vary depending on the amount of degradation, the tape type and specific batch, the exact baking method, and the operating environment after baking. Reports vary from days to years. Certainly baking provides an adequate window for the tape to be transferred to another medium.

Is there any degradation due to the baking process? The most likely problem is print through caused by the elevated temperature. Print through is a time-dependent problem that peaks out at a maximum value after a long time. Heat accelerates the printing. However, stored tape probably has had enough time for the print through to be near the maximum value before baking. As a result, the additional print through caused by the baking may be negligible.

How can sticky shed be avoided? The rate of hydrolysis depends on the storage conditions. Archival storage at a temperature of 60°F

(15°C) and relative humidity of 25% ± 5% is optimal, but few have the luxury of an environmental chamber, so store the tapes in a cool, dry location in the original package standing on edge.

Sticky shed may also produce layer-to-layer adhesion. If you strongly suspect sticky shed, bake the tape before trying any winding operations on a tape transport. This will avoid the total loss of recorded segments due to oxide being ripped from the tape's plastic backing during spooling.

32.4.5.6 Squealing Tape

One of the many ingredients in the coating recipe is a small amount of lubricant. Obviously, the tape cannot be too slippery or else the capstan couldn't maintain constant tape speed. Running the tape completely dry, on the other hand, can produce an audible squeal. The tape undergoes a "stick-slip" phenomenon on the stationary guides and head, creating a jerky motion at a high frequency. The irregular motion can even be measured with a scrape flutter meter.

The squeal results from the loss or failure of the original lubricant. The obvious solution is to replace the lubricant. A can of 10-W30 motor oil isn't appropriate, but another common household lubricant, WD-40, is recommended by Quantegy. Quantegy claims that WD-40 is cheap, available, inert to all recorder components, and a very good lubricant. Apply the oil sparingly by lightly wetting a lintless rag or swab with the oil and holding the applicator against the oxide side of the moving tape at the first guide after the supply reel. A bit of experimenting may be required to find the proper amount of oil that is required to eliminate the squeal without causing slippage and speed irregularities. When you are finished using the tape, prepare the tape for storage by passing the tape over a dry applicator in a medium speed fast wind mode. (Use extreme caution on tape transports that have elastomer coatings on the capstan and/or timing rollers. Lightly lubricate the tape while passing the tape directly from the supply reel to the takeup reel, and then dry the tape with a second pass over a dry applicator before threading the tape over the elastomer components.)

32.5 Analog Circuits and Systems

The transport mechanism, heads, and tape should combine to determine the basic performance limitations of a tape recorder. The analog electronic circuits of the recorder, on the other hand, should exceed the capabilities of the heads and tape in all respects so that only the heads and tape limit the quality of the final signal. In practical terms, this means that the *SNR*, frequency response, distortion, and head room of the electronics are comfortably better than the heads and any tape, including future improved tapes.

The block diagram of the signal electronics of a typical professional audio recorder is shown in Fig. 32-21. In terms of actual hardware, approximately 75% of the audio circuits of a modern professional audio recorder are devoted to operator interfacing and controls; the remaining 25% implement the basic functions of erasing, recording, and playback. Since the variation of features and technology used to implement the interfacing and control functions is too broad to be summarized herein, the following description covers only the latter basic functions.

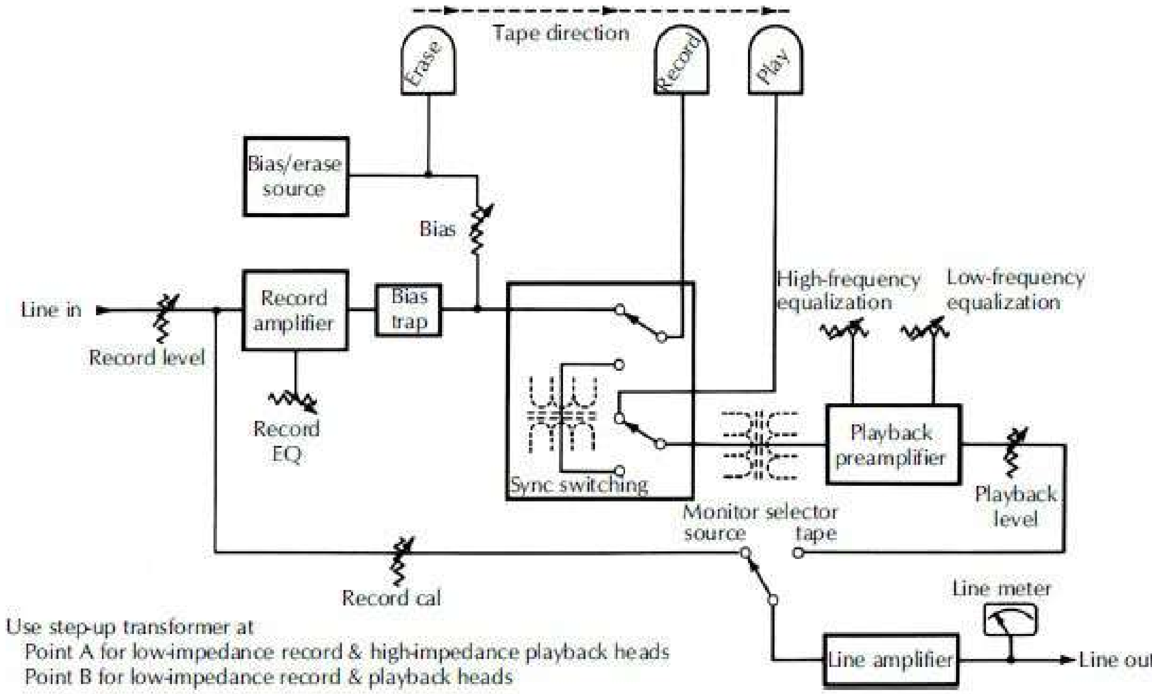


Figure 32-21. Tape recorder signal block diagram.

32.5.1 Playback Amplifiers

The amount of electrical power that can be generated by a magnetic tape passing over the face of a reproduce head is exceedingly small. The output voltage from the head for loud recorded passages will reach no more than a few millivolts, with quiet passages dropping into the microvolt region. This weak signal must be carefully boosted without the introduction of additional noise to a higher, more usable level by the first stage of the playback amplifier. Special low-noise amplifier circuits developed for this purpose provide at least 20dB of gain so that subsequent amplifier stages will not be required to operate near their noise limits.

Since the reproduce head produces an output voltage that is related to the rate of change of the flux on the tape, $d\phi/dt$, the output voltage will rise at a rate of 6dB/octave. A compensating circuit with a falling 6dB/octave response, known as an integrator, is used in the playback amplifier to correct for this rise and give a voltage that is proportional to the value of flux sensed by the head.

When the effects of playback head resonance peaking and gap length, spacing, eddy current, and thickness losses are included, the output from the low-noise amplifier and integrator would follow the falling curve in Fig. 32-22 for 15in/s (38cm/s) operation. This curve must be reshaped by the combined effect of the record and playback equalizers to yield a flat response. The method of partitioning this correction between the record and playback circuits is dictated by the equalization standard chosen by the operator. Since all users of a given equalization standard will be using the same partitioning, the recorded tapes will all be interchangeable.

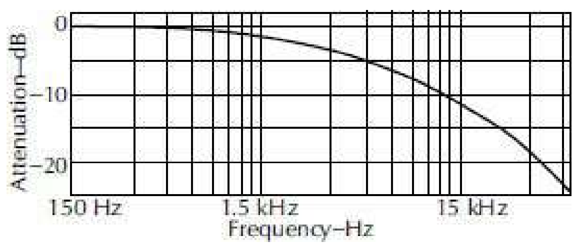


Figure 32-22. Unequalized reproduce head output.

Fig. 32-23 shows a simplified schematic of a typical operational amplifier type of a playback amplifier capable of the necessary playback corrections. The low-frequency-cut circuit is utilized in some NAB and cassette standards to achieve a decrease in low-frequency playback noise below 100Hz at the expense of low-frequency headroom. A typical design would include additional ancillary components for amplifier biasing and stabilization.

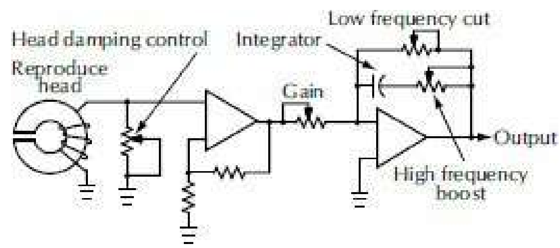


Figure 32-23. Simplified playback amplifier.

With one common exception, the same type of circuitry is utilized in the sync/overdub mode to condition and amplify the playback signal from the record head. The exception is in the form of an added voltage-boosting transformer that is commonly necessary to get the signal above the noise floor of the low-noise input section. This problem arises from the low inductance and few turns of wire that are typically found in a record head. The record head must pass the audio plus the high-frequency bias signal; therefore, the inductance must be kept low enough to avoid self-resonance with the head cables at the bias frequency. When fewer turns are used to reduce the inductance, the output voltage goes down proportionately. In essence, these turns are restored in the transformer by a step-up turns ratio ranging from 3:1 to as high as 10:1.

32.5.2 Record Circuits

The primary task of the amplifier that drives the record head is to convert the input audio signal voltage into a proportional amount of current flowing in the windings of the record head. To accomplish this task, the head driver must overcome the rise in head impedance with frequency that is due to the inductance of the head. A common technique to achieve flat current response, as shown in Fig. 32-24A, is to insert a resistor in series with the head so that the combined series impedance of the resistor and the head remain relatively constant throughout the audio band. If the resistance is chosen to be two to three times the reactance of the head at the upper limit of the desired audio band, the desired constant current characteristics can be closely approximated.

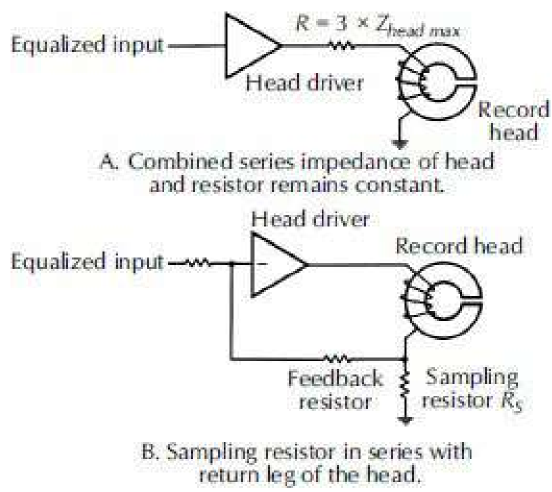


Figure 32-24. Constant-current record head drivers.

The primary disadvantage of the series resistor is the loss of head room due to the extra signal drop across the resistor. This problem can be overcome with an active current feedback circuit that senses the current in the head through a small sampling resistor. Fig. 32-24B shows a sampling resistor R_S in series with the return leg of the head. The voltage generated across R_S by the current flowing in the head is fed back to the inverting input for comparison with the incoming audio signal. The high gain of the driver amplifier necessitates only a very small feedback signal, creating a negligible loss in head room at high frequencies.

The circuits of Fig. 32-24 oversimplify the task of driving the record head since no provisions are included for adding the high-frequency bias signal to the current in the head. A common method of adding the bias and audio signals is shown in Fig. 32-25. The audio driver is isolated from the bias source by a parallel trap tuned to the bias frequency so that the bias signal does not create nonlinearities within the audio driver. The high impedance of the trap at the bias frequency also reduces the loading effect of the audio source on the bias source.

A similar isolation of the bias source is accomplished by the capacitor in series with the bias supply. Since the capacitor looks like a high impedance at audio frequencies, the loading effect of the bias supply on the audio source is minimized. At the higher frequencies of the bias signal, the reactance of the capacitor has dropped to a relatively low value that provides adequate coupling of the bias signal into the record head.

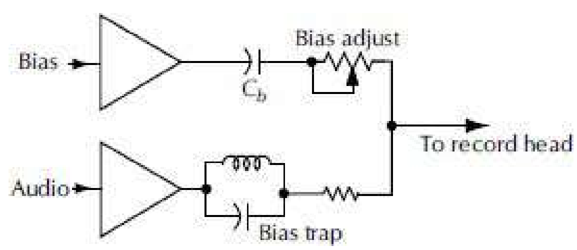


Figure 32-25. Audio and bias coupling to record head.

An alternate approach that eliminates some of the previously mentioned isolation requirements is shown in Fig. 32-26. In this case, the bias and the audio are added together at the input to a combination bias/audio head driver amplifier. If the amplifier has sufficient head room and very low distortion, the two signals can be amplified simultaneously by the same amplifier without any interference. The problem of coupling the output to the head for constant current drive must still be overcome, however, by including either a complex coupling network or an active feedback network.

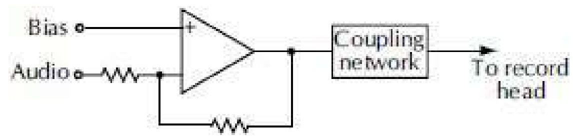


Figure 32-26. Active summer for bias and audio.

In addition to the head driver circuits, which correct for any response droop due to head inductance, the record amplifier must provide deliberate frequency response tailoring to match the desired equalization standards. The standards usually require an adjustable boost at 6dB/octave beginning in the middle of the audio band, with lower tape speeds generally requiring greater boosts to overcome the increased tape thickness and self-erasure losses.

The needed boost is easily implemented by the resistance-capacitance circuit shown in Fig. 32-27A, but the use of a variable capacitor is inconvenient due to the limited range of capacitor adjustment and the awkward size and mounting of the capacitor. Newer designs, therefore, favor operational amplifier configurations that control the amount of boost with a potentiometer. One such circuit, shown in Fig. 32-27B, selectively adds the output of a differentiator circuit, which rises at 6dB/octave, to the main signal path.

A secondary benefit of the differentiator circuit is the phase change introduced by the inverting characteristics of the differentiator amplifier. Unlike most of the loss-correction circuits of the signal path, which introduce signal delay at high frequencies, the inverted differentiator advances the high frequencies. The proper combination of advance and delay can provide less phase distortion in the signal, yielding improved transient response with less overshoot. A similar phase-correcting effect has been implemented in other designs by providing an all-pass, phase-shifting network in the reproduce amplifier.

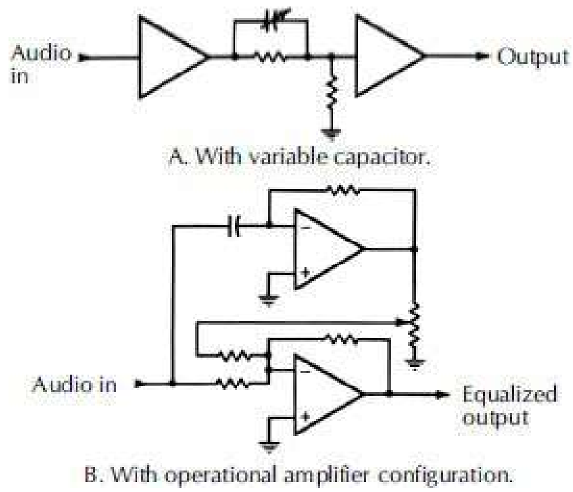


Figure 32-27. High-frequency record boost circuits.

The NAB and compact cassette equalization standards provide an additional record signal boost at low frequencies to overcome the hum and noise limitations of the reproduce heads and amplifiers. Typical circuits for this purpose are shown in Fig. 32-28. Both cases achieve a 6dB/octave rise with decreasing frequency from 50Hz or 100Hz to below 20Hz.

Abrupt changes in the bias and audio signals on the record head must be avoided whenever the record mode is entered or exited. Ramping circuits are employed for this purpose to control the buildup and decay of these signals. Typical methods include the use of analog switching elements such as bipolar-junction or field-effect transistors. The rate of switching of these elements is limited to a value that does not create abrupt transients but, at the same time, is quick enough to avoid annoying delays, overrecordings, or program holes.

32.5.3 Bias and Erase Circuits

The high-frequency signals required for biasing and erasing all tracks of the tape are derived from a single master oscillator so that no interference or beating of multiple oscillators will occur. Older designs generally employ a tuned push-pull multivibrator oscillator; newer designs favor crystal-stabilized oscillators utilizing digital circuitry. Several designs have used separate bias and erase frequencies, with the erase circuit running at one-third the bias frequency to minimize the power dissipation on the erase head.

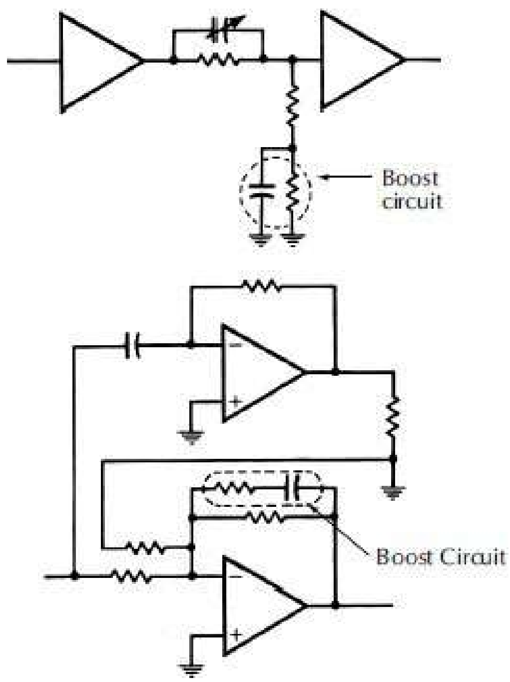


Figure 32-28. Low-frequency record boost circuits.

In all cases, the primary consideration is purity of the bias and erase current waveforms. Any even-order harmonics, including dc, second harmonics, fourth harmonics, and so on, will create a detrimental rise in the background tape noise, reducing the available *SNR* for the recorder. Older designs, such as Fig. 32-29A, relied heavily on push-pull circuits with balancing transformers to minimize these even-order components. Newer designs, such as Fig. 32-29B, favor filtering and feedback control to reduce unwanted components. The divide-by-two flip-flop eliminates any even-order distortion in the oscillator waveform.

The erase head is typically coupled to the erase source with an adjustable series resonating capacitor to minimize the voltage required from the driver and to filter out even-order components. A current sampling resistor is frequently provided in the ground leg of the erase head circuit so that the amplitude of the erase current can be conveniently monitored.

32.5.4 Sync Operation

Multitrack recording requires that artists be able to listen to the previously recorded tracks while simultaneously adding their new performance in synchronism with the prior tracks. Analog recorders accomplish this by using some tracks of the record head as playback sources while simultaneously recording on other tracks of the same head.

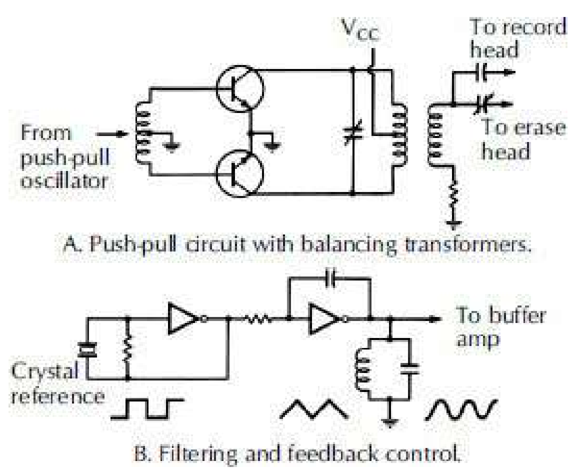


Figure 32-29. Typical bias and erase sources.

32.6 Tape Recorder Transport, Maintenance, and Testing

Maintenance begins with inspection and cleaning. Before starting the cleaning procedure, note the location and type of dirt and debris that has accumulated due to prior use. Excessive debris indicates that your recording tape is being slowly destroyed by the tape transport.

A deposit of very fine, silky threads indicates that the polyester base film of the tape is being scraped off by a sharp edge on a guide flange. Examine all edge guides for grooves cut into the flanges by the tape. Either reposition the guide to place an unworn surface in contact with the tape or install a new guide if the groove is severe.

Deposits of brown or black dust near the guides indicate that the edges of the tape are being scraped or deformed enough to break small chunks of coating from the edge of the tape. Check the tape tension and the height of the guides and reel hubs.

Any caked-on deposits on the surface of the guides or heads are very serious. Inspect the surface of the tape for scratch marks. If the tape surface is being scratched, continued use will destroy the tape. Correct the cause of the scratches before continuing.

Several types of cleaners are available for cleaning tape machines. Older head cleaners usually contained Xylol, a strong solvent, to aggressively dissolve tape residue. Milder isopropyl alcohol is a more popular solvent today, but avoid rubbing alcohol containing 30% water in favor of the 99% pure variety for topical antimicrobial use.

Use a soft swab moistened with cleaner to scrub the contact surfaces of the heads, guides, and capstans. Avoid drenching the swab. If the swab is too wet, solvent may run down the capstan shaft into the top bearing, washing away the bearing's lubrication. Cotton swabs are suitable for most analog tape recorders but not for the delicate heads on a helical scan recorder. Use special lint-free swabs with more pliable sticks for cleaning rotary head machines.

When cleaning the head, always rub the swab in the direction of tape motion, never across the head sideways. Sideways scrubbing may peel away the edge laminations of the cores. Avoid scraping the face of the head with the stick or core of the swab. Allow adequate time, typically 30seconds, for the solvent to evaporate before rethreading the tape. You don't want the leftover solvent dissolving your recording tape.

Xylene head cleaning solvents will attack some plastics including the lenses of optical sensors. Aggressive solvents may either partially dissolve or create a hard, glazed surface on some rubber rollers. If you notice a lot of reside on your swab or rag after wiping a seemingly clean roller, you are probably dissolving the roller, not cleaning it! Use general-purpose cleaners for the plastic components and rubber cleaners for the rubber rollers.

The tape must also be kept completely free of dirt. Keep the surface of the transport clean to avoid dirt being picked up during high-speed spooling. Always return the tape to its storage carton between uses. Do not stick your fingers through the windage cutouts in the flanges of the reel and touch the edges of the tape pack when handling the reel. (Skin debris from fingers is a source of tape dropouts!) In addition:

1. Avoid eating greasy foods while handling tapes.
2. Contamination due to finger oils and debris can be avoided

during editing sessions by wearing lint-free editing gloves, which are available at most camera supply stores.

3. Keep cigarette ashes and other powdery materials far away from the tape.

The cooling system of the tape recorder should be cleaned periodically. Clean all air filters and cooling passageways and remove any dust buildup with a vacuum cleaner. Verify that all inlet or exhaust ports on the bottom of the machine are not obstructed by carpeting or dust and that adequate clearance for free airflow exists at the rear of the machine.

Following cleaning, diagnostic servicing should begin with verification that the tape guiding and tension at the heads is adequate to maintain good tape-to-head contact. Set aside one reel of tape, known as a *shop tape* because it typically comes from the maintenance shop, for testing. Run this tape in all modes while observing the tape at the heads and the guides. The tape should not run hard against either guide flange and there should never be any edge distortion. If edge distortion is noted, check for a bent guide or tension sensor arm. These components can be easily bumped out of alignment by a full reel of tape during loading or unloading.

On many machines a tape tension gauge of the type shown in Fig. 32-30 can be inserted in the tape path near the heads to measure the tension. For other machines that are too crowded in the head area, either the head assembly must be removed or a test location away from the heads must be used. Measure the tension at both the beginning and the end of the reel.

Figure 32-30. Tension measurement. Courtesy Ampex Corp.

Note that the stiffness of a piece of tape varies with the width, base film thickness, and type of tape. The tension gauge must be adjusted before use to read correctly for the specific tape sample being used on the transport. A calibration weight is included with the gauge for this purpose.

The following tape tension values indicate the range of tensions commonly encountered on studio recorders.

¼ inch	[illegible]
½ inch	[illegible]
1 inch	6
2 inch	10

The nominal value for a given model of recorder will be found in the maintenance manual for the machine.

Some manufacturers specify tension measurements with a spring scale and a cord that is wrapped around a tape hub. Follow the recommended procedure.

Verify that the mechanical brakes or dynamic braking logic is stopping the tape smoothly from all modes and speeds without excessive force. A sticky brake solenoid or dirty brake band can quickly ruin your precious tape.

32.6.1 Speed

Absolute tape speed is extremely difficult to measure, even under the controlled conditions of a laboratory. One method available to maintenance personnel is to measure the frequency reproduced from a commercially available speed reference tape. The frequency read on the frequency counter must be corrected for any difference between the tape tension on the playback machine and the tension value used by the manufacturer of the tape during the recording process. A correction table is furnished with the tape for this purpose.

A more common speed test is to check speed uniformity from beginning to end of a reel of a tape. The following procedure outlines the general technique:

1. Using an oscillator that has been operating long enough to reach stable conditions, record a reference tone in the range of 1–5kHz at the head end of the tape.
2. Flip the reels so that the head end becomes the tail.
3. Using the console monitoring provisions of the console, mix the reproduced tone with the oscillator tone, listening for any major pitch differences. (If a significant error is detected, flip the reels again to verify that the oscillator has not shifted frequency.)

A more accurate version of this test is to use a frequency counter to measure the frequency at both ends of the reel. The speed error in percent is then calculated as

$$\%\ \text{speed error} = 2\left(\frac{head - tail}{head + tail}\right) \times 100\% \qquad (32\text{-}4)$$

A speed error of 6% will yield a pitch change of one half-tone step. Typical recorder specifications are in the range of 0.1–0.5%. Machines with constant tape tension will generally have the least error.

Possible causes of speed error include excessive tension variations from beginning to end of the reel, tape slippage due to a worn capstan surface or pinch roller, inadequate pinch roller pressure, and unstable capstan speed.

Assuming that tape tension has already been determined to be correct on both sides of the capstan, the next test is to check pinch roller pressure. First, inspect the pinch roller for glazing of the roller surface or excessive wear. Fig. 32-31 shows roller wear patterns that may reduce the traction between the tape and capstan.

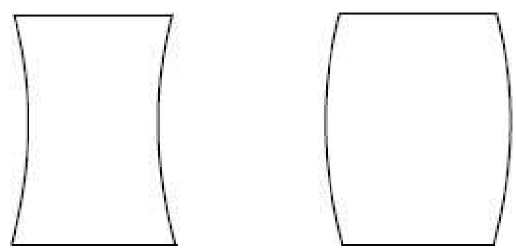


Figure 32-31. Pinch roller wear patterns.

Next, a spring scale is coupled to the top (and the bottom, if possible) of the pinch roller yoke or arm, as shown in Fig. 32-32. The scale is pulled at right angles to the support arm with just enough force to disengage the roller from the capstan. The force reading at disengagement should be compared with the recorder manufacturer's recommended value.

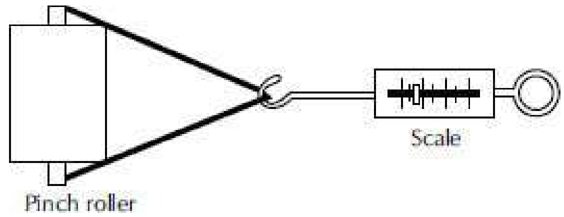


Figure 32-32. Pinch roller force measurements.

For some transports the pinch roller force is set as a fixed number of turns of a nut or screw. For this case the roller linkage is first tightened to bring the roller into light contact with the capstan, and then the recommended clamping force is applied by tightening the adjustment by the specified number of additional turns.

The surface of the capstan may become so highly polished by the abrasive action of the tape that slippage will persist for the correct values of tension and pinch roller pressure. In this case the capstan must be resurfaced by plating or sandblasting or both to restore the required traction.

In very rare cases the capstan motor may actually slow down due to excessive loading caused by bad motor bearings or high tension. Bushing bearings, which are used on many direct-drive ac synchronous capstan motors and some capstan pinch rollers are an especially noteworthy problem. Periodic lubrication of these components is essential to maintain low-friction operation. Although these components may appear to spin freely when turned by hand in an unloaded state, the friction can rise dramatically when the engagement solenoid exerts several pounds of side load on the bearings. The resulting drag and wear due to dry bearings may produce substantial speed errors. One small drop of oil can make all the difference in the world. To avoid problems, follow the manufacturer's recommended lubrication schedule.

A simple strobe light, as shown in Fig. 32-33, can be used to check the running speed of the flywheel or fan on the shaft of the synchronous capstan motors. Package the components inside a discarded plastic pen housing with the tip of the bulb protruding. Hold the light close enough to the rotating device to observe a reflection. The reflected pattern must remain stationary under all conditions of tape pack and speed. Induction motors, which do not run at synchronous speed, will always yield a moving pattern.

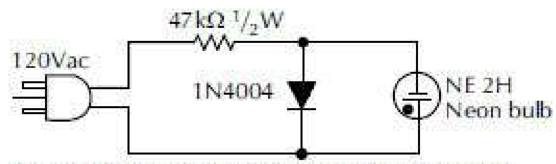


Figure 32-33. Strobe light for speed testing.

Crystal-referenced servos may falsely appear to vary in speed when tested with a strobe light if the frequency of the ac mains driving the strobe varies. An oscilloscope and frequency counter are required to properly verify correct servo operation.

32.6.2 Wow and Flutter

Speed drift represents only the very lowest frequency components of the spectrum of speed errors. Measurement of the higher-frequency flutter components requires a specialized frequency demodulating instrument called a *flutter meter*. A block diagram of a flutter meter is shown in Fig. 32-34. The reference signal from the crystal clock must pass through the record/playback process of a tape recorder before being applied to one of the phase comparator inputs. The low-pass filter and voltage-controlled oscillator simulate a large flywheel that stores the average value of the playback frequency. By applying the average value to the second phase comparator input, the phase comparator output will consist of only the short-term variations from the average speed. These variations are divided into various frequency bands for further analysis. The metering circuit provides a convenient quantitative measurement of the speed variations.

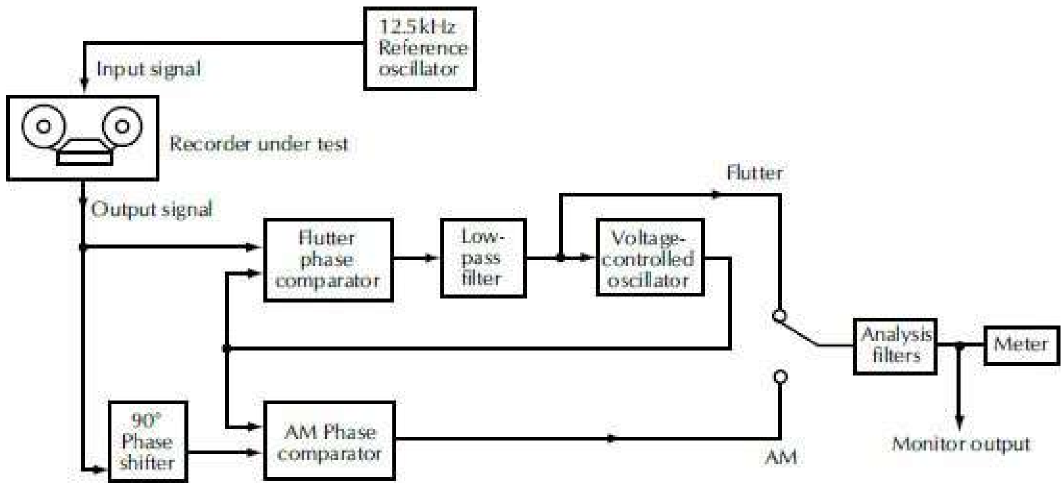


Figure 32-34. Flutter meter block diagram.

Just as the sampling rate of a digital audio system determines the highest possible audio frequency that can be encoded, the frequency of the test tone determines the range of flutter components that can be measured by any frequency demodulator. The typical upper frequency is about 0.4 times the test frequency. Due to the nature of the sidebands that are required to operate the demodulator, a typical 18kHz audio bandwidth can support a 12.5kHz test tone and a flutter bandwidth of 5kHz. This measurement technique, referred to as *high-band flutter measurement*, is supported by Audio Precision.

Unfortunately, most flutter meters use a low-frequency test tone of 3150Hz and cut off all flutter components above 250Hz, ignoring many flutter components caused by modern-day servo systems and virtually all scrape components due to the elastic vibration of the tape. To make matters worse, most flutter specifications are made through a flutter weighting filter that only measures flutter components near 4Hz. Proper maintenance requires that a broader spectrum test be implemented to check for any possible problem.

Two methods of specifying flutter performance are commonly encountered. If a flutter-free test tape is available, the flutter reading obtained in the playback mode can be reported. Most professional recorders, however, have flutter levels that are equal to or better than any available test tapes. In this case, recording and reproducing on the same machine is appropriate. The method of testing should be noted as part of the performance report.

Although test and diagnostic work is commonly conducted with simultaneous record/reproduce, the final testing should always be conducted in the reproduce-only mode. The tape should be started and stopped several times, with the various transport elements reoriented by hand between runs, to achieve a sampling of random combinations of the various record and playback flutter components. The arithmetic average of the maximum values of each sample throughout the reel, excluding any infrequent short-duration bursts, is the reported value.

If the flutter readings are excessive, the next step is to analyze the flutter waveform for information to help pinpoint which tape path component is defective. The following techniques are helpful in isolating the culprit:

1. The human ear and brain form a very versatile spectrum analyzer that frequently can immediately identify the defective component from the characteristics of the flutter signal being

reproduced in a monitor loudspeaker. Take advantage of this free portable instrument that is always at your disposal by listening to the demodulated output from the flutter meter.

2. The various selectable filters of the flutter meter can be used to isolate the general portion of the flutter spectrum in which the offending component is generating flutter.
3. The expected rotational flutter rate from a rotating component can be calculated from the diameter of the component and the tape speed using the expression

$$\%\ \text{speed error} = 2\left(\frac{head - tail}{head + tail}\right) \times 100\% \qquad (32\text{-}4)$$

where,
S_T is the tape speed,
d is the diameter of the component.

These frequencies can range from approximately 0.5Hz for the once-around of full reel of tape to 60Hz for a small-diameter capstan shaft. Some manufacturers include a table of these flutter frequencies in their maintenance manuals. The small balls and retainer clips inside the ball bearings used in many rotating components generate additional not-so-obvious flutter components at frequencies higher than the once-around rate of the bearing.

4. If the flutter is very regular, the flutter pattern displayed on the oscilloscope can be utilized to calculate the frequency of the dominant flutter component. Any flutter components caused by ac motors or power supply ripple will remain stationary on the oscilloscope screen if the sweep triggering mode is set to line.
5. A common search technique is to deliberately create flutter by

attaching a small piece of masking tape to the surface of a rotating component. The rate of the flutter blips created by the masking tape can then be compared with the unknown component to determine if the two rates are identical.

6. Note any change in the flutter spectrum when each of the auxiliary rotating components such as guides and flutter idlers is stalled. Stalling the defective component will cause the offending flutter component to cease. A notable exception to this case is the scrape flutter idler. Stalling a scrape flutter idler should usually double or triple the scrape flutter amplitude. If little or no increase is noted, the idler is not functioning properly. Check for dirty or damaged bearings that would keep the idler from spinning freely.

The following procedure describes a flutter test using a wide-bandwidth flutter meter, such as is shown in Fig. 32-35. The general technique also applies to other meters.

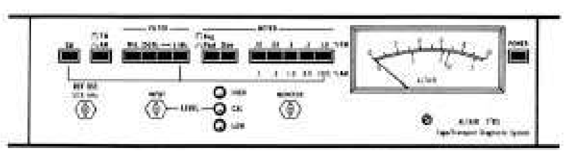


Figure 32-35. Flutter meter. Courtesy MANCO.

1. Connect the reference oscillator output (REF OSC) to the line input of the tape recorder.
2. Connect the demodulator input (INPUT) to the line output of the tape recorder.
3. Connect the demodulated output (MONITOR) to an oscilloscope and an audio monitor.
4. With the tape machine in the record mode, set the recorder's input level control to achieve a playback level of –10VU. The

green Cal light should be illuminated, indicating proper operating level.

5. With the FM/AM and Avg/Peak buttons both out and the 5kHz and 1.0% FM buttons in, depress the Cal button. A reading of 0.68% indicates proper system operation. A 150Hz square-wave tone will be seen on the oscilloscope and heard in the monitor.
6. To begin the actual test of the recorder, select the 250Hz filter and choose the meter sensitivity range that yields a reading near midscale. The meter reading is the composite value of all flutter components in the frequency band of 0.5–250Hz, including flutter due to not only the rotating capstan, roller, and guides and their associated bearings but also any ac-power-related motor torque pulsations.
7. Select the Wtd. filter. The bandwidth is now reduced to 0.5–20Hz, to emphasize the once-around rates due to eccentricities of the rotating components. Capstans and rollers with diameters of ½–2in (12–50mm) are major contributions in this band.
8. Select the 250Hz–5kHz bandpass filter labeled ↔ The dominant component in this range is scrape flutter, which typically peaks at 3–4kHz for most recorders. Instabilities or oscillations of the capstan or spooling servos, which tend to occur in the 100–500Hz range, may also be evident.
9. If the machine is equipped with a scrape flutter idler, stall the idler by pressing the point of a pencil against the top of the idler. The scrape flutter component should typically rise to two or three times the normal value. If little or no rise or even a decrease is noted, the scrape flutter idler is not functioning properly. Clean and lubricate the idler bearings according to the manufacturer's instructions. Use the flutter meter to obtain optimum positioning of the idler after cleaning.

10. Select the 5kHz filter. This overall reading covers the entire range from 0.5Hz–5kHz.

32.7 Tape Testing

Contrary to popular belief, not all tape that reaches the customer's hands is fault free. Although the tape manufacturers are to be commended for the very high standards of excellence that are maintained, the customer must be prepared to deal with the bad rolls of tape that slip through the manufacturer's quality control screening. The problems can usually be traced to one of the seven steps in the manufacturing process:

1. The basic recipe of approximately a dozen major ingredients that form the oxide mixture must be correctly formulated. Each ingredient must be pure and must be measured correctly. Errors in mixing and experimental formula modifications often lead to nondurable oxides that shed debris onto the guides and heads.
2. The mixing of the ingredients must be thorough but not excessive. Inadequate mixing leads to high modulation noise and high background noise. Excessive mixing reduces noise but increases print through.
3. The coating process must apply a uniform coating across the width and length of the tape. The coating is applied to jumbo rolls that range from 18–36in (0.5–1m) in width. To monitor the entire width of one of these rolls fully would require over 400 channels of conventional record/reproduce circuits!
4. The tape is baked to remove solvents by passing the coated web through a multizone oven. Poor temperature control can lead to either brittle or soft oxides.

5. The jumbo roll is run through heated rollers that make the oxide denser to increase output and high-frequency response. This calendaring step is a major factor in determining the modulation noise content of the finished tape.
6. The tape is slit to the final width by a set of rotary shears. Poor slitting can produce ruffled edges, wavy or crooked tape, and excessive oxide and backing debris on the recording surface.
7. The tape is rewound onto reels or hubs, tested, and then packaged for sale. The tape cartons usually pass through a very large degausser so that no residual signals are left on the tape.

Mistakes during the manufacturing process create four types of problems. The most common of these is signal amplitude variations, which are due to either a nonhomogeneous magnetic dispersion or erratic tape-to-head contact due to physical distortions of the tape. Other common problems include excessive noise or distortion and high print through.

A common method of testing the signal instability and dropouts is to observe the amplitude variations of a sine-wave signal on either an oscilloscope or a VU meter. While these techniques give some insight into the performance of the tape, they do not yield a quantitative value that can be used for determining acceptable limits of performance.

A more informative method is to amplitude demodulate the test signal to remove the steady tone and magnify the fluctuations. If the output of the demodulator is properly filtered and fed to a metering circuit, quantitative values for the fluctuations in various test bandwidths can be read.

Unlike other flutter test instruments, the flutter meter shown in

Fig. 32-35 contains amplitude-demodulating circuitry to be used for testing tape. The AM test configuration is identical to the previous flutter setup, except that the FM/AM selector is set for AM mode testing to connect the phase-lock loop as a synchronous amplitude demodulator. The AM meter ranges, which are ten times larger than the flutter ranges, are labeled below the meter ranging pushbuttons.

The AM reading for 15 in/s (38cm/s) operation is typically 0.5% rms for a good roll of tape on a professional recorder. The texture of the demodulation products coming from the audio monitor should be a low rumbling with only occasional moderate bursts. The high-pass filter ↔ should produce a uniform hiss.

Typical symptoms of bad rolls of tape include readings that are approximately three times higher than the normal readings or very large frequent bursts that drive the meter pointer hard against the upper stop. Routine studio tests of large quantities of tape stock over a period of two years has shown that these easily spotted characteristics are good indicators of defective tape.

Although amplitude variations are symptomatic of bad tape, the tape transport and heads are also possible sources. If the tape is not being held snugly against the faces of the heads due to inadequate tape tension, the tape may suffer irregular spacing loss. Other contributors are dirt on the heads or heads that have been worn so flat that the gap is no longer pressed firmly against the tape. Mechanical misalignments, such as a twisted head or improperly positioned guides or scrape flutter idlers, can also degrade the contact between the tape and head.

Misadjustments of the bias amplitude or even-order distortions of the bias or erase waveforms can also produce excessive AM levels. Always verify that the bias levels and tuning are correct before condemning the tape.

A simple method of avoiding embarrassment when a defective roll of tape is suspected is to recheck the machine with a reference roll of the same type of tape that is known to be good. If changing from the reference roll to the suspect roll causes a large increase in AM content, then the tape is the source of the problem.

Since none of the tape manufacturers supplies information that is useful for specifying the AM performance of a tape, the user must generate data by testing several rolls of tape on machines. Once this process is begun, subsequent additions to the database will provide even more insight into the expected range of values.

32.8 Magnetic Head Troubleshooting and Maintenance

Troubleshooting any piece of complex equipment requires a methodical search technique to isolate the source of the problem quickly. The most productive technique is to conduct a series of tests that subdivide the faulty portion of the total system into smaller and smaller parts until the fault source is finally isolated.

Applying this technique to a magnetic tape recorder would lead to partitioning questions such as:

1. Is the problem associated with the tape drive, the audio circuitry, or the control logic?
2. Does the fault occur during recording, playback, and/or input monitoring?
3. Is the problem due to the recorder or the roll of tape?
4. Is the problem similar at both tape speeds?
5. Is the problem the same throughout the reel of tape?

6. Does temperature or running time have an effect?

If the problem relates to the audio signal passing through the recorder, a fundamental question that must be answered is whether the problem is wavelength-dependent or frequency-dependent. Wavelength problems immediately isolate the problem to the interface between the moving tape and the heads. Frequency problems are often related to the audio circuits.

A very useful tool for separating wavelength problems from frequency problems is a simple device known as a *flux loop* shown in Fig. 32-36. The flux loop, which consists of nothing more than a few turns of fine magnet wire driven with a constant current from an audio oscillator, creates a magnetic field that simulates a perfect lossless piece of tape. When the flux loop is attached to the gap region of the playback head, the flux from the loop excites the head much like the primary winding on a transformer excites the secondary winding. This direct excitation eliminates all the wavelength effects associated with gap length, azimuth error, and thickness loss. If the reproduce electronics perform correctly when excited by the flux loop but still fail to reproduce a known-good prerecorded test tape correctly, the problem is a wavelength-dependent error at the head-to-tape interface.

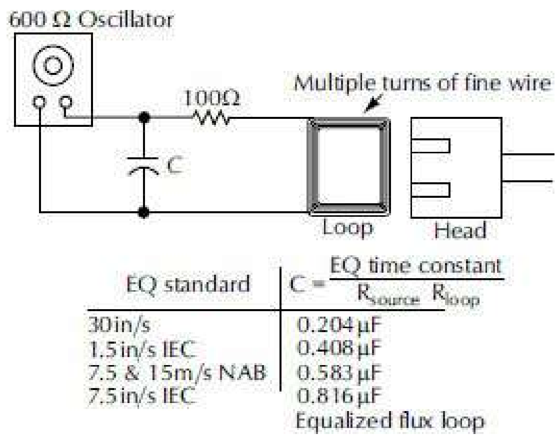


Figure 32-36. Flux loop.

The playback response from a simple flux loop is by no means flat. Since the dominant loss due to the coating thickness is not present for flux loop excitation, the high-frequency response with a flux loop will show a pronounced rise that relates to the particular reproduce equalization standard that is being utilized. NAB low-frequency equalization will also produce a roll-off below 50Hz.

To simplify the measurement process, the oscillator signal feeding the flux loop is usually preequalized to accommodate these effects of the equalization standard. Fig. 32-36 includes a simple circuit for correcting the high end, with capacitor values for several common equalizations. (The 600Ω impedance of the oscillator is part of the filter. For a 50Ω oscillator, multiply the capacitor values by 2.57.) The resulting high-frequency playback response of an equalized flux loop will be flat except for any residual high-frequency discrepancies due to eddy current losses or self-resonance of the playback head and cabling.

The flux loop can also be used in reverse as a pickup device to probe the magnetic fields generated at the gaps of the record and erase heads. If the driving network is disconnected and the loop connected directly to the inputs of an oscilloscope and meter, the relative magnitude of the bias and audio fields can be examined. Care must be exercised to correct for the 6dB/octave rise in flux loop output voltage due to the inductive nature of the flux loop. (A resistor in series with the input and a capacitor shunted across the input can be used to create an integrating low-pass filter that will flatten out this 6dB/octave rise.)

Details regarding the construction and use of a flux loop, along with detailed mechanical alignment procedures for azimuth, height, and tape wrap, are available from the various tape recorder manufacturers.

Head Relapping

The performance characteristics gradually change as the abrasive action of the tape wears away the faces of the heads. The resulting decreases in gap depth will reduce shunting effects, leading to an increase in efficiency for both the record and playback heads. Bias and audio levels must be gradually reduced to offset the rising efficiency. A critical point is reached, however, when the useful face of the head has been completely removed and the length of the gap begins to increase quickly with wear. The top end of the playback response will drop abruptly within a matter of only a few hours of use, rendering the recorder unusable. At this point, the head must be replaced to restore normal performance.

The heads on most recorders require attention long before this point of ultimate failure is reached. On most machines, the tape wears away the rounded apex at the gap of the head, leading to a drop in contact pressure with the tape at the gap. The tape begins to lift off the head slightly, creating erratic short-wavelength performance due to the spacing loss effect.

The common solution is to recontour the face of the head to restore the contact pressure. This process, known as *head relapping*, can be utilized two or three times during the useful life of a head to restore original performance. Although the average technician can be trained in the relapping process, the high cost of a mistake with a 2in (50mm) multitrack head assembly suggests that the more exotic relapping tasks should be handled by relapping specialists.

32.9 Routine Signal Alignment Procedure

A common problem arises with conventional recorders and alignment procedures—namely, that the procedures require a change in each adjustment to verify that the optimum point has been reached. This typically leads to not only the premature demise of many trimmer potentiometers (which are typically rated by the manufacturer for a life of 200 adjustment cycles) and head azimuth hardware, but also many operator errors due to the tedious nature of adjusting a multitrack machine that may have as many as 1000 adjustments.

If the operator is willing to adopt a philosophy that most of the adjustments are probably adequately close to optimum and that they need not be readjusted, then the alignment task shifts to looking for the exceptions to the norm rather than arbitrarily resetting everything. This strategy promotes better results since each iteration of the alignment procedure serves to fine-tune the results rather than to erase all past efforts and start afresh for each alignment with a high probability of error.

A few exceptions to the need for tweaking to verify proper performance are worthy of note. For example, head azimuth can be verified with a differential method that uses alternating test segments that have equal but opposite amounts of deliberate azimuth error. If the drop in level is equal for both directions of tilt, then the head must be correctly aligned to the correct vertical reference. No head adjustments are required if the test results are satisfactory.

A similar noninvasive test procedure for optimizing the bias level can be achieved if the bias system contains a master bias level trimmer that varies the level of bias for all tracks simultaneously. The bias level can be increased and decreased on all tracks with this single control to verify that the proper level of overbias is achieved without resorting to unnecessary adjustments on each track.

The following sequence of steps represents a comprehensive alignment procedure that would be appropriate whenever the proper performance of a recorder must be verified. Since the details of each step vary with machine type, the operator should consult the operator's manual published by the manufacturer of the recorder.:

1. Clean and inspect the tape transport. (Refer to section 32.7).
2. Degauss the heads and guides. (Refer to section 32.3.6). Before using a degausser, always verify that the tips of the unit are covered with a soft material such as plastic or tape that will not scratch the faces of the magnetic heads.
3. Calibrate the reproduce section of the recorder with a test tape of known accuracy. Several brands of standard alignment tapes are available for this purpose. Remember that the final results will be no better than the measurement standard that is being used as a reference.

   First, verify the perpendicular alignment of the reproduce head with the short-wavelength azimuth test tone on the test tape. The azimuth and/or phase alignment of the head can be measured with an oscilloscope using either a Lissajous pattern or a dual-trace display or with a phase meter that reads phase error directly. If no specialized equipment is available, invert one channel and sum the inverted output with another channel that is not inverted. Phase alignment produces a deep null in the summed output. Since phase alignment at one frequency does not eliminate the possibility of a 360° error, check the phase for several lower frequencies. The voice announcements on the alignment tapes provide a convenient multi-frequency sample for this purpose.

Next, establish a convenient reference level for making playback frequency-response measurements. Check and adjust the high-frequency reproduce equalizer at 10kHz to match this reference level. Once the equalizer has been set, sweep through the tones on the tape, noting the maximum deviations from the reference value. Readjust the equalizer and the reference level as necessary to obtain the desired degree of flatness.

When the results are satisfactory, write down the results for later comparison. Having a record of correct performance makes troubleshooting much easier.

Two pitfalls exist when making the previously discussed adjustments: one affects the reference level and the other affects the frequency-response and reference level. Some recorders use different track widths for the record and playback heads. For machines that have wider playback heads, the full-track test tapes used for most of the wide-tape formats will produce an enhanced output during testing. The reference level from the tape must be set above the 0VU reference by the amount of this extra pickup due to the wider head when using the playback head. When setting the reference level for sync/overdub playback, the track width is correct, yielding a true 0 VU level that requires no correction.

If the record head has a wider track, then the normal playback level will be correct and the error will occur on the sync/overdub level.

The second problem is created by the fringing effect of long wavelengths that produces a rise in playback response at low frequencies whenever additional flux is present beyond the area being scanned by the reproduce head. Such a condition exists for playback of a full-track alignment tape and for test and alignment procedures that apply the same low-frequency signal to all tracks of the recorder simultaneously.

The fringing effect will first create a problem in establishing the correct reference level for the midband-level set tone. At 15 in/s and 30in/s (38 cm/s and 76 cm/s) tape speeds, sufficient fringing may exist to create an error of approximately 0.57–1 dB, depending on the track format, tape speed, and geometry of the head cores and shielding. This extra fringing contribution in the reference tone also makes the high-frequency response appear to be deficient, tempting the operator to raise the equalizer adjustment. Consult the operator's manual for the correct procedure and correction factors for a given model of recorder.

The final step in the reproduce alignment procedure is to set the level and equalization of the sync/overdub circuit. The operator may choose to defer the azimuth alignment of the record head until the following record alignment procedure if the heads have not been disturbed.

4. The record alignment begins with the verification and/or

adjustment of the azimuth setting of the record head. Using the playback head as a standard, set the record head alignment while recording a short-wavelength signal such as a 10kHz or 15kHz signal to give minimum azimuth or phase error using whatever method was used for the reproduce alignment procedure. This alignment should be rechecked after the bias and record equalization settings are made, since these adjustments can introduce varying amounts of phase delay.

The bias should be set by adjusting for the desired amount of overbias as recommended by the tape and machine manufacturer for the appropriate type of tape, record head gap width, and tape speed. Note that a 10kHz signal at 30in/s (76cm/s) does not achieve the desired wavelength of 1.5 mils (38μm) that is typically specified for bias adjustment. The test frequency must be changed to match the tape speed.

The bias should first be decreased to achieve deliberate underbias, and then slowly increased to the point at which a peak in the playback level is observed. Continue to increase the bias until the signal drops by the number of decibels desired. Typical overbias settings range from 2–5 dB for professional formats.

Once the bias is correctly adjusted, the input signal should be set to the frequency used as a reference during the playback alignment. The record gain control can then be set to produce the reference level when driven with the appropriate 0VU input level.

Adjust the high-frequency record equalizer to match the record/play response as closely as possible to the alignment tape response noted previously. Smoothness in the midband frequencies is more important than trying to hold small errors at 15kHz or 20kHz.

Recheck the record head azimuth to verify that changes in bias and equalization have not created any phase differences. Readjust as necessary until all parameters are optimized.

Set the record gain preset and the input monitor gain calibration to achieve a 0VU reading in all monitor modes.

5. After the record section has been aligned, a final test and alignment of the low-frequency playback equalizers can be undertaken. To eliminate all the fringing problems previously mentioned, the equalizers should be set in the record/play mode with signal being applied to every other track. Make small adjustments as required to optimize the smoothness of the response.

   If any large discrepancies are noticed, rerun the alignment tape. Any failure in the low-frequency record equalizer circuits, such as a faulty switching component, will create an error that should be obvious if a large correction is required. If any doubt still exists, record a full-frequency sweep and then flip the reels over to play the tape backward. The alignment should be similar within a few tenths of a dB to the values set in the forward direction.

6. The alignment procedure is not completed until the noise level and erasure have been checked. Record a signal at +6VU, rewind the tape, and then erase the signal. Listen on the monitor speakers to the level of the residual signal and to the subjective nature of the tape noise. The tone should be either completely eliminated or well buried in the tape noise. The noise should be a smooth hiss without large or frequent bursts or crackling. All tracks should be similar in performance. Also, check for objectionable clicks and pops when changing modes.

> Although these noise and erasure levels can be read from instruments, the operator should take the time to listen to the machine before issuing his or her stamp of approval. Many sessions have died aborning because the recorder was never given a final listening test after alignment.

The previous procedure does not include several steps that are more appropriately considered to be maintenance routines. Examples include tuning of the bias and erase sources, tuning of bias traps, checking meter calibration, and testing distortion levels. These tests are not required on a day-to-day basis.

As a final note on alignment, never gloss over large discrepancies. The corrections that should be required for this alignment procedure should be on the order of a small part of a dB, not several dB. Whenever a large change seems required, stop long enough to determine why such a large change is necessary. Look for faulty components and recheck your own procedure. Recheck the maintenance log to establish the proper level of performance that should be expected. Heeding the small symptoms may help you avoid a serious catastrophic failure.

32.10 Automated Alignment

The onslaught of digital technology has provided the tools to control the variable alignment adjustments of a tape recorder with a microprocessor. Multiple sets of calibration constants can be stored in nonvolatile memory, permitting rapid changes of operating speeds, equalization standards, reference flux levels, and tape types.

Once the provisions for automated adjustment are made available, three methods of alignment are possible. Under the simplest mode, the operator performs a manual alignment with the calibration constants being stored for later use. This method permits rapids change-overs, but does not simplify bias and equalization adjustments to optimize a specific roll of tape.

If the microprocessor can be provided with input information from the metering devices on the individual tracks, then calibration programs can be automatically executed without operator intervention. The program contains the “strategy” for alignment, including desired amounts of overbias, equalization adjustment frequencies, and operating levels. Beware that such systems use an inferred adjustment technique which does not actually test many of the critical parameters. For example, the recorder will set the bias level for minimum distortion based on an overbias criterion at a specified frequency. In reality, the machine doesn’t have the ability to measure distortion. The strategy only infers that overbiasing by the desired amount corresponds to minimum distortion. Unfortunately, if a malfunction exists that causes abnormal operation, the adjustment routine may not detect the symptoms.

Nearly automatic calibration can be implemented by connecting external automated test equipment such as an Audio Precision System One test set to the machine through an external intelligent controller such as an IBM-compatible computer. A remotely controlled input/output switching matrix will also be necessary for multitrack machines. An operator is still required to adjust nonautomated devices such as head azimuth and to change tape reels for calibration tapes and sample stock. The calibration program of the intelligent controller sequences through a comprehensive set of tests which rigorously exercise the machine. Parameters such as harmonic and intermodulation distortions, crosstalk, erasure, flutter, speed, noise, and phase can be tested against absolute standards of acceptance.

A final word of caution is appropriate at this point. Many operators and test technicians ignore symptoms that indicate problems are developing in a tape recorder. A good example is the frequent need to boost the high-frequency equalization adjustments of a recorder. A properly operating machine should not show such trends, but a gradually deteriorating head would create just such a problem. Simply readjusting without determining the cause of the change wastes an opportunity to fix a problem at an early stage before it grows to catastrophic consequences. Try to avoid problems by fixing things before they break completely.

Chapter 33

MIDI

by David Miles Huber

33.1 Introduction to MIDI

Today, professional and nonprofessional musicians are using the language of the *Musical Instrument Digital Interface* (MIDI) to perform an expanding range of music and automation tasks within audio production, audio for video, film post, stage production, etc. This industry-wide acceptance can, in large part, be attributed to the cost effectiveness, power and general speed of MIDI production. Once a MIDI instrument or device comes into the production picture, there's often less need (if any at all) to hire outside musicians for a project. This alluring factor allows a musician to compose, edit and arrange a piece in an electronic music environment that is extremely flexible. By this, I'm not saying that MIDI replaces (or should replace) the need for acoustic instruments, microphones and the traditional performance setting. In fact, it's a powerful production tool that assists countless musicians in creating music and audio productions in ways that are innovative and highly personal. In short, MIDI is all about control, repeatability, flexibility, cost-effective production power and fun.

33.1.1 What is MIDI?

Simply stated, the Musical Instrument Digital Interface (MIDI) is a digital communications language and compatible specification that allows multiple hardware and software electronic instruments, performance controllers, computers and other related devices to communicate with each other over a connected network. MIDI is used to translate performance- or control-related events (such as playing a keyboard, selecting a patch number, varying a modulation wheel, triggering a staged visual effect, etc.) into equivalent digital messages and then transmit these messages to other MIDI devices where they can be used to control sound generators and other performance/control parameters. The beauty of MIDI is that its data can be recorded into a hardware device or software program (known as a sequencer), where it can then be edited and transmitted to electronic instruments or other devices to create music or control any number of parameters in a performance- or post-performance setting.

In addition to composing and performing a song, musicians can also act as techno-conductors, having complete control over a wide palette of sounds, their timbre (sound and tonal quality), overall blend (level, panning) and other real-time controls. MIDI can also be used to vary the performance and control parameters of electronic instruments, recording devices, control devices and signal processors in the studio, on the road or on the stage.

The term *interface* refers to the actual data communications link and software/hardware systems in a connected MIDI network. Through the use of MIDI, it’s possible for all of the electronic instruments and devices within a network to be addressed through the transmission of real-time performance and control-related

MIDI data messages throughout a system to multiple instruments and devices through one or more data lines (which can be chained from device to device). This is possible because a single data cable is capable of transmitting performance and control messages over 16 discrete channels. This simple fact allows electronic musicians to record, overdub, mix and play back their performances in a working environment that loosely resembles the multitrack recording process. Once mastered, MIDI surpasses this analogy by allowing a composition to be edited, controlled, altered and called up with complete automation and repeatability all of this providing production challenges and possibilities that are well beyond the capabilities of the traditional tape-based multitrack recording process.

33.1.2 What MIDI Isn't

For starters, let's dispel one of MIDI's greatest myths: MIDI DOESN'T communicate audio, nor can it create sounds! It is strictly a digital language that instructs a device or program to create, play back or alter the parameters of sound or control function. It is a data protocol that communicates on/off triggering and a wide range of parameters to instruct an instrument or device to generate, reproduce or control audio or production-related functions. Because of these differences, the MIDI data path and the audio routing paths are entirely separate from each another, Fig. 33-1. Even if they digitally share the same transmission cable (such as through FireWire or USB), the actual data paths and formats are completely separate.

In short, MIDI communicates information that instructs an instrument to play or a device to carry out a function. It can be likened to the dots on a player-piano roll; when we put the paper roll up to our ears, we hear nothing, but when the cut-out dots pass over the sensors on a player piano, the instrument itself begins to make beautiful music. It's exactly the same with MIDI. A MIDI file or datastream is simply a set of instructions that pass down a wire in a serial fashion, but when an electronic instrument interprets the data, we begin to hear sound.

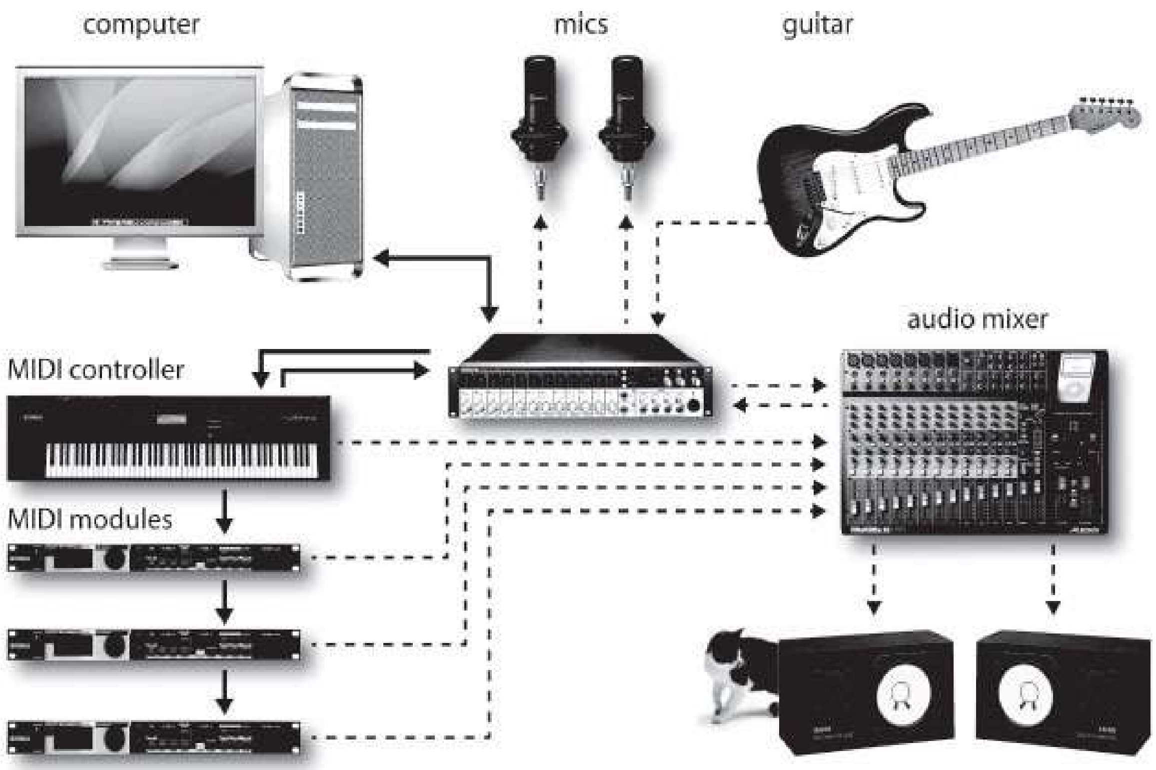


Figure 33-1. Example of a typical MIDI system with the MIDI network connections highlighted in solid lines.

33.2 System Interconnections within MIDI Production

As a data transmission medium, MIDI is able to pack performance (16 discrete channels), controller and timing information and then transmit it in one direction, using data densities that are economically small and easy to manage. In this way, it's possible for MIDI messages to be communicated from a specific source (such as a keyboard or MIDI sequencer) to any number of devices within a connected network over a single MIDI data chain. In addition, MIDI is flexible enough that multiple MIDI data lines can be used to interconnect devices in a wide range of possible system configurations; for example, multiple MIDI lines can be used to transmit data to instruments and devices over 32, 48, 128 or more discrete MIDI channels!

Of course, those of you who are familiar with the concept of MIDI know that, over the years, the concept of interconnecting electronic instruments and other devices together has changed. These days, you're more likely to connect a device to a computer by using a USB, Firewire or Thunderbolt cable than the older standard cabling systems. In fact, often these interconnections are made "under the virtual hood", within the overall system and cable connectivity limitations aren't usually an issue. It's for this reason that it's important that we have an understanding of how these data connections are made "at a basic level"...therefore, I present to you, the MIDI cabling system.

33.2.1 The MIDI Cable

A MIDI cable, Fig. 33-2, consists of a shielded, twisted pair of conductor wires that has a male 5-pin DIN plug located at each of its ends. The MIDI specification currently uses only three of the five pins, with pins 4 and 5 being used as conductors for MIDI data; pin 2 is used to connect the cable's shield to equipment ground. Pins 1 and 3 are currently not in use, although the next section describes an ingenious system for powering devices through these pins that's known as MIDI phantom power. The cables use twisted cable and metal shield groundings to reduce outside interference, such as radio-frequency interference (RFI) or electrostatic interference, both of which can serve to distort or disrupt the transmission of MIDI messages.

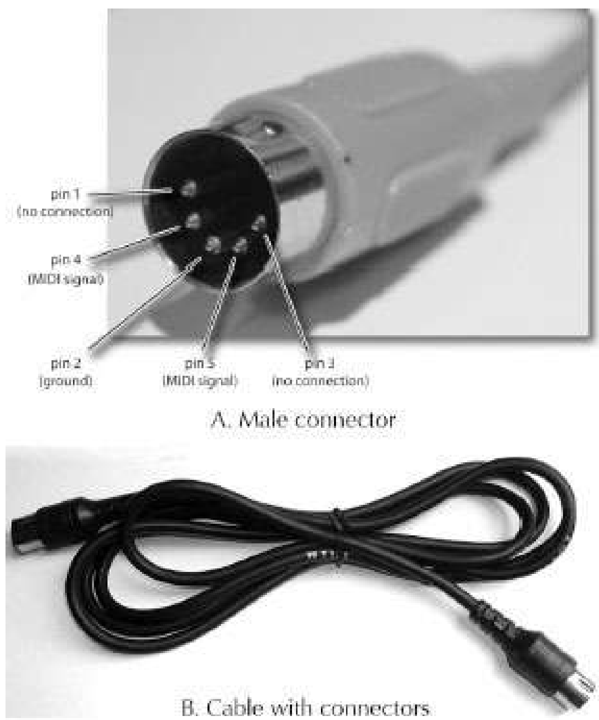


Figure 33-2. The MIDI cable: (A) wiring diagram; (B) the cable itself.

33.2.2 MIDI Pin Description

- Pin 1 is not used in most cases; however, it can be used to provide the VB (ground return) of a MIDI phantom power supply.
- Pin 2 is connected to the shield or ground cable, which protects the signal from radio and electromagnetic interference.
- Pin 3 is not used in most cases; however, it can be used to provide

the +V (+9 to +15V) of a MIDI phantom power supply.
- Pin 4 is a MIDI data line.
- Pin 5 is a MIDI data line.

MIDI cables come prefabricated in lengths of 2, 6, 10, 20 and 50 feet and can commonly be obtained from music stores that specialize in MIDI equipment. To reduce signal degradations and external interference that tends to occur over extended cable runs, 50 feet is the maximum length specified by the MIDI spec.

Again, It should be noted, however, that in modern-day MIDI production, it's become increasingly common for MIDI data to be transmitted throughout a system network through USB, Firewire and wireless interconnections. Although the data isn't transmitted through traditional MIDI cabling, the data format adheres to the MIDI protocol.

33.2.3 MIDI Phantom Power

In December 1989, Craig Anderton (musician and audio guru) submitted an article to *EM* proposing an idea that provides a standardized 12Vdc power supply to instruments and MIDI devices directly through pins 1 and 3 of a basic MIDI cable. Although pins 1 and 3 are technically reserved for possible changes in future MIDI applications (which never really came about), over the years several forward-thinking manufacturers (and project enthusiasts) have begun to implement MIDI phantom power directly into their studio and on-stage systems.

33.2.4 Wireless MIDI

Several companies have begun to manufacture wireless MIDI transmitters that allow a battery-operated MIDI guitar, wind controller, etc., to be footloose and fancy free on-stage and in the studio. Working at distances of up to 500 feet, these battery-powered transmitter/receiver systems introduce very low delay latencies and can be switched over a number of radio channel frequencies.

33.2.5 MIDI Jacks

MIDI is distributed from device to device using three types of MIDI jacks: MIDI In, MIDI Out and MIDI Thru, Fig. 33-3. These three connectors use 5-pin DIN jacks as a way to connect MIDI instruments, devices and computers into a music or production network system. As a side note, it's nice to know that these ports (as strictly defined by the MIDI 1.0 spec) are optically isolated to eliminate possible ground loops that might occur when connecting numerous devices together.

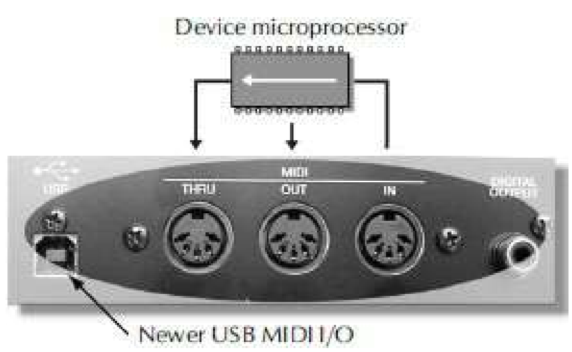


Figure 33-3. MIDI In, Out and Thru ports, showing the device's signal path routing.

MIDI In jack. The MIDI In jack receives messages from an external source and communicates this performance, control and timing data to the device's internal microprocessor, allowing an instrument to be played or a device to be controlled. More than one

MIDI In jack can be designed into a system to provide for MIDI merging functions or for devices that can support more than 16 channels. Other devices (such as a controller) might not have a MIDI In jack at all, but only sport a MIDI Out.

MIDI Out jack. The MIDI Out jack is used to transmit MIDI performance, control messages or SysEx from one device to another MIDI instrument or device. More than one MIDI Out jack can be designed into a system, giving it the advantage of controlling and distributing data over multiple MIDI paths using more than just 16 channels (i.e., 16 channels × N MIDI port paths).

MIDI Thru jack. The MIDI Thru jack retransmits an exact copy of the data that's being received at the MIDI In jack. This process is important, because it allows data to pass directly through an instrument or device to the next device in the MIDI chain. Keep in mind that this jack is used to relay an exact copy of the MIDI In datastream, which isn't merged with data being transmitted from the MIDI Out jack.

MIDI Echo. Certain MIDI devices may not include a MIDI Thru jack at all. They may, however, give the option of switching the MIDI Out between being an actual MIDI Out jack and a MIDI Echo jack, Fig. 33-4. As with the MIDI Thru jack, a MIDI Echo option can be used to retransmit an exact copy of any information that's received at the MIDI In port and route this data to the MIDI Out/Echo jack. Unlike a dedicated MIDI Out jack, the MIDI Echo function can often be selected to merge incoming data with performance data that's being generated by the device itself. In this way, more than one controller can be placed in a MIDI system at one time. Note that, although performance and timing data can be echoed to a MIDI Out/Echo jack, not all devices are capable of echoing SysEx data.

33.2.6 Typical Configurations

Although electronic studio production equipment and setups are rarely alike (or even similar), there are a number of general rules that make it easy for MIDI devices to be connected to a functional network. These common configurations allow MIDI data to be distributed in the most efficient and understandable manner possible.

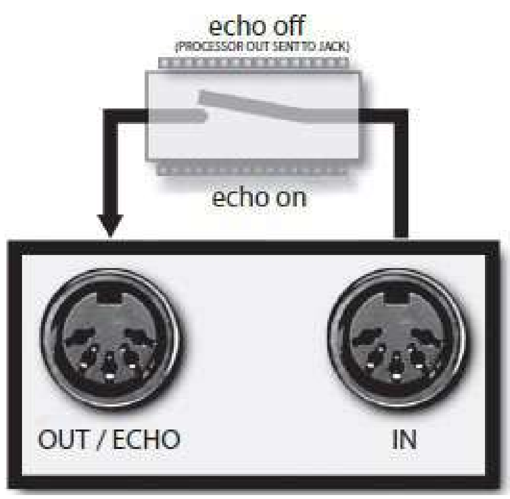


Figure 33-4. MIDI echo configuration.

As a primary rule, there are only two valid ways to connect one MIDI device to another within a MIDI cable chain. Fig. 33-5:

- The MIDI Out jack of a source device (controller or sequencer/computer) must be connected to the MIDI In of a second device in the chain.
- The MIDI Thru jack of the second device must be connected to the MIDI In jack of the third device in the chain… following this

same Thru-to-In convention until the end of the chain is reached.

Figure 33-5. The two valid means of connecting one MIDI device to another.

33.2.7 The Daisy Chain

One of the simplest and most common ways to distribute data throughout a MIDI system is the daisy chain. This method relays MIDI data from a source device (controller or sequencer/computer) to the MIDI In jack of the next device in the chain (which receives and acts on this data). This next device then relays an exact copy of the incoming data onwards to its MIDI Thru jack, which is then relayed to the next device in the chain, and so on through the successive devices. In this way, up to 16 channels of MIDI data can be chained from one device to the next within a connected data network and it's precisely this concept of transmitting multiple channels through a single MIDI line that makes the whole concept work! Let's try to understand this system better by looking at a few examples.

Fig. 33-6A shows a simple (and common) example of a MIDI daisy chain whereby data flows from a controller (MIDI Out jack of the source device) to a synth module (MIDI In jack of the second device in the chain). An exact copy of the data that flows into the second device is then relayed to its MIDI Thru jack to another synth (MIDI In jack of the third device in the chain). If our controller is set to transmit on MIDI channel 3, the second synth in the chain (which is set to channel 2) will ignore the messages and not play, while the third synth (which is set to channel 3) will be playing its heart out. The moral of this story is that, although there is only one connected data line, a wide range of instruments and channel voices can be played in a surprisingly large number of combinations—all by using individual channel assignments along a daisy chain.

Another example, Fig. 33-6B, shows how a computer can easily be designated as the master source within a daisy chain so a sequencing program can be used to control the entire playback and channel routing functions within a daisy-chained system. In this situation, the MIDI data flows from a master controller/synth to the MIDI In jack of a computer's MIDI interface (where the data can be played into, processed and rerouted through a MIDI sequencer). The MIDI Out of the interface is then routed back to the MIDI In jack of the master controller/synth (which receives and acts on this data). The controller then relays an exact copy of this incoming data out to its MIDI Thru jack (which is then relayed to the next device in the chain) and so on, until the end of the chain is reached. When we stop and think about it, we can see that the controller is essentially used as a "performance tool" for entering data into the MIDI sequencer, which is then used to communicate this data out to the various instruments throughout the connected MIDI chain.

33.2.8 iConnections

In addition to wired connections via standard MIDI, USB and Firewire cables, MIDI can also be connected to many pad and phone apps and devices using any number of wireless and wired methods for connecting portable on-the-go apps to your music production system. Any number of YouTube videos and online resources can be found on this ever-expanding subject.

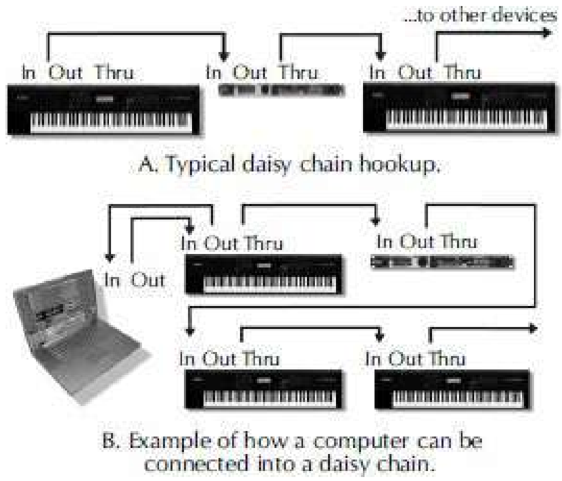


Figure 33-6. Example of a connected MIDI system using a daisy chain: (A) typical daisy chain hookup; (B) example of how a computer can be connected into a daisy chain.

33.3 Connecting MIDI to the Peripheral World

An important event in the evolution of personal computing has been the maturation of hardware and processing peripherals. With the development of the USB (www.usb.org) and FireWire (www.1394ta.org) protocols, hardware devices such as mice, keyboards, cameras, audio interfaces, MIDI interfaces, CD and hard drives, MP3 players and even portable fans can be plugged into an available port without any need to change frustrating hardware settings or open up the box. External peripherals are generally hardware devices that are designed to do a specific task or range of production tasks. For example, an audio interface is capable of translating analog audio (and often MIDI, control and other media) into digital data that can be understood by the computer. Other peripheral devices can perform such useful functions as printing, media interfacing (video and MIDI), scanning, memory card interfacing, portable hard disk storage…the list could literally fill pages.

The MIDI Interface

Although computers and electronic instruments both communicate using the digital language of 1s and 0s, computers simply can't understand the language of MIDI without the use of a device that translates the serial messages into a data structure that computers can comprehend. Such a device is known as the *MIDI interface*. A wide range of MIDI interfaces currently exist that can be used with most computer system and OS platforms. For the casual and professional musician, interfacing MIDI into a production system can be done in a number of ways. Probably the most common way to access MIDI In, Out (and sometimes Thru) jacks is on modern-day USB or FireWire audio interface or controller surface. It's become a common matter for portable devices to offer 16 channels of I/O (on one port), while multichannel interfaces often include multiple MIDI I/O ports that can give you access to 32 or more channels.

The next option is to choose a USB MIDI interface that can range from simpler devices that include a single port to multiple-port systems that can easily handle up to 64 channels over four I/O ports. The multiport MIDI interface, Fig. 33-7, is often the device of choice for most professional electronic musicians who require added routing and synchronization capabilities. These USB devices can easily be ganged together to provide eight or more independent MIDI Ins and Outs to distribute MIDI data through separate lines over a connected network.

Figure 33-7. M-Audio MIDISPORT 4x4 MIDI interface. Courtesy M-Audio, a division of Avid Technology, Inc., www.m-audio.com.

In addition to the above interface types, a number of MIDI keyboard controllers and synth instruments have been designed with MIDI ports and jacks built right into them. For those getting started, this useful and cost-saving feature makes it easy to integrate your existing instruments into your DAW and sequencing environment.

33.4 The MIDI Message

MIDI is a specified data format that must be strictly adhered to by those who design and manufacture MIDI-equipped instruments and devices. Because the format is standardized, you don't have to worry about whether the MIDI output of one device will be understood by the MIDI in port of a device that is made by another manufacturer. As long as the data ports say and/or communicate MIDI, you can be assured that the data (at least the basic performance functions) will be transmitted and understood by all devices within the connected system. In this way, the user need only consider the day-to-day dealings that go hand-in-hand with using electronic instruments, without having to be concerned with data compatibility between devices.

MIDI digitally communicates musical performance data between devices as a string of MIDI messages. These messages are traditionally transmitted through a standard MIDI line in a serial fashion at a speed of 31,250 bits/sec. Within a serial data transmission line, data travels in a single-file fashion through a single conductor cable; a parallel data connection, on the other hand, is able to simultaneously transmit digital bits in a synchronous fashion over multiple wires.

MIDI messages are made up of groups of 8-bit words (known as bytes), which are transmitted in a serial fashion to convey a series of instructions to one or all MIDI devices within a system.

Only two types of bytes are defined by the MIDI specification: the status byte and the data byte.

- A status byte is used to identify what type of MIDI function is to be performed by a device or program. It is also used to encode channel data (allowing the instruction to be received by a device that is set to respond to the selected channel).
- A data byte is used to associate a value to the event that is given by the accompanying status byte.

Although a byte is made up of 8 bits, the most significant bit (MSB; the leftmost binary bit within a digital word) is used solely to identify the byte type. The MSB of a status byte is always 1, while the MSB of a data byte is always 0. For example, a 3-byte MIDI Note-On message (which is used to signal the beginning of a MIDI note) might read in binary form as a 3-byte Note-On message of (10010100) (01000000) (01011001). This particular example transmits instructions that would be read as: Transmitting a Note-On message over MIDI channel #5, using keynote #64, with an attack velocity [volume level of a note] of 89.

33.4.1 MIDI Channels

Just as a public speaker might single out and communicate a message to one individual in a crowd, MIDI messages can be directed to communicate information to a specific device or range of devices within a MIDI system. This is done by embedding a channel-related nibble (4 bits) within the status/channel number byte. This process makes it possible for up to 16 channels of performance or control information to be communicated to a specific device, or a sound generator through a single MIDI data cable.

Since this nibble is 4 bits wide, up to 16 discrete MIDI channels can be transmitted through a single MIDI cable or designated port.

0000 = CH#1	0100 = CH#5	1000 = CH#9	1100 = CH#13
0001 = CH#2	0101 = CH#6	1001 = CH#10	1101 = CH#14
0010 = CH#3	0110 = CH#7	1010 = CH#11	1110 = CH#15
0011 = CH#4	0111 = CH#8	1011 = CH#12	1111 = CH#16

Whenever a MIDI device, sound generator within a device or program function is instructed to respond to a specific channel number, it will only react to messages that are transmitted on that channel (i.e., it ignores all channel messages that are transmitted on any other channel). For example, let's assume that we are going to create a short song using a synthesizer that has a built-in sequencer (a device or program that is capable of recording, editing and playing back MIDI data) and two other "synths," Fig. 33-8.

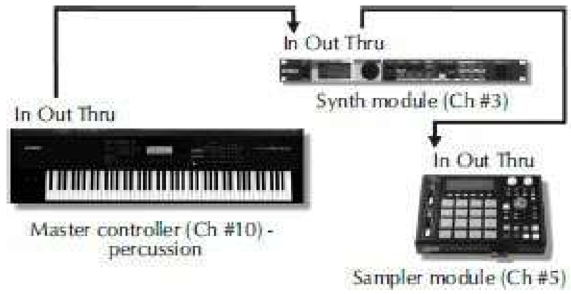


Figure 33-8. MIDI setup showing a set of MIDI channel assignments.

We could easily start off by recording a simple drum pattern track into the master synth on channel #10 (numerous synths are pre-assigned to output drum/percussion sounds on this channel).

Once recorded, the sequence will thereafter transmit the notes and data over channel 10, allowing the synth's percussion section to be played.

Next, we could set a synth module to channel #3 and instruct the master synth to transmit on the same channel (since the synth module is set to respond to data on channel 3, its generators will sound whenever the master keyboard is played). We can now begin recording a melody line into the sequencer's next track.

Playing back the sequence will then transmit data to both the master synth (perc section #10) and the module (melody line #3) over their respective channels. At this point, our song is beginning to take shape.

Now we can set the other synth (or other instrument type) to respond to channel #5 and instruct the master synth to transmit on the same channel, allowing us to further embellish the song.

Now that the song's beginning to take shape, the sequencer can play the musical parts to the synths on their respective MIDI channels all in an environment that gives us complete control over voicing, volume, panning and a wide range of edit functions over each instrument. In short, we have created a true multichannel production environment.

It goes without saying that the above example is just one of the infinite setup and channel possibilities that can be encountered in a production environment. It's often true, however, that even the most complex MIDI and production rooms will have a system a basic channel and overall layout that makes the day-to-day operation of making music easier. This layout and the basic decisions that you might make in your own room are, of course, up to you. Streamlining a system to work both efficiently and easily will come with time, experience and practice.

33.4.2 MIDI Modes

Electronic instruments often vary in the number of sounds and notes that can be simultaneously produced by their internal sound-generating circuitry. For example, certain instruments can only produce one note at a single time (known as a monophonic instrument), while others can generate 16, 32 and even 64 notes at once (these are known as polyphonic instruments). The latter type can easily play chords or more than one musical line on a single instrument at a time.

In addition, some instruments are only capable of producing a single generated sound patch (often referred to as a "voice") at any one time. Its generating circuitry could be polyphonic, allowing the player to lay down chords and bass or melody lines, but it can only produce these notes using a single, characteristic sound at any one time (e.g., an electric piano, a synth bass or a string patch). However, the vast majority of newer synths differ from this in that they are multitimbral in nature, meaning that they can generate numerous sound patches at any one time (e.g., an electric piano, a synth bass and a string patch). That is to say that it's common to run across electronic instruments that can simultaneously generate a number of voices, each offering its own control over a wide range of parameters. Best of all, it's also common for different sounds to be assigned to their own MIDI channels, allowing multiple patches to be internally mixed within the device to a stereo output bus or independent outputs.

The following list and figures explain the four modes that are supported by the MIDI spec:

- **Mode 1 (Omni On/Poly):** In this mode, an instrument will respond to data that is being received on any MIDI channel and then redirect this data to the instrument's base channel. In essence, the device will play back everything that is presented at its input in a polyphonic fashion, regardless of the incoming channel designations. As you might guess, this mode is rarely used.
- **Mode 2 (Omni On/Mono):** As in Mode 1, an instrument will respond to all data that's being received at its input, without regard to channel designations; however, this device will only be able to play one note at a time. Mode 2 is used even more rarely than Mode 1, as the device can't discriminate channel designations and can only play one note at a time.
- **Mode 3 (Omni Off/Poly):** In this mode, an instrument will only respond to data that matches its assigned base channel in a polyphonic fashion. Data that's assigned to any other channel will be ignored. This mode is by far the most commonly used, as it allows the voices within a multitimbral instrument to be

individually controlled by messages that are being received on their assigned MIDI channels. For example, each of the 16 channels in a MIDI line could be used to independently play each of the parts in a 16-voice, multitimbral synth.

- **Mode 4 (Omni Off/Mono):** As with Mode 3, an instrument will be able to respond to performance data that is transmitted over a single, dedicated channel; however, each voice will only be able to generate one MIDI note at a time. A practical example of this mode is often used in MIDI guitar systems, where MIDI data is monophonically transmitted over six consecutive channels (one channel/voice per string).

33.4.3 Channel Voice Messages

Channel Voice messages are used to transmit real-time performance data throughout a connected MIDI system. They are generated whenever a MIDI instrument's controller is played, selected or varied by the performer. Examples of such control changes could be the playing of a keyboard, pressing of program selection buttons or movement of modulation or pitch wheels. Each Channel Voice message contains a MIDI channel number within its status byte, meaning that only devices that are assigned to the same channel number will respond to these commands. There are seven Channel Voice message types: Note-On, Note-Off, Polyphonic Key Pressure, Channel Pressure, Program Change, Pitch Bend Change and Control Change:

Note-On messages. Indicate the beginning of a MIDI note, Fig. 33-9. This message is generated each time a note is triggered on a keyboard, drum machine or other MIDI instrument (by pressing a key, striking a drum pad, etc.). A Note-On message consists of three bytes of information: a MIDI channel number, a MIDI pitch number, and an attack velocity value (messages that are used to transmit the individually played volume levels [0B127] of each note).

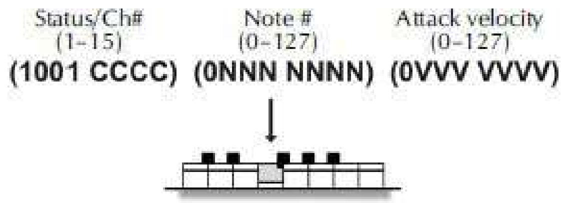


Figure 33-9. Byte structure of a MIDI Note-On message.

Note-Off messages. Indicate the release (end) of a MIDI note. Each note played through a Note-On message is sustained until a corresponding Note-Off message is received. A Note-Off message doesn't cut off a sound; it merely stops playing it. If the patch being played has a release (or final decay) stage, it begins that stage upon receiving this message. It should be noted that many systems will actually use a Note-On message with a velocity 0 to denote a Note-Off message.

Polyphonic Key Pressure messages. Transmitted by instruments that can respond to pressure changes applied to the individual keys of a keyboard, Fig. 33-10. A Polyphonic Key Pressure message consists of three bytes of information: a MIDI channel number, a MIDI pitch number and a pressure value.

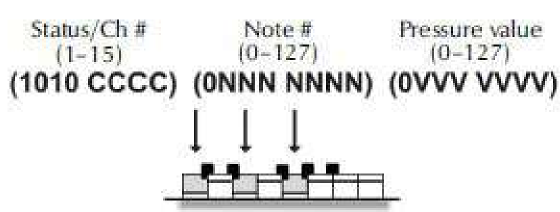


Figure 33-10. Byte structure of a MIDI Polyphonic Key Pressure

message (generated when additional pressure is applied to each key that is played).

Channel Pressure (or Aftertouch) messages. Transmitted and received by instruments that respond to a single, overall pressure applied to the keys, Fig. 33-11. In this way, additional pressure on the keys can be assigned to control such variables as pitch bend, modulation and panning.

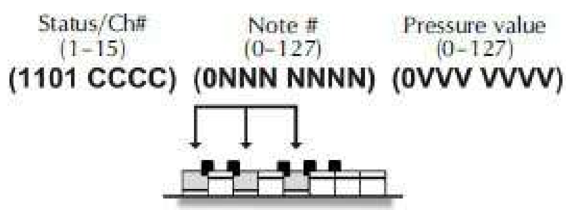


Figure 33-11. Byte structure of a MIDI Channel Pressure message (simultaneously affect all notes that are transmitted over a MIDI channel).

Program Change messages. Change the active voice (generated sound) or preset program number in a MIDI instrument or device, Fig. 33.12. Using this message format, up to 128 presets (a user- or factory-defined number that activates a specific sound-generating patch or system setup) can be selected. A Program Change message consists of two bytes of information: a MIDI channel number (1B16) and a program ID number (0B127).

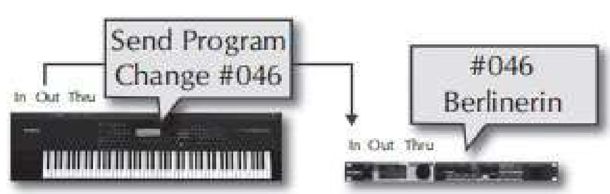


Figure 33-12. Program Change messages can be used to change sound patches from a sequencer or from a remote controller.

Pitch Bend Change messages. Transmitted by an instrument whenever its pitch bend wheel is moved in either the positive (raise pitch) or negative (lower pitch) direction from its central (no pitch bend) position, Fig. 33.13.

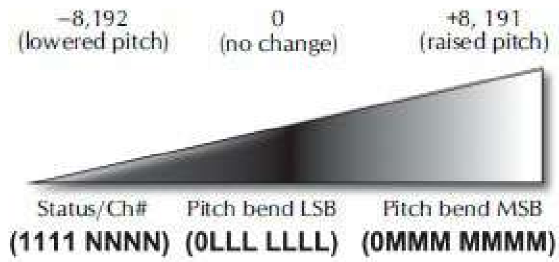


Figure 33-13. Byte structure of a Pitch Bend Change message.

Control Change messages. Transmit information that relates to real-time control over a MIDI instrument's performance parameters (such as modulation, main volume, balance and panning), Fig. 33-14. Three types of real-time controls can be communicated through control change messages: continuous controllers, which communicate a continuous range of control settings, generally with values ranging from 0B127; switches (controls having an ON or OFF state with no intermediate settings); and data controllers, which enter data either through numerical keypads or stepped up/down entry buttons.

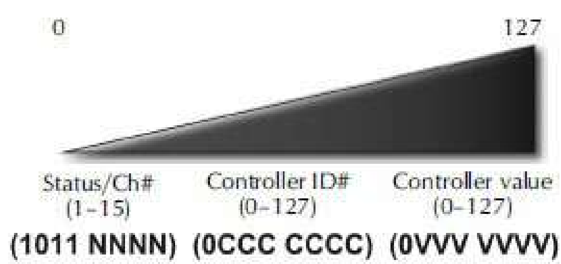


Figure 33-14. Control Change message: (a) byte structure; (b) control messages can be varied in real time or under automation using a number of input methods.

33.4.4 Explanation of Control ID Parameters

As you can see in Fig. 33-14, the 2nd byte of the Control Change message is used to denote the *controller ID number*. This all-important value is used to specify which of the device's program or performance parameters are to be addressed.

The following section details the general categories and conventions for assigning controller ID numbers to an associated parameter (as specified by the 1995 update of the MMA (MIDI Manufacturers Association, www.midi.org). An overview of these controllers can be seen in Table 33-1. This is definitely an important table to earmark, because these numbers will be an important guide toward knowing and/or finding the right ID number that can help you on your path toward finding that perfect variable for making it sound right.

Table 33-1. Listing of Controller ID Numbers, Outlining Both the Defined Format and Convention and Controller Assignments

Control Number	**Parameter**
14 Bit Controllers Coarse/MSB (most significant bit)	
0	Bank Select 0–127 MSB
1	Modulation Wheel or Lever 0–127 MSB
2	Breath Controller 0–127 MSB
3	Undefined 0–127 MSB
4	Foot Controller 0–127 MSB
5	Portamento Time 0–127 MSB
6	Data Entry MSB 0–127 MSB
7	Channel Volume (formerly Main Volume) 0–127 MSB
8	Balance 0–127 MSB
9	Undefined 0–127 MSB
10	Pan 0–127 MSB

11	Expression Controller 0–127 MSB
12	Effect Control 1 0–127 MSB
13	Effect Control 2 0–127 MSB
14	Undefined 0–127 MSB
15	Undefined 0–127 MSB
16–19	General Purpose Controllers 1–4 0–127 MSB
20–31	Undefined 0–127 MSB
14-bit Controllers Fine/LSB (least significant bit)	
32	LSB for Control 0 (Bank Select) 0–127 LSB
33	LSB for Control 1 (Modulation Wheel or Lever) 0–127 LSB
34	LSB for Control 2 (Breath Controller) 0–127 LSB
35	LSB for Control 3 (Undefined) 0–127 LSB
36	LSB for Control 4 (Foot Controller) 0–127 LSB
37	LSB for Control 5 (Portamento Time) 0–127 LSB
38	LSB for Control 6 (Data Entry) 0–127 LSB
39	LSB for Control 7 (Channel Volume, formerly Main Volume) 0–127 LSB
40	LSB for Control 8 (Balance) 0–127 LSB
41	LSB for Control 9 (Undefined) 0–127 LSB
42	LSB for Control 10 (Pan) 0–127 LSB
43	LSB for Control 11 (Expression Controller) 0–127 LSB
44	LSB for Control 12 (Effect control 1) 0–127 LSB
45	LSB for Control 13 (Effect control 2) 0–127 LSB
46–47	LSB for Control 14–15 (Undefined) 0–127 LSB
48–51	LSB for Control 16–19 (General Purpose Controllers 1–4) 0–127 LSB
52–63	LSB for Control 20–31 (Undefined) 0–127 LSB
7-bit Controllers	
64	Damper Pedal On/Off (Sustain) <63 off, >64 on
65	Portamento On/Off <63 off, >64 on
66	Sustenuto On/Off <63 off, >64 on

67	Soft Pedal On/Off <63 off, >64 on
68	Legato Footswitch <63 Normal, >64 Legato
69	Hold 2 <63 off, >64 on
70	Sound Controller 1 (Default: Sound Variation) 0–127 LSB
71	Sound Controller 2 (Default: Timbre/Harmonic Intensity) 0–127 LSB
72	Sound Controller 3 (Default: Release Time) 0–127 LSB
73	Sound Controller 4 (Default: Attack Time) 0–127 LSB
74	Sound Controller 5 (Default: Brightness) 0–127 LSB
75	Sound Controller 6 (Default: Decay Time—see MMA RP-021) 0–127 LSB
76	Sound Controller 7 (Default: Vibrato Rate—see MMA RP-021) 0–127 LSB
77	Sound Controller 8 (Default: Vibrato Depth—see MMA RP–021) 0–127 LSB
78	Sound Controller 9 (Default: Vibrato Delay—see MMA RP–021) 0-127 LSB
79	Sound Controller 10 (Default undefined—see MMA RP–021) 0-127 LSB
80–83	General Purpose Controller 5–8 0–127 LSB
84	Portamento Control 0–127 LSB
85–90	Undefined
91	Effects 1 Depth (Default: Reverb Send Level) 0–127 LSB
92	Effects 2 Depth (Default: tremolo Level) 0–127 LSB
93	Effects 3 Depth (Default: Chorus Send Level) 0–127 LSB
94	Effects 4 Depth (Default: Celesta [Detune] Depth) 0–127 LSB
95	Effects 5 Depth (Default: Phaser Depth) 0–127 LSB
Parameter Value Controllers	
96	Data Increment (Data Entry +1)
97	Data Decrement (Data Entry recorded at the level1)

98	Non-Registered Parameter Number (NRPN): LSB 0–127 LSB
99	Non-Registered Parameter Number (NRPN): MSB 0–127 MSB
100	Registered Parameter Number (RPN): LSB* 0–127 LSB
101	Registered Parameter Number (RPN): MSB* 0–127 MSB
102–119	Undefined
Reserved for Channel Mode Messages	
120	All Sound Off 0
121	Reset All Controllers
122	Local Control On/Off 0 off, 127 on
123	All Notes Off
124	Omni Mode Off (+ all notes off)
125	Omni Mode On (+ all notes off)
126	Poly Mode On/Off (+ all notes off)
127	Poly Mode On (+ mono off +all notes off)

33.4.5 System Messages

As the name implies, System messages are globally transmitted to every MIDI device in the MIDI chain. This is accomplished because MIDI channel numbers aren't addressed within the byte structure of a System message. Thus, any device will respond to these messages, regardless of its MIDI channel assignment. The three System message types are System-Common messages, System Real-Time messages, and System-Exclusive messages.

System-Common messages are used to transmit MIDI timecode, song position pointer, song select, tune request and end-of-exclusive data messages throughout the MIDI system or 16 channels of a specified MIDI port:

- **MIDI timecode (MTC) messages**. Provide a cost-effective and easily implemented way to translate SMPTE (a standardized synchronization timecode) into an equivalent code that conforms to the MIDI 1.0 spec. It allows time-based codes and commands to be distributed throughout the MIDI chain in a cheap, stable and easy-to-implement way. MTC Quarter-Frame messages are transmitted and recognized by MIDI devices that can understand and execute MTC commands. A grouping of eight quarter frames is used to denote a complete timecode address (in hours, minutes, seconds, and frames), allowing the SMPTE address to be updated every two frames.
- **Song Position Pointer (SPP) messages**. Allow a sequencer or drum machine to be synchronized to an external source (such as a tape machine) from any measure position within a song. This complex timing protocol isn't commonly used, because most users and design layouts currently favor MTC.
- **Song Select messages**. Use an identifying song ID number to request a specific song from a sequence or controller source. After being selected, the song responds to MIDI Start, Stop and Continue messages.
- **Tune Request messages**. Used to request that an equipped MIDI instrument initiate its internal tuning routine.
- **End of Exclusive (EOX) messages**. Indicate the end of a System-Exclusive message.

System Real-Time messages provide the precise timing element required to synchronize all of the MIDI devices in a connected system. To avoid timing delays, the MIDI specification allows System Real-Time messages to be inserted at any point in the data stream, even between other MIDI messages.

- **Timing Clock messages**. The MIDI Timing Clock message is transmitted within the MIDI datastream at various resolution rates. It is used to synchronize the internal timing clocks of each MIDI device within the system and is transmitted in both the start and stop modes at the currently defined tempo rate. In the early days of MIDI, these rates (which are measured in pulses per quarter note [ppq]) ranged from 24 to 128 ppq; however, continued advances in technology have brought these rates up to 240, 480 or even 960 ppq.
- **Start messages**. Upon receipt of a timing clock message, the MIDI Start command instructs all connected MIDI devices to begin playing from their internal sequences' initial start point. Should a program be in midsequence, the start command will reposition the sequence back to its beginning, at which point it will begin to play.
- **Stop messages**. Upon receipt of a MIDI Stop command, all devices within the system will stop playing at their current position point.
- **Continue messages**. After receiving a MIDI Stop command, a MIDI Continue message will instruct all connected devices to resume playing their internal sequences from the precise point at which it was stopped.
- **Active Sensing messages**. When in the Stop mode, an optional Active Sensing message can be transmitted throughout the MIDI datastream every 300 ms. This instructs devices that can recognize this message that they are still connected to an active MIDI data stream.
- **System Reset messages**. A System Reset message is manually transmitted in order to reset a MIDI device or instrument back to its initial power-up default settings (commonly mode 1, local

control on and all notes off).

The System-Exclusive (Sys-Ex) message allows MIDI manufacturers, programmers and designers to communicate customized MIDI messages between MIDI devices. The purpose of these messages is to give manufacturers, programmers and designers the freedom to communicate any device-specific data of an unrestricted length, as they see fit. Most commonly, sys-ex data are used for the bulk transmission and reception of program/patch data and sample data, as well as real-time control over a device's parameters. The transmission format of a sys-ex message, Fig. 33-15, as defined by the MIDI standard, includes a sys-ex status header, manufacturer's ID number, any number of sys-ex data bytes and an EOX byte. When a sys-ex message is received, the identification number is read by a MIDI device to determine whether or not the following messages are relevant. This is easily accomplished by the assignment of a unique 1- or 3-byte ID number to each registered MIDI manufacturer and make. If this number doesn't match the receiving MIDI device, the subsequent data bytes will be ignored. Once a valid stream of sys-ex data has been transmitted, a final EOX message is sent, after which the device will again begin to respond normally to incoming MIDI performance messages.

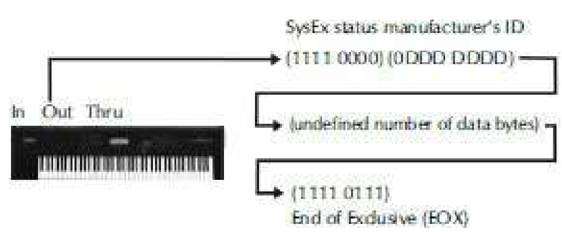


Figure 33-15. System-exclusive ID data and controller format.

In actual practice, the general idea behind sys-ex is that it uses MIDI messages to transmit and receive program, patch and sample data or real-time parameter information between devices. It is sort of like having an instrument or device that's a musical chameleon. One moment it can be configured with a certain set of sound patches and setup data and then, after it receives a new sys-ex data dump, you could easily end up with an instrument that is literally full of new and exciting (or not-so-exciting) sounds and settings. Here are a few examples of how sys-ex can be put to good use:

- **Transmitting patch data between synths**. Sys-ex can be used to transmit patch and overall setup data between synths of identical make and (most often) model. Let's say that we have a Brand X Model Z synthesizer, and, as it turns out, you have a buddy across town that also has a Brand X Model Z. That's cool, except your buddy has a completely different set of sound patches in her synth… and you want them! Sys-ex to the rescue! All you need to do is go over and transfer your buddy's patch data into your synth (to make life easier, make sure you take the instruction manual along).
- **Backing up your current patch data**. This can be done by transmitting a sys-ex dump of your synth's entire patch and setup data to disk, to a sys-ex utility program (often shareware) or to your DAW/MIDI sequencer. This is important: Back up your factory preset or current patch data before attempting a sys-ex dump! If you forget and download a sys-ex dump, your previous settings will be lost until you contact the manufacturer, download the dump from their website or take your synth back to your favorite music store to reload the data.
- **Getting patch data from the Web**. One of the biggest

repositories of sys-ex data is the Internet. To surf the Web for sys-ex patch data, all you need to do is log on to your favorite search engine site and enter the name of your synth. You will probably be amazed at how many hits will come across the screen, many of which are chock-full of sys-ex dumps that can be downloaded into your synth.

- **Varying sys-ex controller or patch data in real time**. Patch editors or hardware MIDI controllers can be used to vary system and sound-generating parameters, in real time. Both of these controller types can ease the job of experimenting with parameter values or changing mix moves by giving you physical or on-screen controls that are often more intuitive and easier to deal with than programming electronic instruments that'll often leave you dangling in cursor and 3-inch LCD screen hell.

Over the last several years, it seems that a single unified standard has begun to emerge from the fray that is so simple that it's amazing it wasn't universally adopted from the start. This system simply records a sys-ex dump as data on a single MIDI track file. Before recording a dump to a sequencer track, you may need to consult the manual to make sure that sys-ex filtering is turned off. Once this is done, simply place the track into record mode, initiate the dump and save the track in an all-important sys-ex dump directory. Using this approach, it would also be possible to:

- Import the appropriate sys-ex dump track (or set of tracks) into the current working session so as to automatically program the instruments before the sequence is played back.
- Import the appropriate sys-ex dump track (or set of tracks) into separate MIDI tracks that can be muted or unassigned. Should

the need arise, the track(s) can be activated and/or assigned in order to dump the data into the appropriate instruments.

33.5 Electronic Instruments

Since their inception in the early 1980s, MIDI-based electronic instruments have played a central and important role in the development of music technology and production. These devices (which fall into almost every instrument category), along with the advent of cost-effective analog and digital audio recording systems, have probably been the most important technological advances to shape the industry into what it is today. In fact, the combination of hardware and newer software plug-in technologies has turned the personal project studio into one of the most important driving forces behind modern-day music production.

The following is a sample listing of the many hardware MIDI instrument types that are currently available on the market.

33.5.1 The Synth

A synthesizer (or synth) is an electronic instrument that uses multiple sound generators, filters and oscillator blocks to create complex waveforms that can be combined into countless sonic variations. These synthesized sounds have become a basic staple of modern music and range from those that sound "cheesy" to ones that realistically mimic traditional instruments… and all the way to those that generate otherworldly, ethereal sounds that literally defy classification.

Synthesizers generate sounds using a number of different technologies or program algorithms. Examples of these include:

- **FM synthesis**. This technique generally makes use of at least two signal generators (commonly referred to as “operators”) to create and modify a voice. It often does this by generating a signal that modulates or changes the tonal and amplitude characteristics of a base carrier signal. More sophisticated FM synths use up to four or six operators per voice, each using filters and variable amplifier types to alter a signal’s characteristics.
- **Wavetable synthesis**. This technique works by storing small segments of digitally sampled sound into a memory media. Various sample-based and synthesis techniques make use of looping, mathematical interpolation, pitch shifting and digital filtering to create extended and richly textured sounds that use a surprisingly small amount of sample memory, allowing hundreds if not thousands of samples and sound variations to be stored in a single device or program.
- **Additive synthesis**. This technique makes use of combined waveforms that are generated, mixed and varied in level over time to create new timbres that are composed of multiple and complex harmonics, which vary over time. Subtractive synthesis makes extensive use of filtering to alter and subtract overtones from a generated waveform (or series of waveforms).

Of course, synths come in all shapes and sizes and use a wide range of patented synthesis techniques for generating and shaping complex waveforms, in a polyphonic fashion using 16, 32 or even 64 simultaneous voices, Fig. 33-16. In addition, many synths often include a percussion section that can play a full range of drum and “perc” sounds, in a number of styles. Reverb and other basic effects are also commonly built into the architecture of these devices, reducing the need for using extensive outboard effects when being played on-stage or out of the box. Speaking of “out of the box,” a number of synth systems are referred to as being “workstations.” Such beasties are designed (at least in theory) to handle many of your basic production needs (including basic sound generation, MIDI sequencing, effects, etc.) all in one neat little package.

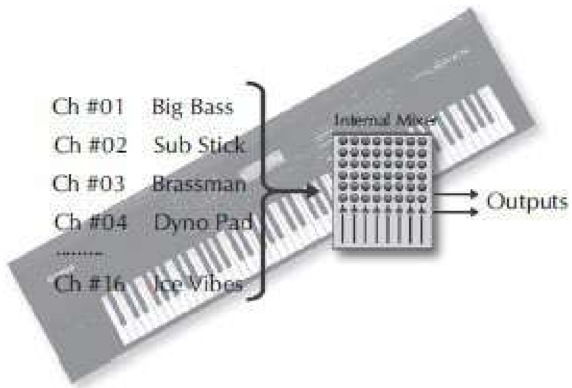


Figure 33-16. Multitimbral instruments are virtual bands-in-a-box that can simultaneously generate multiple patches, each of which can be assigned to its own MIDI channel.

33.5.2 The Sampler

A sampler, Fig. 33-17, is a hardware or software system that can import and manipulate data within random access memory (RAM). Once audio has been sampled or loaded into RAM (from disk, disc or diskette), segments of sampled audio can then be edited, transposed, processed and played in a polyphonic, musical fashion. In short, a sampler can be thought of as a digital audio memory device that lets you record, edit and reload samples into RAM. Once loaded, these sounds (whose length and complexity are often only limited by memory size and your imagination) can be looped, modulated, filtered and amplified (according to user or factory setup parameters) in a way that allows the waveshapes and envelopes to be modified. Signal processing capabilities, such as basic editing, looping, gain changing, reverse, sample-rate conversion, pitch change and digital mixing can also be easily applied to change the sounds in an almost infinite number of ways.

Figure 33-17. Kontact Virtual Sampler. Courtesy of Native Instruments GmbH, www.native-instruments.com.

Samplers derive their sounds from recorded and/or imported audio data stored as digital audio data within a personal computer or other media storage device. Using its digital signal processing (DSP) capabilities, most software samplers are able to store and access samples within the internal memory to:

- Import previously recorded sound files (often in .wav, .aif and other common formats).
- Edit and loop sounds into a usable form.
- Vary envelope parameters (e.g., dynamics over time).
- Vary processing parameters.
- Save the edited sample performance setup as a file for later recall.

A sample can be played according to the standard Western musical scale (or any other scale, for that matter) by altering the playback sample rate over the controller's note range. For example, pressing a low-pitched key on the keyboard will cause the sample to be played back at a lower sample rate, while pressing a high-pitched one will cause the sample to be played back at rates that would put Mickey Mouse to shame. By choosing the proper sample rate ratios, these sounds can be polyphonically played (whereby multiple notes are sounded at once) at pitches that correspond to standard musical chords and intervals.

A sampler (or synth) with a specific number of voices (e.g., 64 voices) simply means that up to 64 notes can be simultaneously played on a keyboard at any one time. Each sample in a multiple-voice system can be assigned across a performance keyboard, using a process known as splitting or mapping. In this way, a sound can be assigned to play across the performance surface of a controller over a range of notes, known as a zone. In addition to grouping samples into various zones, velocity can enter into the equation by allowing multiple samples to be layered across the same keys of a controller, according to how soft or hard they are played. For example, a single key might be layered so that pressing the key lightly would reproduce a softly recorded sample, while pressing it harder would produce a louder sample with a sharp, percussive attack. In this way, mapping can be used to create a more realistic instrument or wild set of soundscapes that change not only with the played keys but with different velocities as well. Most samplers have extensive edit capabilities that allow the sounds to be modified in much the same way as a synthesizer, using such modifiers as:

- Velocity.
- Panning.

- Expression (modulation and user control variations).
- Low-frequency oscillation (LFO).
- Attack, delay, sustain and release (ADSR) and other envelope processing parameters.
- Keyboard scaling.
- Aftertouch.

Many sampling systems will often include such features as integrated signal processing, multiple outputs (offering isolated channel outputs for added live mixing and signal processing power or for recording individual voices to a multitrack recording system) and integrated MIDI sequencing capabilities.

The MIDI Keyboard Controller. Although a MIDI controller is not an instrument, it is a device that is expressly designed to control other devices (be they for musical, light or mechanical control) within a connected MIDI system. These devices contain no internal tone generators or sound-producing elements but often include a high-quality control surface and a wide range of controls for handling control, trigger and device-switching events. Since controllers have become an integral part of music production and are available in many incarnations to control and emulate many types of musical instrument types, don’t be surprised to find controllers of various incarnations popping up all over this book and within electronic music production. These devices commonly include such controls as:

- Music keyboard surface.
- Variable parameter controls.
- Fader, mixing and transport controls.
- Switching controls.

- Tactile trigger and control surfaces.

As you might imagine, controllers vary widely in the number of features that are offered, Fig. 33-18. For starters, the number of keys can vary from the sporty, portable 25-key models to those having 49 and 61 keys and all the way to the full 88-key models that can play the entire range of a full-size grand piano. The keys may be fully or partially weighted and in a number of models the keys might be much smaller than the full piano key size often making a performance a bit difficult. Beyond the standard pitch and modulation wheels (or similar-type controller), the number of options and general features is up to the manufacturers. With the increased need for control over electronic instruments and music production systems, many model types offer up a wide range of physical controllers for varying an ever-widening range of expressive parameters.

Figure 33-18. Novation 25SL USB MIDI Controller. Courtesy Novation Digital Music Systems, Ltd., www.novationmusic.com.

33.5.3 The Drum Machine

In its most basic form, the drum machine uses ROM-based, prerecorded waveform samples to reproduce high-quality drum sounds from its internal memory. These factory-loaded sounds often include a wide assortment of drum sets, percussion sets and rare, wacky percussion hits, and effected drum sets (e.g., reverberated, gated). Who knows you might even encounter scream hits by the venerable King of Soul, James Brown. These prerecorded samples can be assigned to a series of playable keypads that are generally located on the machine's top face, providing a straightforward controller surface that often sports velocity and aftertouch dynamics. Sampled voices can be assigned to each pad and edited using control parameters such as tuning, level, output assignment and panning position.

Because of new cost-effective technology, many drum machines now include basic sampling technology, which allows sounds to be imported, edited and triggered directly from the box, Fig. 33-19. As with the traditional "beat box," these samples can be easily mapped and played from the traditionally styled surface trigger pads. Of course, virtual software drum and groove machines are part of the present-day landscape and can be used in a stand-alone, plug-in and rewired production environment.

Figure 33-19. Alesis SR-18 stereo drum machine. Courtesy Alesis, www.alesis.com.

The MIDI Drum Controller

MIDI drum controllers are used to translate the voicing and expressiveness of a percussion performance into MIDI data. These devices are great for capturing the feel of a live performance, while giving you the flexibility of automating or sequencing a live event. These devices range from having larger pads and trigger points on a larger performance surface to drum-machine-type pads/buttons. Coming under the "Don't try this at home" category, these controller pads are generally too small and not durable enough to withstand drumsticks or mallets. For this reason, they're generally played with the fingers. It's long been a popular misconception that MIDI drum controllers have to be expensive. This simply isn't true. There are quite a few instruments that are perceived by many to be toys but, in fact, are fully implemented with MIDI and can be easily used as a controller. A few of the ways to perform and sequence percussion include:

- **Drum machine button pads**. One of the most straightforward of all drum controllers is the drum button pad design that's built into most drum machines, portable percussion controllers, Fig. 33-20, and certain keyboard controllers. By calling up the desired setup and voice parameters, these small footprint triggers let you go about the business of using your fingers to do the walking through a performance or sequenced track.
- **The keyboard as a percussion controller**. Since drum machines respond to external MIDI data, probably the most commonly used device for triggering percussion and drum voices is a standard MIDI keyboard controller. One advantage of playing percussion sounds from a key-board is that sounds can be triggered more quickly because the playing surface is designed for

fast finger movements and doesn't require full hand/wrist motions. Another advantage is its ability to express velocity over the entire range of possible values (0B127), instead of the limited number of velocity steps that are available on certain drum pad models.

- **Drum pad controllers**. In more advanced MIDI project studios or live stage rigs, it's often necessary for a percussionist to have access to a playing surface that can be played like a real instrument. In these situations, a dedicated drum pad controller would be better for the job. Drum controllers vary widely in design. They can be built into a single, semi-portable case, often having between six and eight playing pads, or the trigger pads can be individual pads that can be fitted onto a special rack, traditional drum floor stand or drum set.
- **MIDI drums**. Another way to MIDI-fy an acoustic drum is through the use of trigger technology. Put simply, triggering is carried out by using a transducer pickup (such as a mic or contract pickup) to change the acoustic energy of a percussion or drum instrument into an electrical voltage. Using a MIDI trigger device, Fig. 33-20, a number of pickup inputs can be translated into MIDI so as to trigger programmed sounds or samples from an instrument for use on stage or in the studio.

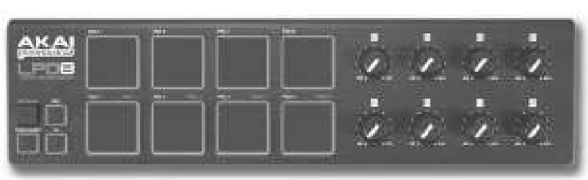


Figure 33-20. Akai LPD8 portable USB Drum Controller. Courtesy Akai Professional, LP, www.akaipro.com.

Other MIDI Instrument and Controller Types. There are literally tons of instruments and controller types out there that are capable of translating a performance or general body movements into MIDI. You'd be surprised what you'll find searching the Web for wild and wacky controllers both those that are commercially available and those that are made by soldering iron junkies. A few of the more traditional controllers include MIDI guitars and basses, wind controllers, MIDI vibraphones… the list goes on.

33.6 Sequencing

Apart from our computers, DAWs and venerable electronic instruments, one of the most important tools that can be found in the modern-day project studio is the MIDI sequencer. Basically, a sequencer is a digital device or software application that's used to record, edit and output MIDI messages in a sequential fashion. These messages are generally arranged in a track-based format that follows the modern production concept of having instruments (and/or instrument voices) located on separate tracks. This traditional interface makes it easy for us humans to view MIDI data as tracks on a digital audio workstation (DAW) or analog tape recorder that follow along a straightforward linear time line.

These tracks contain MIDI-related performance and control events that are made up of such channel and system messages as Note-On, Note-Off, Velocity, Modulation, Aftertouch and Program/Continuous Controller messages. Once a performance has been recorded into a sequencer's memory, these events can be graphically arranged and edited into a musical performance. The data can then be saved as a file or within a DAW session and recalled at any time, allowing the data to be played back in its originally recorded or edited order.

Because most sequencers are designed to function in a way that is similar to the multitrack tape recorder, they give us a familiar operating environment in which each instrument, set of layered instruments or controller data can be recorded onto separate, synchronously arranged tracks. Like its multitrack cousin, each track can be re-recorded, erased, copied and varied in level during playback. However, since the recorded data is digital in nature, a MIDI sequencer is far more flexible in its editing speed and control in that it offers all the cut-and-paste, signal processing and channel-routing features that we have come to expect from a digital production application.

By far, the most common sequencer type is the software sequencing program, Fig. 33-21. These programs run on all types of personal and laptop computers and take advantage of the hardware and software versatility that only a computer can offer in the way of speed, hardware flexibility, memory management, signal routing and digital signal processing. However, a type of keyboard synth and sampler system known as a keyboard workstation will often include much of the necessary production hardware that is required for music production, including effects and an integrated hardware sequencer. These systems have the advantage of letting you take your instrument and sequencer on the road without having to drag your whole system along.

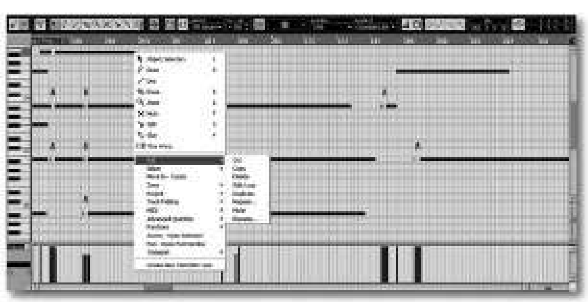


Figure 33-21. MIDI edit window within Cubase audio production software. Courtesy Steinberg Media Technologies GmbH, a division of Yamaha Corporation, www.steinberg.net.

33.6.1 A Basic Introduction to Sequencing

When dealing with any type of sequencer, one of the most important concepts to grasp is that these devices don't store sound directly; instead, they encode MIDI messages that instruct instruments as to what note is to be played, over what channel, at what velocity and at what, if any, optional controller values. In other words, a sequencer simply stores command instructions that follow in a sequential order. These instructions tell instruments and/or devices how their voices are to be played or controlled. This means that the amount of encoded data is a great deal less memory intensive than its digital audio or digital video recording counterparts. Because of this, the data overhead that's required by MIDI is very small, allowing a computer-based sequencer to work simultaneously with the playback of digital audio tracks, video images, Internet browsing, etc., all without unduly slowing down the computer's CPU. For this reason, MIDI and the MIDI sequencer provide a media environment that plays well with other computer-based production media.

33.6.1.1 Recording

Commonly, a MIDI sequencer is an application within a digital production workspace for creating personal compositions in environments that range from the bedroom to more elaborate professional and project studios. Whether hardware or software based, most sequencers use a working interface that is roughly designed to emulate a traditional multitrack-based environment. A tape-like set of transport controls lets us move from one location to the next using standard Play, Stop, Fast Forward, Rewind and Record command buttons. Beyond using the traditional Record-Enable button to select the track or tracks that we want to record onto, all we need to do is select the MIDI input (source) port, output (destination) port, MIDI channel, instrument patch and other setup requirements. Then press the record button and begin laying down the track.

Once you’ve finished laying down a track, you can jump back to the beginning of the recorded passage and listen to it. From this point, you could then “arm” (a term used to denote placing a track into the record-ready mode) the next track and go about the process of laying down additional tracks until a song begins to form.

When beginning a MIDI session, one of the first aspects to consider is the tempo and time signature. The beats-per-minute (bpm) value will set the general tempo speed for the overall session. This is important to set at the beginning of the session, so as to lock the overall “bars and beats” timing elements to this initial speed that’s often essential in electronic music production. The tempo of a MIDI production can often be easily changed without worrying about changing the program’s pitch or real-time control parameters. In short, once you know how to avoid potential conflicts and pitfalls, tempo variations can be made after the fact with relative ease.

Although only one MIDI track is commonly recorded at a time, most mid- and professional-level sequencers allow us to record multiple tracks at one time. This feature makes it possible for a multi-trigger instrument or for several performers to record to a sequence in one, live pass. For example, such an arrangement would allow for each trigger pad of a MIDI drum controller to be recorded to its own track (with each track being assigned to a different MIDI channel on a single port). Alternatively, several instruments of an on-stage electronic band could be captured to a sequence during a live performance and then laid into a DAW session for the making of an album project.

33.6.1.2 Editing

One of the more important features that a sequencer (or MIDI track within a DAW) has to offer is its ability to edit sequenced tracks or blocks within a track. Of course, these editing functions and capabilities often vary from one sequencer to another. The main track window of a sequencer or MIDI track on a DAW is used to display such track information as the existence of track data, track names, MIDI port assignments for each track, program change assignments, volume controller values and other transport commands.

Depending on the sequencer, the existence of MIDI data on a particular track at a particular measure point (or over a range of measures) is indicated by the highlighting of a track range in a way that's extremely visible. By navigating around the various data display and parameter boxes, it's possible to use cut-and-paste and/or edit techniques to vary note values and parameters for almost every facet of a section or musical composition. For example, let's say that we really screwed up a few notes when laying down an otherwise killer bass riff. With MIDI, fixing the problem is totally a no-brainer. Simply highlight each fudged note and drag it to its proper note location—we can even change the beginning and endpoints in the process. In addition, tons of other parameters can be changed, including velocity, modulation and pitch bend, note and song transposition, quantization and humanizing (factors that eliminate or introduce human timing errors that are generally present in a live performance), in addition to full control over program and continuous controller messages… the list goes on.

33.6.1.3 Playback

Once a sequence is composed and saved to disk, all of the sequence tracks can be transmitted through the various MIDI ports and channels to the instruments or devices to make music, create sound effects for film tracks or control device parameters in real time. Because MIDI data exists as encoded real-time control commands and not as audio, you can listen to the sequence and make changes at any time. You could change the patch voices, alter the final mix or change and experiment with such controllers as pitch bend or modulation even change the tempo and key signature.

In short, this medium is infinitely flexible in the number of versions that can be created, saved, folded, spindled and mutilated until you have arrived at the overall sound and feel you want. Once done, you'll have the option of using the data for live performance or mixing the tracks down to a final recorded media, either in the studio or at home.

Almost all DAW and sequencer types will let you mix a sequence in the MIDI domain using various controller message types. This is usually done by creating a software interface that incorporates these controls into a virtual on-screen mixer environment that often integrates with the main DAW mix screen. Instead of directly mixing the audio signals that make up a sequence, these controls are able to directly access such track controllers as Main Volume (controller 7), Pan (controller 10), and Balance (controller 8), most often in an environment that completely integrates into the workstation's overall mix controls. Since the mix-related data is simply MIDI controller messages, an entire mix can be easily stored within a sequence file. Therefore, even with the most basic sequencer, you'll be able to mix and remix your sequences with complete automation and total settings recall whenever a new sequence is opened. As is almost always the case with a DAW's audio and MIDI graphical user interface (GUI), the controller and mix interface will almost always have moving faders and variable controls, Fig. 33-22.

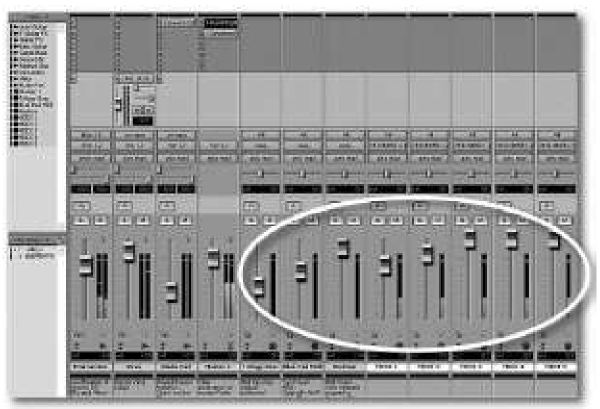


Figure 33-22. MIDI tracks can be added into the mix for real-time parameter control over hardware and/or software devices. Courtesy Avid Technology, Inc., www.avid.com.

33.7 Groove Tools

The expression "getting into the groove" of a piece of music often refers to a feeling that's derived from the underlying foundation of music: rhythm. With the introduction and maturation of MIDI and digital audio, new and wondrous tools have made their way into the mainstream of music production. These tools can help us to use technology to forge, fold, mutilate and create compositions that make direct use of rhythm and other building blocks of music through the use of looping technology. Of course, the cyclic nature of loops can be repeat–repeat–repetitive in nature, but new toys and techniques for looping can inject added flexibility, control and real-time processing into a project in ways that can be used by artists in wondrously expressive ways.

The basic idea behind groove-based tools rests with tempo matching, the idea that various rhythms, grooves, pads and any other imaginable sounds of various tempos, lengths (and often musical keys) can be artfully crafted together into a single, working song session.

Because groove-based tools often deal with rhythms and cyclic-based loops that are pulled from various musical sources, the factors that need to be managed are:

- Sync.
- Tempo and length.
- Time and pitch change techniques.

The aspect of sync relates to the fact that the various loops in a groove project will need to sync up with each other (or in multiple lengths and timings of each other). It almost goes without saying that multiple loops that are successively or simultaneously triggered must have a synchronous timing relationship with one another otherwise, it's a jumbled mess of sound.

Another relationship relates to the aspect of tempo. Just as sync is imperative, it's also necessary for the files to be adjusted in pitch (resampling) and/or length (time stretching), so that they precisely match the currently selected tempo (or are programmed to be a relative multiple of the session's tempo).

A final aspect is associated with time and pitch change techniques. This is the process of altering a sound file (often which is rhythmically repetitive and short in length) to match the current session tempo and to synchronously align them within the software by using variable sample- and pitch-shifting techniques. Using these basic digital signal processing (DSP) tools, it's possible to alter a sound file's duration (varying the length of a program by raising or lowering its playback sample rate) or to alter its relative pitch (either up or down). In this way, these loops can be matched up or musically combined by using any of three possible time and pitch change combinations:

- **Time change**. A program's length can be altered without affecting its pitch.
- **Pitch change**. A program's length can remain the same while pitch is shifted either up or down.
- **Both**. Both a program's pitch and length can be altered using resampling techniques.

By setting the loop program to a master tempo (or a specific tempo at that point in the song), an audio segment or file can be imported, examined as to sample rate or length and then recalculated to a new relative tempo that matches the current session tempo. Voilà! We now have a defined segment of audio that matches the tempo of all of the other segments, allowing it to play and interact in relative sync with the other defined segments and/or loop files.

Loop-based audio software

Loop-based audio editors are groove-driven music programs, Figs. 33-23 and 33-24, that are designed to let you drag and drop prerecorded or user-created loops and audio tracks into a graphic multitrack production interface. At their basic level, these programs differ conceptually from their traditional DAW counterpart in that the pitch- and time-shift architecture is so variable and dynamic that, even after the basic rhythmic, percussive and melodic grooves have been created, their tempo, track patterns, pitch, session key, etc., can be quickly and easily changed at any time. With the help of custom, royalty-free loops (available from many manufacturer and third-party companies), users can quickly and easily experiment with setting up grooves, backing tracks and creating a sonic ambience by simply dragging the loops into the program's main sound file view, where they can be arranged, edited, processed, saved and exported.

Figure 33-23. Apple GarageBand. Courtesy Apple Computers, Inc., www.apple.com.

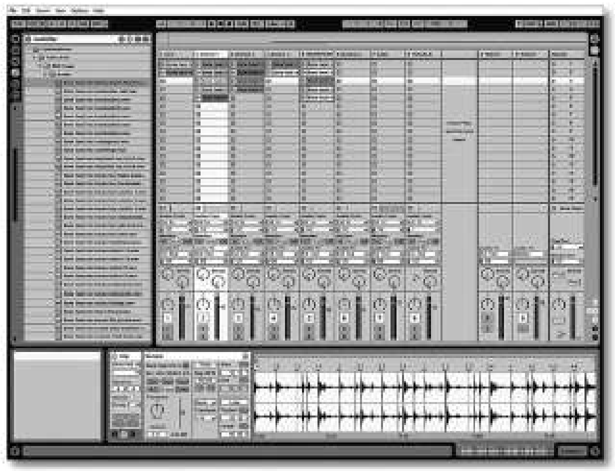


Figure 33-24. Ableton Live performance audio workstation in arrangement view. Courtesy Ableton AG, www.ableton.com.

One of the most interesting aspects of the loop-based editor is its ability to match the tempo of a specially programmed loop to the tempo of the current session. Amazingly enough, this process isn't that difficult to perform, because the program extracts the length, native tempo, and pitch information from the imported files and, using various digital time- and pitch-change techniques, adjusts the loop to fit the native time and pitch parameters of the current session. This means that loops of various tempos and musical keys can be automatically adjusted in length and pitch so as to fit in time with previously existing loops... just drag, drop and go!

33.8 Music Printing Programs

Over the past few decades, the field of transcribing musical scores and arrangements has been strongly affected by both the computer and MIDI technology. This process has been greatly enhanced through the use of newer generations of computer software that make it possible for music notation data to be entered into a computer either manually (by placing the notes onto the screen via keyboard or mouse movements), by direct MIDI input or by sheet music scanning technology. Once entered, these notes can be edited in an on-screen environment that lets you change and configure a musical score or lead sheet using standard cut-and-paste editing techniques. In addition, most programs allow the score data to be played directly from the score by electronic instruments via MIDI. A final and important program feature is their ability to quickly print out hard copies of a score or lead sheets in a wide number of print formats and styles.

A music printing program (also known as a music notation program) lets you enter musical data into a computerized score in a number of manual and automated ways (often with varying degrees of complexity and ease). Programs of this type, Fig. 33-25, offer a wide range of notation symbols and type styles that can be entered either from a computer keyboard or mouse. In addition to entering a score manually, most music transcription programs will generally accept MIDI input, allowing a part to be played directly into a score. This can be done in real time (by playing a MIDI instrument/controller or finished sequence into the program) or in step time (by entering the notes of a score one note at a time from a MIDI controller) or by entering a standard MIDI file into the program (which uses a sequenced file as the notation source).

Figure 33-25. Finale music composition, arranging and printing program. Courtesy MakeMusic, Inc., www.finalemusic.com.

In addition to dedicated music printing programs, most DAW or sequencer packages will often include a basic music notation application that allows the sequenced data within a track or defined region to be displayed and edited directly within the program, Fig. 33-26, from which it can be printed in a limited score-like fashion. However, a number of high-level workstations offer scoring features that allow sequenced track data to be notated and edited in a professional fashion into a fully printable music score.

As you might expect, music printing programs will often vary widely in their capabilities, ease of use and offered features. These differences often center around the graphical user interface (GUI), methods for inputting and editing data, the number of instrumental parts that can be placed into a score, the overall selection of musical symbols, the number of musical staves (the lines that music notes are placed onto) that can be entered into a single page or overall score, the ability to enter text or lyrics into a score, etc. As with most programs that deal with artistic production, the range of choices and general functionality reflect the style and viewpoints of the manufacturer, so care should be taken when choosing a professional music notation program, to see which one would be right for your personal working style.

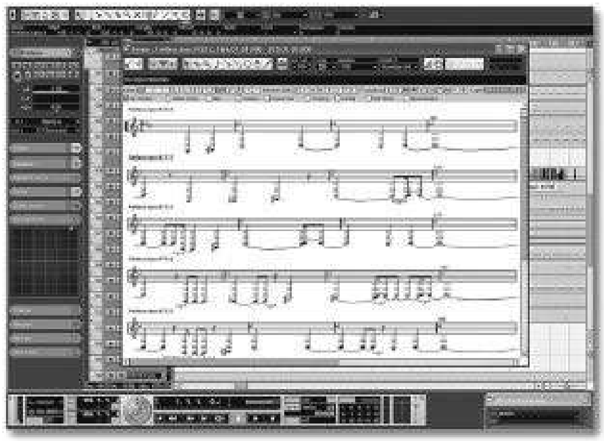


Figure 33-26. Steinberg Cubase/Nuendo score window. Courtesy of Steinberg Media Technologies GmbH, a division of Yamaha Corporation, www.steinberg.net.

33.9 Multimedia and the Web

It's no secret that modern-day computers, portable devices, game-stations and even televisions have gotten faster, sleeker, more touchable and sexier in their overall design. In addition to their ability to act as a multifunctional production workhorse, one of the crowning achievements of modern work and entertainment devices is the degree of media and networking integration that has made its way into our collective consciousness in a way that has come to be universally known as the household buzzword multimedia.

The combination of working and playing with multimedia has found its way into modern media and computer culture through the use of various hardware and software systems that work in a multitasking environment and combine to bring you an experience that seamlessly involves such media types as:

- Audio and music.
- Video and video streaming.
- Graphics.
- Musical instrument digital interface (MIDI).
- Text.

The obvious reason for creating and integrating these media types is the human desire to share and communicate one's experiences with others. This has been done for centuries in the form of books and, in relatively more recent decades, through movies and television. In the here and now, the amazingly powerful and versatile presence of the Web can be added to this communications list. Nothing allows individuals and corporate entities alike to reach millions so easily. Perhaps most importantly, the Web is a multimedia experience that each individual can manipulate, learn from and even respond to in an interactive fashion. The Web has indeed unlocked the potential for experiencing multimedia events and information in a way that makes each of us a participant not just a passive spectator. This is the true revolution occurring at the dawn of the 21st century!

33.9.1 MIDI in Multimedia

One of the unique advantages of MIDI as it applies to multimedia is the rich diversity of musical instruments and program styles that can be played back in real time, while requiring almost no overhead processing from the computer's CPU. This makes MIDI a perfect candidate for playing back soundtracks from multimedia games or over the phone, internet, gaming devices, etc. Still, as one might expect, MIDI has taken a back seat to digital audio as a serious music playback format for multimedia. Most likely, this is due to several factors, including:

- A basic misunderstanding of the medium.
- The fact that producing MIDI content requires a fundamental knowledge of music.
- The frequent difficulty of synchronizing digital audio to MIDI in a multimedia environment.
- The fact that soundcards, phones, etc. often include poorly designed FM synthesizers (although most operating systems now include higher quality software synths).

Fortunately, a number of companies have taken up the banner of embedding MIDI within their media projects and have helped push MIDI a bit more into the Web mainstream. As a result, it's becoming more common for your PC to begin playing back a MIDI score on its own or perhaps in conjunction with a game or more data-intensive program.

33.9.2 Standard MIDI Files

The accepted format for transmitting files or real-time MIDI information in multimedia (or between sequencers from different manufacturers) is the standard MIDI file. This file type (which is stored with a .mid or .smf extension) is used to distribute MIDI data, song, track, time signature and tempo information to the general masses. Standard MIDI files can support both single and multichannel sequence data and can be loaded into, edited and then directly saved from almost any sequencer package. When exporting a standard MIDI file, keep in mind that they come in two basic flavors—type 0 and type 1:

- Type 0 is used whenever all of the tracks in a sequence need to be merged into a single MIDI track. All of the notes will have a channel number attached to them (i.e., will play various instruments within a sequence); however, the data will have no definitive track assignments. This type might be the best choice when creating a MIDI sequence for the Internet (where the sequencer or MIDI player application might not know or care about dealing with multiple tracks).
- Type 1, on the other hand, will retain its original track information structure and can be imported into another sequencer type with its basic track information and assignments left intact.

33.9.3 General MIDI

One of the most interesting aspects of MIDI production is the absolute uniqueness of each professional and even semipro project studio. In fact, no two studios will be alike (unless they have been specifically designed to be the same or there's some unlikely coincidence). Each artist will be unique as to his or her own favorite equipment, supporting hardware, way of routing channels and tracks and assigning patches. The fact that each system setup is unique and personal has placed MIDI at odds with the need for system and setup compatibility in the world of multimedia. For example, after importing a MIDI file over the Net that's been created in another studio, the song will most likely attempt to play with a totally irrelevant set of sound patches (it might sound interesting, but it won't sound anything like what was originally intended). If the MIDI file is loaded into another setup, the sequence will again sound completely different, with patches that are so irrelevant that the guitar track might sound like a bunch of machine-gun shots from the planet Gloop.

To eliminate (or at least reduce) the basic differences that exist between systems, a standardized set of patch settings, known as General MIDI (GM), was created. In short, General MIDI assigns a specific instrument patch to each of the 128 available program change numbers. Since all electronic instruments that conform to the GM format must use these patch assignments, placing GM program change commands at the header of each track will automatically instruct the sequence to play with its originally intended sounds and general song settings. In this way, no matter what sequencer and system setup is used to play the file back, as long as the receiving instrument conforms to the GM spec, the sequence will be heard using its intended instrumentation.

Tables 33-2 and 33-3 detail the program numbers and patch names that conform to the GM format. These patches include sounds that imitate synthesizers, ethnic instruments or sound effects that have been derived from early Roland synth patch maps. Although the GM spec states that a synth must respond to all 16 MIDI channels, the first 9 channels are reserved for instruments, while GM restricts the percussion track to MIDI channel 10.

Table 33-2. GM Nonpercussion Instrument (Program Change) Patch Map

1. Acoustic Grand Piano 2. Bright Acoustic Piano 3. Electric Grand Piano 4. Honky-tonk Piano 5. Electric Piano 1

6. Electric Piano 2
7. Harpsichord
8. Clavi
9. Celesta
10. Glockenspiel
10. Music Box
12. Vibraphone
13. Marimba
14. Xylophone
15. Tubular Bells
16. Dulcimer
17. Drawbar Organ
18. Percussive Organ
19. Rock Organ
20. Church Organ
21. Reed Organ
22. Accordion
23. Harmonica
24. Tango Accordion
25. Acoustic Guitar (nylon)
26. Acoustic Guitar (steel)
27. Electric Guitar (jazz)
28. Electric Guitar (clean)
29. Electric Guitar (muted)
30. Overdriven Guitar
31. Distortion Guitar
32. Guitar harmonics
33. Acoustic Bass
34. Electric Bass (finger)
35. Electric Bass (pick)
36. Fretless Bass
37. Slap Bass 1
38. Slap Bass 2
39. Synth Bass 1
40. Synth Bass 2
41. Violin
42. Viola

43. Cello
44. Contrabass
45. Tremolo Strings
46. Pizzicato Strings
47. Orchestral Harp
48. Timpani
49. String Ensemble 1
50. String Ensemble 2
51. SynthStrings 1
52. SynthStrings 2
53. Choir Aahs
54. Voice Oohs
55. Synth Voice
56. Orchestra Hit
57. Trumpet
58. Trombone
59. Tuba
60. Muted Trumpet
61. French Horn
62. Brass Section
63. SynthBrass 1
64. SynthBrass 2
65. Soprano Sax
66. Alto Sax
67. Tenor Sax
68. Baritone Sax
69. Oboe
70. English Horn
71. Bassoon
72. Clarinet
73. Piccolo
74. Flute
75. Recorder
76. Pan Flute
77. Blown Bottle
78. Shakuhachi
79. Whistle

80. Ocarina 81. Lead 1 (square) 82. Lead 2 (sawtooth) 83. Lead 3 (calliope) 84. Lead 4 (chiff) 85. Lead 5 (charang) 86. Lead 6 (voice) 87. Lead 7 (fifths) 88. Lead 8 (bass b lead) 89. Pad 1 (new age) 90. Pad 2 (warm) 91. Pad 3 (polysynth) 92. Pad 4 (choir) 93. Pad 5 (bowed) 94. Pad 6 (metallic) 95. Pad 7 (halo) 96. Pad 8 (sweep) 97. FX 1 (rain) 98. FX 2 (soundtrack) 99. FX 3 (crystal) 100. FX 4 (atmosphere) 101. FX 5 (brightness) 102. FX 6 (goblins) 103. FX 7 (echoes) 104. FX 8 (sci-fi) 105. Sitar 106. Banjo 107. Shamisen

Table 33-3. GM Percussion Instrument (Key Number/Musical Note) Patch Map (Channel 10)

35/BO. Acoustic Bass Drum 36/C1. Bass Drum 1 37/C#1. Side Stick 38/D1. Acoustic Snare 39/Eb1. Hand Clap

40/E1. Electric Snare
41/F1. Low Floor Tom
42/F#1. Closed Hi-Hat
43/G1. High Floor Tom
44/Ab1. Pedal Hi-Hat
45/A1. Low Tom
46/Bb1. Open Hi-Hat
47/B1. Low-Mid Tom
48/C2. Hi Mid Tom
49/C#1. Crash Cymbal 1
50/D2. High Tom
51/Eb2. Ride Cymbal 1
52/E2. Chinese Cymbal
53/F2. Ride Bell
54/F#2. Tambourine
55/G2. Splash Cymbal
56/Ab2. Cowbell
57/A2. Crash Cymbal 2
58/Bb. Vibraslap
59/B2. Ride Cymbal 2
60/C3. Hi Bongo
61/C#3. Low Bongo
62/D3. Mute Hi Conga
63/Eb3. Open Hi Conga
64/E3. Low Conga
65/F3. High Timbale
66/F#3. Low Timbale
67/G3. High Agogo
68/Ab3. Low Agogo
69/A3. Cabasa
70/Bb3. Maracas
71/B3. Short Whistle
72/C4. Long Whistle
73/C#4. Short Guiro
74/D4. Long Guiro
75/Eb4. Claves
76/E4. Hi Wood Block

77/F4. Low Wood Block
78/F#4. Mute Cuica
79/G4. Open Cuica
80/Ab4. Mute Triangle
81/A4. Open Triangle

Note: In contrast to Table 33-2, the numbers in Table 33-3 represent the percussion keynote numbers on a MIDI keyboard, not program change numbers.

33.10 MIDI Time Code

In earlier times, the synchronizing of audio devices to other video and/or audio devices was a very ex-pensive proposition, far beyond the budget of most project or independent production houses. Today, however, an easy-to-use and inexpensive standard makes use of MIDI to transmit sync and timecode data throughout a connected production system, Fig. 33-27. This has made it possible for even the most budget-minded bedrooms to be able to synchronize media devices and software using time-code.

Figure 33-27. Many time-based media devices in the studio can be cost effectively connected via MIDI timecode (MTC).

MIDI timecode (MTC) was developed to allow electronic musicians, project studios, video facilities and virtually all other production environments to cost effectively and easily translate timecode into time-stamped messages that can be transmitted over MIDI data lines. Created by Chris Meyer and Evan Brooks, MIDI timecode allows SMPTE-based timecode to be distributed throughout the MIDI chain to devices or instruments that are capable of synchronizing to and executing MTC commands. MIDI timecode is an extension of MIDI 1.0, which makes use of existing Sys-ex message types that were either previously undefined or were being used for other, non-conflicting purposes.

Since most modern recording systems include MIDI in their design, there's often no need for external hardware when making direct connections. Simply chain the MIDI data lines from the master to the appropriate slaves within the system (via physical cables, USB or virtual internal routing). Although MTC uses a reasonably small percentage of MIDI's available bandwidth (about 7.68% at 30 fr/sec), it's customary (but not necessary) to separate these lines from those that are communicating performance data when using physical MIDI cables. As with conventional SMPTE, only one master can exist within an MTC system, while any number of slaves can be assigned to follow, locate and chase to the master's speed and position. Because MTC is easy to use and is often included free in many system and program designs, this technology has grown to become the most straightforward and commonly used way to lock together such devices as DAWs, modular digital multitracks and MIDI sequencers, as well as analog and videotape machines (by using a MIDI interface that includes a SMPTE-to-MTC converter).

33.10.1 MIDI Timecode Messages

The MIDI timecode format can be divided into two parts:

- Timecode.

- MIDI cueing.

The timecode capabilities of MTC are relatively straightforward and allow devices to be synchronously locked or triggered to SMPTE timecode. MIDI cueing is a format that informs a MIDI device of an upcoming event that is to be performed at a specific time (such as load, play, stop, punch in/out, reset). This protocol envisions the use of intelligent MIDI devices that can prepare for a specific event in advance and then execute the command on cue.

MIDI timecode is made up of three message types:

- **Quarter-frame messages**. These are transmitted only while the system is running in real or variable speed time, in either forward or reverse direction. True to its name, four quarter-frame messages are generated for each timecode frame. Since 8 quarter-frame messages are required to encode a full SMPTE address (in hours, minutes, seconds and frames: 00:00:00:00), the complete SMPTE address time is updated once every two frames (In other words, MIDI timecode actually has half the resolution accuracy of its SMPTE timecode counterpart). Each quarter-frame message contains 2 bytes. The first byte is F1, the quarter-frame common header; the second byte contains a nibble (four hits) that represents the message number (0 through 7) and a nibble for encoding the time field digit.
- **Full messages**. Quarter-frame messages are not sent in the fast-forward, rewind or locate modes, because this would unnecessarily clog a MIDI data line. When the system is in any of these shuttle modes, a full message is used to encode a complete timecode address. After a fast shuttle mode is entered, the system generates a full address message and then places itself in a pause

mode until the time-encoded slaves have located to the correct position. Once playback has resumed, MTC will again begin sending incremental quarter-frame messages.

- **MIDI cueing messages**. MIDI cueing messages are designed to address individual devices or programs within a system. These 13-bit messages can be used to compile a cue or edit decision list, which in turn instructs one or more devices to play, punch in, load, stop, and so on, at a specific time. Each instruction within a cueing message contains a unique number, time, name, type and space for additional information. At the present time, only a small percentage of the possible 128 cueing event types have been defined.

33.10.2 SMPTE-to-MIDI Conversion

Although MIDI time code connections can be directly made between compatible MIDI devices, a SMPTE-to-MIDI converter is required to read incoming MTC SMPTE timecode and convert it into MIDI timecode (and vice versa) for other device types. These conversion systems are available as a stand-alone device or as an integrated part of an audio interface or multiport MIDI interface/patch bay/synchronizer system, Fig. 33-28.

Figure 33-28. SMPTE timecode can often be generated throughout a production system, possibly as either MTC or as MTC

via a capable MIDI or audio interface.

Chapter 34

Optical Disc Formats for Audio Reproduction and Recording

by Ken Pohlmann

34.1 Introduction

The digital storage of audio signals presents a technical challenge. For example, a 60-minute stereo music selection, recorded with a sampling frequency of 44.1kHz and 16bit pulse code modulation, generates over 5 billion bits. Recordings with a higher sampling frequency, longer word length, and additional channels require correspondingly greater storage capacity. Optical discs provide high-capacity storage with low-cost replication as well as durability, convenience, random access, and relatively small size. Optical discs play a major role in the dissemination of audio (and video) data.

The Compact Disc (CD) was the first optical disc format widely used to store digital audio data. A CD can hold over an hour of high-fidelity music on a robust and economically manufactured disc. A number of alternative CD formats were developed. A CD-ROM disc may hold several hours of music, along with video and text information. Write-once and recordable/erasable formats (CD-R and CD-RW) were also developed.

The desire for higher performance specifications and multichannel sound stimulated development of the Super Audio CD (SACD) format; it uses direct stream digital coding in place of PCM coding to store either stereo or multichannel audio signals on a multilayer disc. The need for increased storage capacity, particularly for the storage of high-quality digital video, encouraged development of the DVD format. A DVD disc may store from 4.7 to 17Gbytes of data, using one or multiple data layers. The DVD-Video format is used to store motion pictures, DVD-Audio is used for high-quality stereo and multichannel music, DVD-ROM is used for computer applications, and a variety of DVD formats have been devised for recording applications. The Blu-ray disc format uses shorter wavelength lasers and higher resolution optics to dramatically increase storage density, allowing storage of high-definition video and audio. Disc formats such as these extend the opportunities of optical disc storage for professional and consumer applications.

34.2 Compact Disc Format

The Compact Disc Digital Audio (CD-DA) format is sometimes known as the Red Book standard and is codified in the ISO/IEC 908 standard. The diameter of a CD is 120 millimeters (mm) (4.7in), its center hole diameter is 15mm (0.59in), and its thickness is 1.2mm (0.047in). The innermost diameter does not hold data; it provides a clamping area for the player to secure the disc to the spindle motor shaft. Data is recorded on a 35.5mm (1.4in) wide area. A lead-in area occupies the innermost data radius, and a lead-out area occupies the outermost radius; they contain nonaudio data used to control the player's operation.

A transparent polycarbonate plastic substrate forms most of a disc's 1.2mm thickness, as shown in Fig. 34-1. Data is physically contained in pits that are impressed on the top surface of the substrate. The pit surface is covered with a very thin 50nm to

100nm (nanometer) metal (e.g., aluminum or gold) layer and another thin 10μm to 30μm (micrometer) protective plastic layer, with the 5μm identifying label printed on top. A laser beam is used to read the data. It is applied from below, passes through the transparent substrate, reflects from the metallized pit surface, and passes back through the substrate. The laser beam is focused on the metallized surface embedded inside the disc.

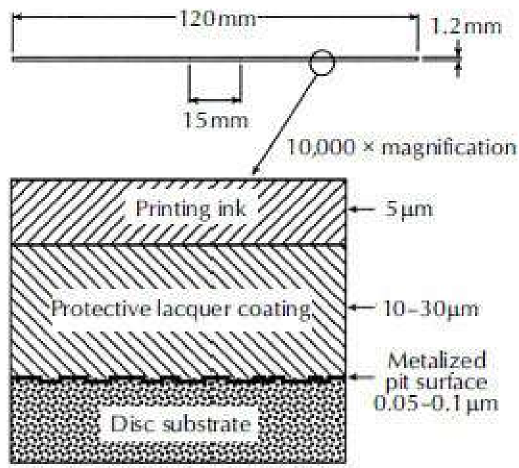


Figure 34-1. CD construction showing substrate, metallized surface, protective layer, and label.

34.2.1 Pit Track

Data is arranged as a pit track in a continuous spiral running from the inner circumference to the outer. A pit is about 0.6μm wide. A photograph of a pit surface, taken with a scanning electron microscope, is shown in Fig. 34-2. The track pitch, the distance between successive tracks, is 1.6μm. There is a maximum of 20,188 revolutions across the disc's standard data surface width of 35.5mm.

The linear dimensions of a track are the same at the beginning of a spiral as at the end. This means that a CD rotates with a constant linear velocity (CLV), a condition in which a uniform relative velocity is maintained between the data spiral and the pickup. To accomplish this, the rotation speed of a disc varies depending on the radial position of the pickup. Because each outer track revolution contains more pits than each inner track revolution, the disc must be slowed as it plays outward to maintain a constant rate of data. In particular, the disc rotates at a speed of about 500rpm when the pickup is reading the inner circumference, and as the pickup moves outward, the rotational speed gradually decreases to about 200rpm. A constant linear velocity is maintained through a CLV servo system; the player reads frame synchronization from the stored data and varies the disc speed to maintain a constant data rate. The CD standard permits a maximum of 74 minutes, 33 seconds of audio playing time on a disc. However, by reducing parameters such as track pitch and linear velocity, it is possible to manufacture discs with over 80 minutes of music.

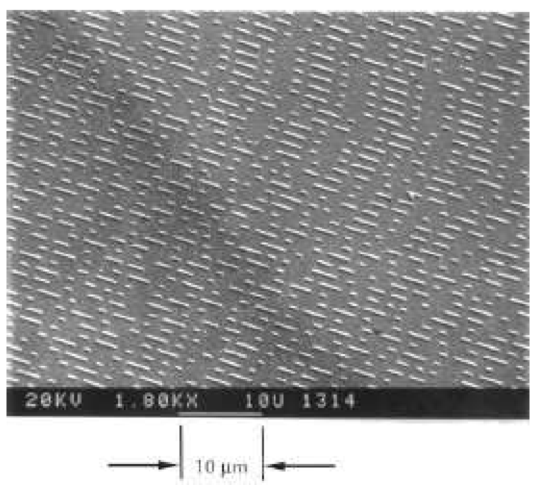


Figure 34-2. Scanning electron microscope photograph of the CD data surface. Courtesy University of Miami.

The fact that the disc data surface is physically separated from the reading side of the substrate provides a significant asset. Damage and dust on the outer surface do not lie in the focal plane of the reading laser beam and hence their effect is minimized.

The polycarbonate substrate has a refractive index of 1.55; the velocity of light slows from 3×0^5km/s, to 1.9×10^5km/s. Because of the bending from the refractive index and thickness of the substrate, and the numerical aperture (NA) of 0.45 of the laser pickup's lens, the diameter of the laser spot is reduced from approximately 800µm on the disc surface to approximately 1µm at the pit surface. The laser beam is thus focused to a point larger than a pit width.

The reflective data pit surface, known as land, causes almost 90% of the laser light to be reflected back into the optical pickup. When viewed from the laser's underside perspective, the pits appear as bumps. The height of each bump is between 0.11 and 0.13µm (110 and 130nm.) This dimension is slightly smaller than the laser beam's wavelength in air of 780nm. Inside the polycarbonate substrate, the laser's wavelength is about 500nm. The height of the bumps is thus approximately one-quarter of the laser's wavelength in the substrate.

There is a phase difference between the part of the beam reflected from the bump, and the part reflected from the surrounding land. The phase difference causes destructive interference in the reflected beam. In theory, when the beam strikes an area between pits virtually all of its light is reflected, and when it strikes a pit virtually all of the light returning to the pickup is canceled, hence virtually none is reflected. In practice, the laser spot is larger than required for complete cancellation between pit and land reflections, and pits are made slightly shallower than a quarter wavelength; this yields a better tracking signal, among other things. Typically the presence of a bump reduces reflective power by about 25%. In any case, the data surface varies the intensity of the reflected laser beam. Thus the data physically encoded on the disc can be recovered by the laser, and converted to an electrical signal using a photodiode.

34.2.2 Data Encoding

The audio program on a CD is encoded from a master recording. Various media are used to hold master recordings. Originally, many CDs were mastered from data recorded on ¾-inch U-matic videotape cassettes using a digital audio processor. In many cases, Exabyte 8-mm data tapes are used to hold the master recording. For audio mastering, the DDP (Disk Description Protocol) file format may be used to hold Red Book and PQ subcode data. Both DDP 1.0 and DDP 2.0 are used; the 2.0 specification writes the TOC to the end of the tape. It is generally recommended to supply a replication plant with an Exabyte tape with DDP files (including PQ and ISRC data). With Exabyte tapes, glass masters may be created at faster than real time speeds. In some cases, audio data is written to a master CD-ROM (CD read-only memory) disc as 24-bit WAV or AIFF files. CD-R discs can be used as masters, but their relatively higher error rates and susceptibility to damage make them nonideal. An analog tape can also be used as the master. Digital recordings made at a different sampling rate must be passed through a sample rate converter.

CD encoding is the process of placing audio data in a format suitable for storage on the disc. A frame structure provides a means to distinguish the data types. The information contained in a CD frame (prior to modulation) contains a 27-bit sync word, 8-bit subcode, 192 data bits, and 64 parity bits. Encoding begins with the audio data. Six 32-bit PCM audio sampling periods (alternating from 16-bit left and right channels) are grouped in a frame, left channel preceding right. Each 32-bit sampling period is divided to yield four 8-bit audio symbols. Subsequent signal processing prepares the audio data for storage on the disc surface. In particular, error correction encoding must be accomplished.

The raw error rate from a CD is around 10^{-5} to 10^{-6}, or about one error for every 0.1 to 1.0 million channel (stored) bits. Considering that a disc outputs 4.3218 million channel bits per second, the need for error correction is obvious. With error correction, 220 errors per second can be completely corrected; interleaving distributes errors, and parity corrects them.

The Cross Interleave Reed-Solomon Code (CIRC) algorithm is used for error correction in the CD system. The CIRC algorithm uses two correction codes for correcting capability, and three interleaving stages to encode data before it is placed on a disc and to decode the data during playback. Because of cross interleaving, the separation of two error correction codes by an interleaving stage, one Reed-Solomon code can check the validity of the other code. The Reed-Solomon code used in CIRC is well suited for the CD system because its decoding requirements are relatively simple. The complete CIRC encoding scheme is shown in Fig. 34-3. With this encoding algorithm, data (twenty-four 8-bit symbols) from the audio signal are cross-interleaved, and two encoding stages generate 8-bits of parity.

34.2.3 Subcode

Following CIRC encoding, an 8-bit CD subcode symbol is added to each frame. The eight subcode bits are designated as P, Q, R, S, T, U, V, and W. Only the P or Q bits are required in the audio format. The CD player collects subcode symbols from 98 consecutive frames to form a subcode block, with eight 98-bit words. Thus the eight subcode bits (P through W) are used as eight different channels with each CD frame containing 1 P bit, 1 Q bit, etc. A subcode block is complete with a synchronization word, instruction and data, commands, and parity. The start of each subcode block is denoted by sync patterns in the first symbol positions of two successive blocks.

The P channel contains a flag bit originally designed for use by simple players to access disc information. In practice, players ignore the P bit and use information in the more comprehensive Q channel. The Q subcode channel is vital for reading audio data on the disc. The Q channel contains four kinds of information: control, address, Q data, and cyclic redundancy check code (CRCC). Each subcode block contains 72-bits of Q data and 16-bits for CRCC, used for error detection on the control, address, and Q data information. The control information flag bits handle several player functions:

1. The number of audio channels (two or four) is indicated; this distinguishes between a two- and four channel CD recording (the latter not implemented).
2. Preemphasis (on/off) is indicated; a CD track may be encoded with preemphasis, a noise suppression method (this is rarely employed).
3. Digital copy prohibited (yes/no) is indicated.
4. Audio or data content is indicated.

The address information consists of four bits designating the three modes for the Q data bits. Primarily, Mode 1 contains the number and start times of tracks, Mode 2 contains a catalog number, and Mode 3 contains other product codes. Mode 1 stores information in the disc lead-in area, program area, and lead-out area; the data format in the lead-in area differs from that in the other areas. Mode 1 lead-in information is contained in the CD table of contents (TOC). The TOC stores data indicating the number of music selections (up to 99) as a track number and the starting points of the tracks in disc running time. The TOC is read during disc initialization, before the disc begins playing audio data.

In the program and lead-out areas, Mode 1 contains track numbers, indices (subdivision numbers) within a track, time within a track, and absolute time. A time count is set to zero at the beginning of each track and increases to the end of the track. At the beginning of a pause, a time count decreases ending with zero at the end of the pause. The absolute time is set to zero at the beginning of the program area and increases to the start of the lead-out area. Time and absolute time are expressed in minutes, seconds, and frames (75 frames per second). Modes 2 and 3 are optional in the subcode.

The other six channels (R, S, T, U, V, and W), which account for about 20 megabytes of 8-bit storage, are available for other data storage. In some discs, this capacity is used to hold CD-Text, a feature that was appended to the original Red Book specification. With CD-Text, album title, song titles, artist names, and other text information are coded prior to manufacture. Compatible players can read and display CD Text information.

34.2.4 EFM Encoding

After the audio, parity, and subcode data is assembled, the bit stream is modulated using EFM (eight-to-fourteen modulation). Blocks of eight data bits are translated into blocks of 14 channel bits, assigning an arbitrary and unambiguous word of 14 bits to each 8-bit word. By choosing select 14-bit words with a low number (and known rate) of 1/0 transitions, greater data density can be achieved. It would be inefficient to store the 8-bit symbols directly on the disc; the large number of 1/0 transitions would demand many pits. In addition, 8-bit symbols have many similar patterns. With 14-bit words, more unique patterns can be selected. EFM thus expedites error correction.

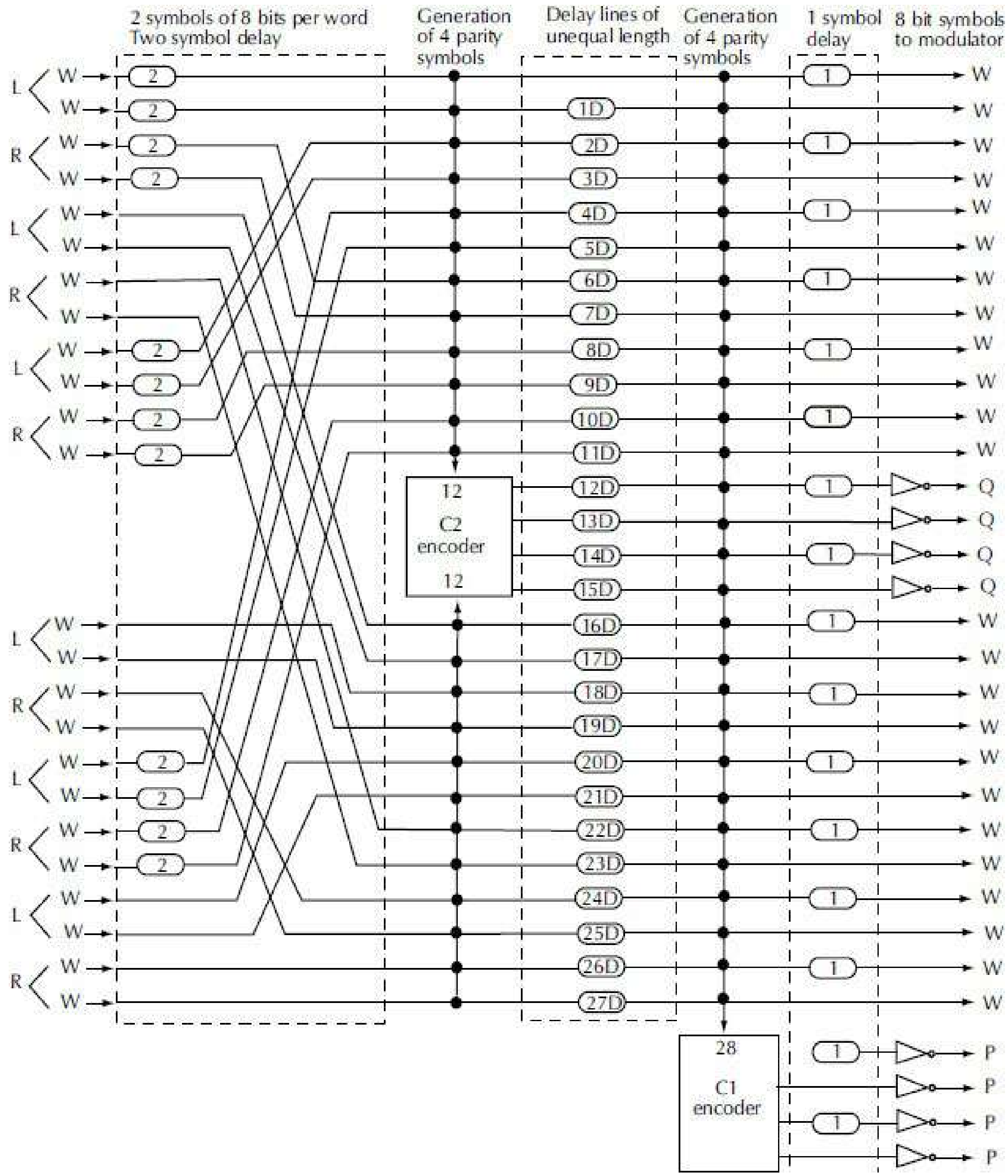


Figure 34-3. CIRC encoding algorithm.

Blocks of 14 bits are linked by three merging bits; two merging bits (always 0s) are required to prevent the possibility of successive 1s between serial words (a violation of the EFM coding scheme). The additional merging bit (either a 1 or a 0, depending on the preceding and succeeding patterns) is added to each code pattern to aid in clock synchronization and to suppress the signal's low-frequency component. The latter is accomplished by selecting merging bits that maintain the signal's average digital sum value at zero. The ratio of bits before and after modulation is 8:17. During demodulation, only 14 bits will be processed, the three merging bits are discarded.

The eight data bits require 28 or 256 different code patterns. However the 14-bit channel word can offer 16,384 combinations. To achieve pits of controlled length, only those combinations are selected in which more than two but less than ten 0s appear continuously. In addition, unique patterns are sought. Only 267 combinations satisfy these criteria. Because only 256 patterns are needed, 11 of the 267 patterns are discarded (two of them are used for subcode synchronization words).

The resulting channel stream produces pits and lands that comprise at least two (3T) but no more than ten (11T) successive 0s in length. It is the combination of these varying dimensions that physically encodes the data. The selection of EFM bit patterns defines the physical relationship of the pit dimensions. The pits and intervening reflective land on the CD surface do not directly designate 1s and 0s. Rather, each pit edge whether leading or trailing, is a 1 and all increments in between, whether inside or outside a pit, are 0s, as shown in Fig. 34-4.

With EFM there are more bits to accommodate, but with modulation the highest frequency in the output signal is decreased. Therefore a lower track velocity can be utilized and longer playing time is achieved. This is an efficient encoding method because the number of bits transmitted divided by the number of transitions needed on the medium to convey them is high.

34.3 CD Player Design

A CD player hardware architecture may be considered as five functional elements: Optical readout, servo system, spindle motor, control and display, and decoding circuits. The data path directs the modulated light from the pickup through a series of processing circuits, ultimately yielding a stereo analog signal. The data path typically consists of elements such as data separator, deinterleaving RAM, error detection, correction and concealment circuits, oversampling filters, D/A converters, and analog output filters. The servo, control, and display system must direct mechanical operation of the disc, including spindle drive, auto-tracking, and auto-focusing, and handle user interface with the player's controls and displays. A block diagram of the data path is shown in Fig. 34-5.

34.3.1 Optical Pickup

The CD optical pickup must focus, track, and read the data spiral. The entire lens assembly, a combination of the laser source and the reader, must be small enough to move laterally beneath the disc in response to tracking information and user access demands. Furthermore, the pickup must maintain focusing and tracking even under adverse playing conditions such as a dirty disc or impact and vibration.

To achieve focus on the data surface and intensity modulation, a laser is used as the light source. CD pickups use an AlGaAs semiconductor laser irradiating a coherent-phase laser beam with a 780nm wavelength. CD players can employ either single-beam or three-beam pickups; three-beam designs are more prevalent. A three-beam pickup uses a center beam for reading data and focusing, and two secondary beams for tracking. The design of a three-beam pickup is shown in Fig. 34-6. To generate additional beams, the laser light passes through a diffraction grating. As the beam passes through the grating, the light diffracts; when the resulting collection is again focused, it will appear as a single bright centered beam with a series of successively less intense beams on either side. Three beams from this diffraction pattern usefully strike the disc. As noted, when a laser spot strikes land, the smooth interval between two pits, the light is almost totally reflected; when it strikes a pit (seen as a bump by the laser), destructive interference and diffraction causes less light to be reflected into the pickup. The intensity-modulated light is collected by the objective lens and passes through the reading portion of the pickup.

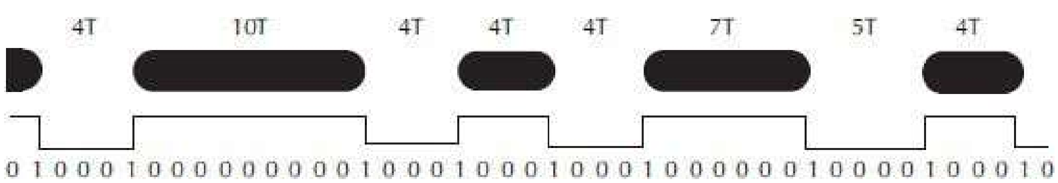


Figure 34-4. Pit/land edges represent logical 1 data.

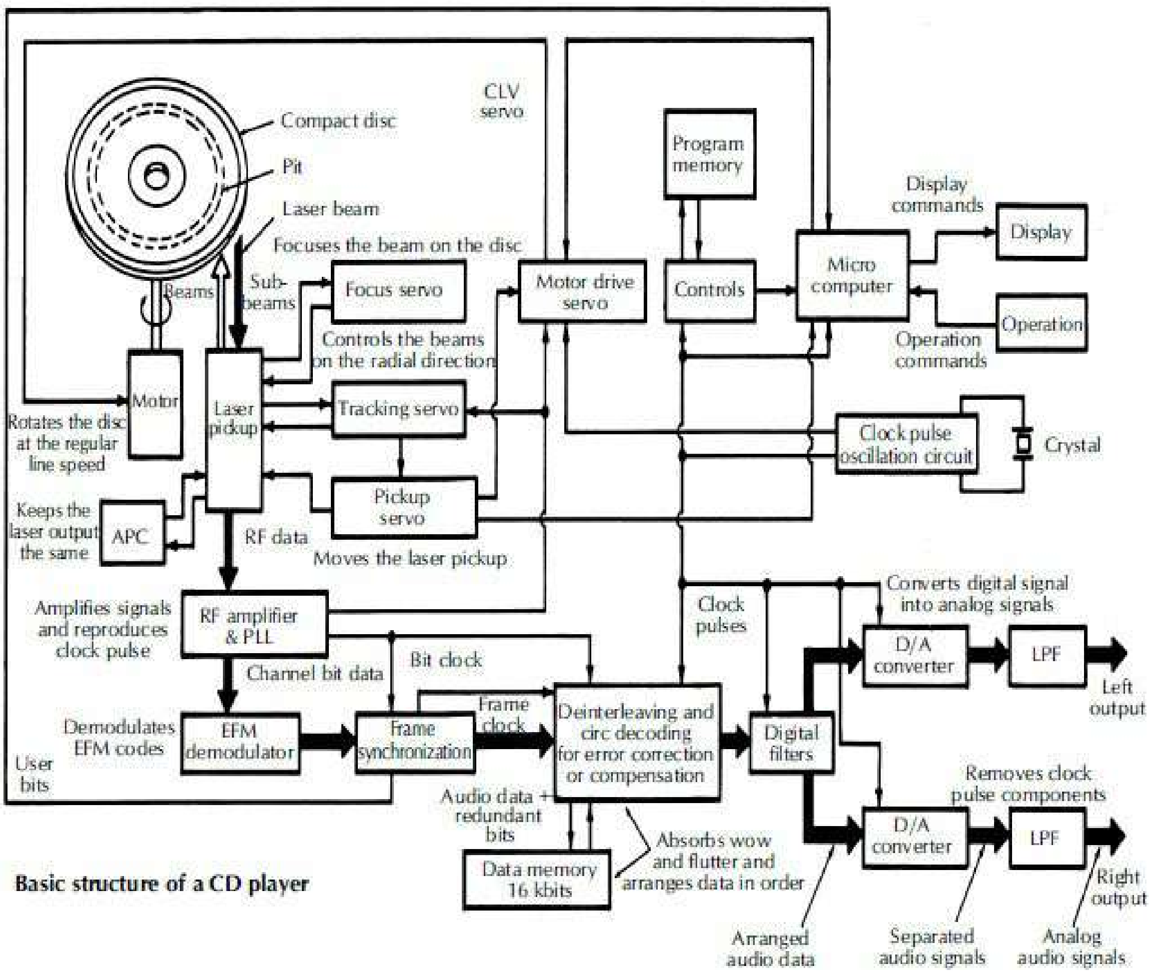


Figure 34-5. CD player block diagram showing optical processing and output signal processing.

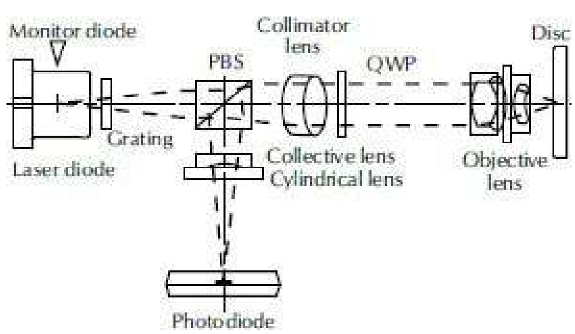


Figure 34-6. Three-beam optical pickup showing diffraction grating, objective lens, and photodiode.

In many three-beam designs, the property of astigmatism is used to achieve auto-focusing. A cylindrical lens is used to detect an out-of-focus condition. As the distance between the objective lens and disc reflective surface varies, the focal point of the optical system also changes, and the image projected by the cylindrical lens changes its shape, as shown in Fig. 34-7. That change in the image on a four-quadrant photodiode generates the focus correction signal. For example, if the disc were too near to the pickup's objective lens, the focal length would be shortened and astigmatism from the cylindrical lens would cause the reflected laser spot to be flattened and rotated to one side. This would cause more light to fall on two (opposite) pairs of photodiodes than on the other pairs. This generates a voltage interpreted by the servo system as a command to pull the lens down from the disc. This provides the correct focal path length where astigmatism would not affect the beam. Hence, it would have a round shape, and an equal amount of light would fall on each part of the four-quadrant photodiode, providing a neutral signal to the servo system. When the disc is too far from the lens, the laser spot rotates in the opposite direction, generating a voltage that pushes the lens upward. In practice, the process in this servo loop is a dynamic one, with the objective lens moving in constant accord with disc deviations to provide a correct focal path length.

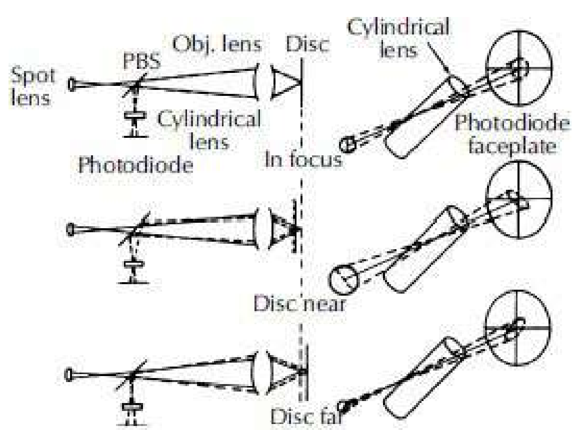


Figure 34-7. Astigmatism used in auto-focusing.

In three-beam pickups, the two secondary beams are used for auto-tracking. The central beam spot covers the pit track while the two tracking beams are aligned above and below and to either side. When the beam is tracking the disc properly, part of each tracking beam is aligned on the pit edge; the other part covers the land between pit tracks. The main beam strikes a four-quadrant photodiode, and the two tracking beams strike two separate photodiodes mounted on either side of the main photodiode.

If the three spots drift to either side of the pit track, the amount of light reflected from the tracking beams varies. There is less average light intensity reflected by the tracking beam that encounters more pit area and greater reflected light intensity from the tracking beam that encounters less pit area. The relative output voltages from the two tracking photodiodes thus form a tracking correction signal, as shown in Fig. 34-8. Operating similarly to the signal used in the auto-focus servo loop, this tracking signal forms a control voltage for the auto-tracking servo mechanism. For example, when the pickup's objective lens drifts to the right of the pit track, the right tracking beam encounters more reflective land and its reflected intensity is greater. When this brighter spot strikes the right tracking photodiode, a voltage greater than that on the left photodiode is generated. This voltage shift causes the servo system to move the pickup to the left, toward the pit track center. Likewise, the opposite occurs when the pickup drifts to the left. In this dynamic process the servo system continually moves the pickup to compensate for track deviations.

In addition to auto-focus and auto-tracking, a CD pickup uses other motor systems to move the pickup across the disc surface in response to user commands. For example, the pickup must search rapidly across the disc as it reads data, or move from one track to another. These functions are handled using control signals derived from the auto-tracking and auto-focus circuits; however, separate motors are used to move the pickup itself. Three-beam pickups are mounted on a sled that moves across the disc surface. In many designs, linear motors move the pickup and position it to within capture range of the auto-tracking circuit, which takes control when the selected disc location is found. A spindle motor is used to rotate the disc with constant linear velocity. Thus the player must vary the disc speed depending on where the pickup is located on the disc surface. This is accomplished with yet another servo loop; information from the data stream recovered by the laser pickup is used to determine correct rotating speed, and the spindle motor is regulated accordingly.

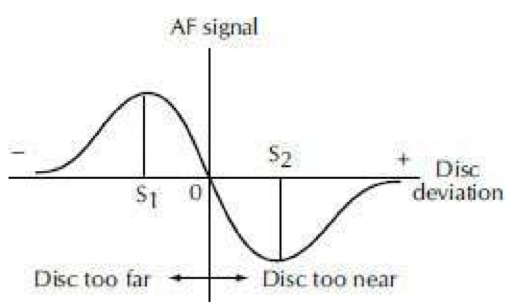


Figure 34-8. Auto-tracking correction signal.

34.3.2 Data Decoding

The photodiode array and its processing circuits produce a signal resembling a series of high-frequency sinusoids called the EFM signal. A collection of EFM waveforms (forming an eye pattern) is shown in Fig. 34-9. The digital data can be recovered from the EFM signal if it can be determined when the signal crosses the zero axis, relative to the timing constraints created by the EFM encoding rules.

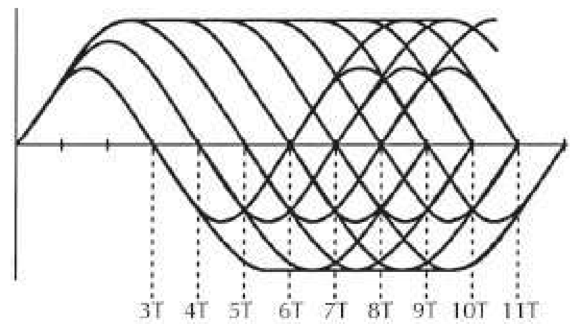


Figure 34-9. EFM eye pattern.

CD data decoding follows a procedure that essentially duplicates, in reverse order, the encoding process. The first data to be extracted from the signal is synchronization words. This information is used to synchronize the thirty-three 17-bit symbols of channel information in each frame, and a synchronization pulse is generated to aid in locating the zero crossing of the EFM pattern and to generate a transition at those points to produce a binary signal.

The EFM signal is demodulated so that every 17-bit EFM word is reconverted to an 8-bit word. Demodulation can be accomplished by logic circuitry or a look-up table. A buffer is used to remove the effect of disc rotational irregularities; data input to the buffer may be irregular in time but clocking ensures that the buffer output is precise. To guarantee that the buffer neither overflows nor underflows, a correction signal is generated and used to regulate the disc rotating speed.

Following demodulation, data is sent to a CIRC decoder for deinterleaving, error detection, and correction. The CIRC decoding process reverses the processing steps accomplished during encoding. The CIRC decoder accepts one frame of thirty-two 8-bit symbols; twenty four are audio symbols, and eight are parity symbols. One frame of twenty-four 8-bit symbols is output. The decoder utilizes parity from two Reed-Solomon decoders and deinterleaving. The first error correction decoder is designed to correct random errors, and to detect burst errors. It flags all burst errors, to alert the second error correction decoder.

Error concealment algorithms, employing interpolation and muting circuits, follow the CIRC decoder. Uncorrected words are detected through flags and dealt with, while valid data passes through unprocessed. Using error flags, the player's signal-processing circuits determine whether to output the data directly, to interpolate it, or to mute the sound.

For continuous errors, muting is employed as a last resort; invalid data passed on to the D/A converter could result in an audible click. Muting is accomplished by beginning attenuation many samples before the invalid data, smoothly muting the invalid data, and then smoothly restoring the signal level. This method of muting is often largely inaudible.

34.3.3 Signal Reconstruction

At the output stage, the digital data is converted to a stereo analog audio signal. This reconstruction requires low-pass filtering to suppress high-frequency image components, and D/A conversion. An oversampling digital filter uses samples from the disc as input and then computes interpolation samples, digitally implementing the response of a low-pass filter. A transversal filter can be used to oversample; image components appear at multiples of the new sampling rate. Because the separation between the baseband and sidebands is greater, a low-order analog filter can be used to remove the images. The type of oversampling filter found in CD players is an example of a wider class of FIR (finite impulse response) digital filters used in many applications. The transversal filter used in CD players resamples and filters through interpolation. Resampling acts to increase the sampling rate; for example, in an eight-times oversampling filter, seven zero values are inserted for every data value output from the disc. This increases the sampling rate from 44.1kHz to 352.8kHz.

Interpolation is used to generate the values of intermediate sample points—for example, seven intermediate samples for each original sample. These samples are computed using coefficients derived from a low-pass filter response. In this way, when these samples are summed with other such samples, the output data stream corresponds to the sin (x)/(x) impulse response processing of an ideal low-pass filter. Following this processing, the data is converted into a format appropriate for the type of D/A converter used in the player. In most CD players, sigma-delta D/A converters are used, employing techniques such as short word lengths, very high oversampling rates and noise shaping.

Also present in the audio output stage of every player is an audio deemphasis circuit. Some CDs are encoded with audio preemphasis characteristic. On playback, this is detected and deemphasis is automatically carried out, resulting in an improvement in signal-to-noise ratio.

34.4 Other CD Formats

The CD's economy and robustness make it an excellent music carrier. However, its utility is not limited to music playback. Other formats, including computer-based storage and recordable formats, have been derived from the original Red Book standard. In particular, the CD-ROM, CD-R, and CD-RW formats are widely used in computer applications as well as audio applications.

34.4.1 CD-ROM

The CD read-only memory (CD-ROM) standard, sometimes called the Yellow Book standard, is codified as the ISO/IEC 10149 standard. It is derived from the CD audio standard but defines a format for general data storage and is not tied to any specific application. Ninety-eight CD frames are summed to form a data block of 2352 bytes (24 bytes × 98 frames). Each disc holds 330,000 blocks. The first 12 bytes of a block form a synchronization pattern, and the next four bytes form a header field for time and address flags. The header contains three address bytes, represented as disc times, storing minutes, seconds, and block numbers within the second. The header also contains a mode byte; depending on the mode selected, the remaining 2336 bytes can store user data, or 2048 bytes of user data with extended error correction.

The mode byte identifies three modes and is used for two different data types, shown in Fig. 34-10. Mode 1 permits 2048 bytes of user data in each block. Each block contains 2 Kbytes (2 × 1024) of user data; 280 bytes are given to extended error detection and correction (EDC/ECC). A Mode 1 CD-ROM holds 682 million bytes of user information (333,000 blocks × 2048 bytes). Mode 2 gives the full 2336 bytes to user data. A CD-ROM bit stream is applied to conventional CD encoding so that CIRC, EFM, and other processing is applied. Mode 1 thus has two independent layers of error correction (EDC/ECC and CIRC) whereas Mode 2 uses only CIRC error correction.

Because of its extended error correction, EDC/ECC data independently supplements the CIRC error correction code applied to the frame structure, improving the error rate over that of audio CD. Mode 1 is employed for numerical data storage, which is more critical than audio data. In EDC/ECC encoding, a GF(28) Reed-Solomon product code (RS-PC) codes each block. It produces P and Q parity bytes with (26,24) and (45,43) code words respectively.

The CD-ROM/XA format is an extension to the Mode 2 standard and defines an XA data track that can contain diverse computer data, and compressed audio and video. However, CD-ROM/XA differs from CD-ROM Mode 2; XA provides a subheader that defines two types of blocks: Form 1 for computer data and Form 2 for compressed audio/video data. The former provides a 2048-byte user area, and the latter provides 2324 bytes.

Hybrid audio/data CD formats such as CD Extra and Mixed Mode CD combine different format types (such as CD audio and CD-ROM/XA) on one disc. A CD Extra disc contains CD audio data in the first session, and CD-ROM-XA mode 2 data in the second session. In Mixed Mode CDs, ROM data is placed in track 1, and CD audio data is placed in subsequent tracks. To make sure an audio player does not access the ROM track, a pregap may be used so that ROM data is placed after the disc table of contents (TOC), but before the first music track. CD-ROM data is placed between Index 0 and Index 1 of Track 1, while the music starts at Track 1, Index 1. An audio player thus skips the data, starting playback at the first music track. However, the pregap area is not accessible to all drive software.

Unlike the CD audio standard, the CD-ROM standard does not stipulate how content is defined. Subsequently the ISO/DIS 9660 standard was devised; it specifies how computer data is placed on a CD-ROM; to read the data, the computer operating system must read the ISO 9660 file structure. Content on CD-ROM discs can be authored for multiple platforms; however, executable files can only run on the appropriate platform.

34.4.2 3CD-R

The CD recordable (CD-R) format allows users to permanently record audio or other data to a CD. The format is technically named CD-WO (Write Once), as codified in the Orange Book Part II. CD-R discs that carry audio and nonaudio data prior to CD replication can be written with the PMCD (premastered CD) format; the disc contains index and other information. CD-R discs with up to 80 minutes of playing time (about 700Mbytes) are available.

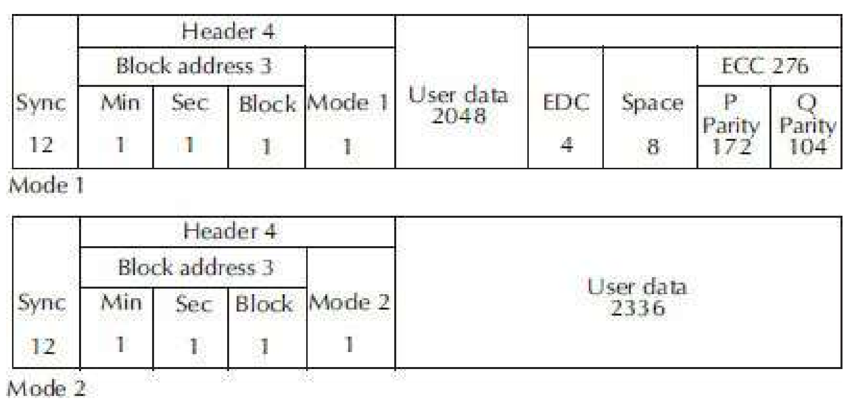


Figure 34-10. CD-ROM Mode 1 contains 2048 bytes of user data with extended error correction, and Mode 2 contains 2336 bytes of user data.

CD-R discs are physically different from Red Book CDs. CD-R discs are manufactured with a pregrooved spiral track with 0.6μm width and 1.6μm pitch; it guides the recording laser along the track. The pregroove is physically modulated with a ±0.03μm sinusoidal wobble with a frequency of 22.05kHz. Recorders use the wobble to control the disc CLV rotation speed. The 22.05kHz groove wobble is also frequency modulated with a ± 1kHz signal; this creates an ATIP (absolute time in pregroove) clocking signal.

CD-R discs are manufactured on a polycarbonate substrate, and contain a metal (e.g., gold or silver) reflective layer, an organic dye recording layer, and a top protective layer. The recording layer is placed between the substrate and reflective layer as shown in Fig. 34-11. Together with the reflective layer it provides a reflectivity of about 73%. A writing laser with wavelength of 775 nm to 795 nm passes through the polycarbonate substrate and heats the recording layer to approximately 250°C, causing it to melt and/or chemically decompose to form a depression or mark in the recording layer. Simultaneously, the reflective layer is deformed. These depressions or marks have a decreased reflectivity. During readout, the same laser, reduced in power, is reflected from the data surface and its changing intensity is monitored.

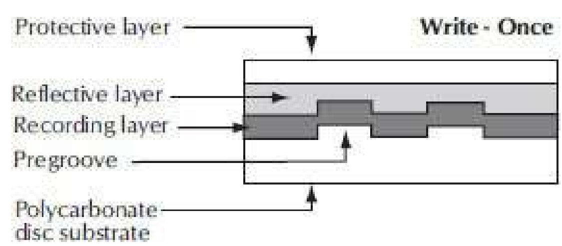


Figure 34-11. CD-R disc construction showing embedded recording layer.

Either cyanine or phthalocyanine organic dye polymers are often used for the recording layer. They are designed to absorb light at about 780nm. Cyanine dye has a relatively broad range of sensitivity to light and is generally reliable in a wide range of recorders and laser powers and writing speeds. Phthalocyanine-based media are generally said to have greater longevity because it is less sensitive to ordinary light and is stable. In some cases, metallized azo dye is used as the recording layer in CD-R media. Organic dye layers are affected by aging. The dye layer will deteriorate over time because of oxidation, material impurities, or exposure to ultraviolet light. CD-R discs will play back on most CD audio players, but the reduced data layer reflectivity can cause playback incompatibility.

Two areas are written to the inner radius (22.35–23mm) of CD-R discs, both inside the Red Book lead-in radius. The PMA (program memory area) contains data describing the recorded tracks, a temporary table of contents, and track skip information. When the disc is finalized, this data is transferred to the TOC. On the innermost radius, the PCA (power calibration area) is used by the recording laser to make an optimal power calibration test recording to determine proper laser recording power. A recording is complete when a lead-in area (with TOC), user data, and lead-out area are written.

The CD-R standard defines both single-session and multisession recording (a session is a recording with lead-in, data, and lead-out areas). In single-session recording, sometimes called disc-at-once recording, an entire disc program is recorded without interruption. Track-at-once recording allows single or multiple tracks to be written in a session. Recorders using track-at-once can also write a single-session CD-R. In multisession recording, sessions can be recorded one or a few at a time. Tracks can be written singly and recording can be stopped after each track. Separate recording sessions are allowed, each with its own lead-in TOC, data, and lead-out areas. Track-at-once recorders allow both multisession and single-session recording. In track-at-once recording, multiple tracks can be written to a session, adding data one track at a time; no lead-in or lead-out is written until the session is closed. CD audio players can read only the first session on a multi-session disc. A partially recorded disc can be played on the CD-R recorder but cannot be played on a CD audio player until the session ends when the final TOC and lead-out areas are recorded. Using the CD portion of the universal disk format (CD-UDF), CD-R recorders can perform packet writing; this allows small amounts of data to be written efficiently without high overhead. Data in a file can be appended and updated without rewriting the entire file.

34.4.3 CD-RW

The CD Rewritable (CD-RW) format allows data to be written and read, and erased and rewritten. The format is technically named CD-E and is described in the Orange Book Part III standard. A CD-RW drive can read, write, and erase CD-RW media, read and write CD-R media, and read CD-ROM and CD audio media. Thousands of rewrite cycles are possible. Any type of data can be written. A CD-RW disc has five layers on a polycarbonate substrate: a dielectric layer, a recording layer, another dielectric layer, a reflective aluminum layer, and a top acrylic protective layer, as shown in Fig. 34-12. As in CD-R, the writing and reading laser follows a pregroove spiral track.

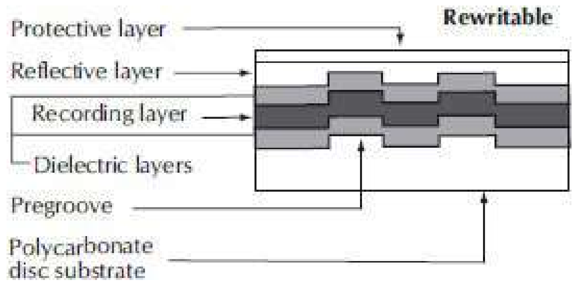


Figure 34-12. CD-RW disc construction showing embedded recording and dielectric layers.

The CD-RW format employs a phase-change recording method, using materials that exhibit a reversible crystalline/amorphous phase change when recorded at one temperature and erased at another. In most cases, a high-reflectivity (crystalline) to low-reflectivity (amorphous) phase change is used to record data, and the reverse to erase. Data is recorded by heating an area of the crystalline layer to a temperature slightly above its melting point and cooled rapidly. The area is amorphous when it solidifies, and the decreased reflectivity is detected by a low power reading laser. Because the crystalline form is more stable, the material will tend to change back to this form. Thus when the area is heated to just below its melting temperature and cooled slowly, it returns to a crystalline state, erasing the data. In some cases, the recording layer comprises gallium antimonide and indium antimonide; other systems use tellurium alloyed with elements such as germanium and indium. The dielectric layers control the optical response of the media and increase the efficiency of the laser by containing the heat in the recording layer. The dielectric layers also thermally insulate and protect the pregroove, substrate, and reflective layer.

The reflectivity of CD-RW discs is only about 15% (amorphous state) and 25% (crystalline). Discs will not play in most CD audio players or CD-ROM drives; however, many DVD players can play

CD-RW discs. Multiread drives are capable of reading lower reflectivity CD-RW discs. They use an AGC (automatic gain control) circuit to boost the gain of the signal output from the photodiodes and compensate for the lower reflectivity and decreased signal modulation. When CD-RW discs are appropriately formatted, the CD-Universal Device Format (CD-UDF) specification permits easy file-by-file rewriting; for example, users can write to CD-RW discs with dragging and dropping.

34.5 Super Audio Compact Disc Format

The Super Audio Compact Disc (SACD) standard provides high-capacity storage to support two-channel and multichannel SACD audio recordings as well as two-channel CD recordings. SACD recordings use 1-bit direct stream digital (DSD) coding with a high sampling frequency to achieve a frequency response to 100kHz and a dynamic range of 120dB in the 0 to 20kHz band. Hybrid SACD discs can hold both a high-density DSD data layer (containing both a 5.1-channel mix and a stereo mix) as well as a Red Book compatible (44.1kHz/16-bit) data layer. SACD players play both SACD and CD discs. To achieve this, dual laser pickups operate at both the SACD 650nm wavelength and the CD 780 nm wavelength. The SACD format also specifies a lossless coding algorithm known as direct stream transfer (DST); it uses an adaptive prediction filter and arithmetic coding to effectively double disc capacity. The SACD standard is described in the Scarlet Book.

34.5.1 SACD Specifications

SACD discs have a 12cm diameter and 1.2mm thickness, the same as a CD. Other specifications allow greater density; the laser wavelength is 650nm, the lens numerical aperture (NA) is 0.60, the minimum pit/land length is 0.40μm, and the track pitch is 0.74μm. A single-layer SACD disc holds 4.7 Gbytes of data; this provides about 110 minutes of playing time for a two-channel stereo DSD recording. Several disc types are specified in the SACD format including single-layer, dual-layer, and hybrid disc constructions. The single-layer disc contains one layer of DSD content (4.7Gbytes); the dual layer contains one or two layers of DSD content (8.5Gbytes with two layers); and the hybrid disc is a dual layer disc that contains one inner layer of DSD content (4.7Gbytes) and one outer layer of Red Book CD content (780Mbytes) that can be played in ordinary CD players. In dual-layer discs, two 0.6mm substrates are bonded together. There is only one data side in all implementations. A semireflective layer (20–40% reflective) covers the embedded inner data layer, and a fully reflective top metal layer (at least 70% reflective) covers the outer data surface. The outer data surface is protected by an acrylic layer and a printed label. Fig. 34-13 shows a hybrid disc and a dual pickup (650and 780nm) reading SACD and CD layers.

SACD players can play both SACD and CD discs (and hybrid SACD discs). CD data is passed to the digital filter and SACD data is applied to the DSD decoder. DSD data is output as a 1-bit signal and applied to a pulse density modulation processor. The data signal is converted to a complementary signal; each logical 1 creates a wide pulse and each logical 0 creates a narrow pulse. A current pulse D/A converter converts the voltage pulse output into a current pulse. This signal is applied to an analog low-pass filter to create the analog audio waveform.

34.5.2 Direct Stream Digital Coding

SACD recordings employ direct stream digital (DSD) coding which uses 1-bit pulse density representation and sigma-delta modulation to code audio signals. Many A/D converters use sigma-delta techniques to sample the input signal at a high sampling frequency. The signal is applied to a decimation filter and quantized for output as a PCM signal at a nominal sampling frequency of 44.1kHz (for CD) and up to 192kHz (for DVD-Audio). Similarly, many D/A converters use oversampling to increase the sampling rate of the output signal and thus move the image spectra from the audio band. DSD coding uses a high sampling frequency, but does not require decimation filtering and multibit PCM quantization; instead, the original sampling frequency is retained. One-bit data is recorded directly on the disc. Moreover, DSD does not employ interpolation (oversampling) filtering during playback.

DSD uses sigma-delta modulation (SDM) and noise shaping. In a simple SDM encoder, the 1-bit output signal is used as a compensation signal. It is delayed by one sample and subtracted from the input analog signal using a negative feedback loop. If the input waveform rises above the value accumulated in the negative feedback loop during the previous sample, the converter outputs a logical 1. Similarly, if the waveform falls relative to the accumulated value, a logical 0 is output. The output pulses represent the magnitude of the input signal; pulse density modulation can be used. Because the integrator in the SDM encoder acts as a low-pass filter, the low-frequency error content is reduced while the high-frequency content is increased. Higher-order noise shaping feedback filters can further decrease error in the audible range of frequencies. In principle, a low-pass filter can decode SDM signals, and also remove high-frequency noise resulting from noise shaping.

On SACD recordings, the DSD modulation employs a sampling frequency of 2.8224MHz and each sample is quantized as a 2-bit word. Overall, the bit rate is thus four times higher than on a CD. In principle, the Nyquist frequency of DSD is 1.4112MHz. However, to remove high-frequency noise introduced by noise shaping, some SACD players incorporate a 50kHz low-pass filter (e.g., −3dB at 50kHz) for use with conventional power amplifiers and speakers. A 20kHz low-pass filter is recommended when making SACD audio measurements. The 1-bit DSD signal can be converted to standard multibit PCM sampling rates.

34.6 DVD Format

In its early development, the DVD format was envisioned as a consumer video disc playback system. Subsequent development expanded the scope of the standard. The resulting family of DVD optical disc formats encompasses video, audio, and computer applications, with both playback-only and recordable technologies. Although its outer physical dimensions are identical to that of CD, one DVD data layer provides about seven times the storage capacity of a CD. This increase is due to the shorter wavelength laser, higher numerical aperture, smaller track pitch, and other aspects. During development, the format was sometimes called Digital Versatile Disc, but that name was never accepted. Instead, the format is simply called DVD.

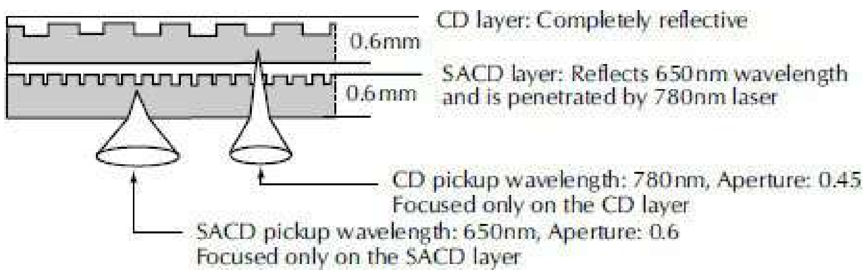


Figure 34-13. A hybrid SACD disc contains two data layers holding CD and SACD data.

The DVD family contains six DVD books: Book A is DVD-ROM (read only), Book B is DVD-Video, Book C is DVD-Audio, Book D is DVD-R (write-once), Book E is DVD-RAM (random access memory), and Book F is DVD-RW (rewritable). In each book, Part 1 defines the physical specifications, Part 2 defines the file system specifications, and subsequent parts define specific applications and extensions. For example, Part 3 defines the video application, Part 4 defines the audio application, and Part 5 defines the VAN extension.

The physical specifications for the DVD-ROM, DVD-Video, and DVD-Audio discs are identical and these read-only formats share disc construction, modulation code, error correction, etc. Discs are 120mm or 80mm in diameter and 1.2mm in thickness, and have two bonded substrates with single or dual data layer per substrate. DVD discs use a pit/land structure to store data. The DVD track pitch is 0.74µm. The track constant linear velocity (CLV) is 3.49m/s on a single layer and 3.84m/s on a dual layer. Minimum/maximum pit length is 0.40/1.87µ m (single layer) and 0.44/2.05µm (dual layer). The laser beam used to read DVDs uses a wavelength of 635nm or 650nm. The objective lens has a numerical aperture of 0.6. A DVD layer can store 4.37Gbytes (measured in 8-bit bytes) of data and multiple data layers provide greater capacity.

34.6.1 DVD Disc Manufacturing

A DVD disc thickness of 1.2mm comprises two 0.6mm substrates, bonded together with the data layers placed near the internal interface. Thinner substrates are optically more resistant to tracking errors that result when a disc is slightly tilted relative to the laser pickup. The dual substrate construction allows manufacturing variants, yielding five types of playback-only discs: DVD-5 (single side, single layer), DVD-9 (single side, dual layer), DVD-10 (dual side, single layer), DVD-14 (dual side, mixed layers with single layer on one side and dual layer on the other side), and DVD-18 (dual side, dual layer). As the nomenclature loosely suggests, five disc capacities are supported: 4.37, 7.95, 8.75, 12.33, and 15.91Gbytes (expressed in 8-bit bytes). When the average data output bit rate is 4.8Mbps, the approximate video playing times are DVD-5 (133min), DVD-9 (241min), DVD-10 (266min), DVD-14 (375min), and DVD-18 (482min).

A single-layer, single-sided DVD-5 disc uses one substrate with a data surface and one blank substrate. Two substrates with data surfaces can be bonded together to form a single-layer, dual-sided DVD-10 disc; the disc is turned over to access the opposite layer. The DVD standard allows data to be placed on two layers in a substrate to create a dual-layer disc that is read from one side comprising a DVD-9 disc. The layers are separated by a clear resin and a very thin semitransparent (semireflective from 25% to 40%) sputtered layer of gold or silicon. Both layers are read from one disc side by moving the objective lens and focusing the reading laser on either layer. The beam either reflects from the lower semireflective layer or passes through it and reflects from the top reflective layer. Because the *SNR* and reflectivity of the interior layer are slightly reduced, the layer uses a faster linear velocity (3.84m/s versus 3.49m/s). Thus the pit length is longer (e.g., the minimum pit length is 0.44μm versus 0.4μm). The interior layer thus has less capacity than the top data layer.

In the manufacture of dual-sided discs, two polycarbonate substrates are independently formed and then bonded together using a hot-melt adhesive or UV-curable bonding. Dual-layer discs can be formed from two 0.6mm substrates; one layer is fully metallized and the other is semireflectively metallized. The two substrates are then bonded together with a layer of UV-cured optically clear photopolymer. This technique can be used to manufacture single-sided discs (such as some DVD-9 discs). Alternatively, a single-layer substrate can be coated with a semitransparent layer followed by a layer of liquid photopolymer that is molded by a second stamper and hardened by exposure to ultraviolet light. After the layer is hardened, a fully reflective metal layer is applied and the substrate is bonded to a second substrate. This technique is used for some DVD-9 and DVD-18 discs. Construction of a dual-layer/dual-side DVD-18 disc is shown in Fig. 34-14.

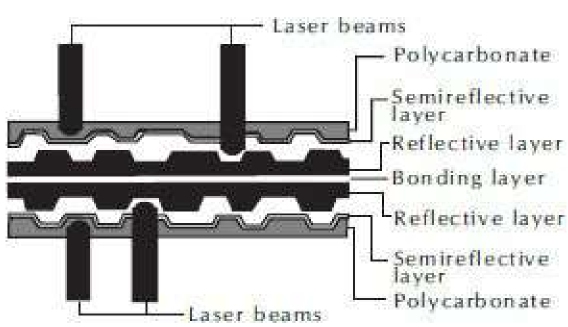


Figure 34-14. Construction of a dual-layer/dual-side DVD-18 disc.

34.6.2 DVD File Format and Coding

The DVD specification describes Universal Disc Format (UDF) Bridge, a file format specifically designed for optical disc storage. Read-only DVDs (DVD-ROM, DVD-Video, and DVD-Audio) use UDF for volume structure and file format, and UDF applies to the write-once and recordable disc formats. However, application-specific parameters are unique to both DVD-Video and DVD-Audio. UDF Bridge is a simplified version based on Part 4 of ISO/IEC 13346 and conforms to both UDF and ISO 9660 (the file format used in CD-ROM). UDF Bridge defines data structures such as volumes, files, blocks, sectors, CRCs, paths, records, allocation tables, partitions, and character sets, as well as methods for reading, writing, and other operations. It is a flexible format that has been adapted to DVD and is backward compatible to existing ISO-9660 operating system software. However, a DVD-Video or DVD-Audio player supports only UDF and not ISO-9660.

In read-only DVD formats, data is stored in files within directories. DVD data is placed on a disc in physical sectors that run continuously without gap from the lead-in to the lead-out area. A DVD data sector comprises 2064 bytes, with 2048 bytes of main data and 16 header bytes; the latter comprises 4 bytes of identification (ID), 8 bytes of other data, and 4 bytes of error-detection code (EDC) data. The 4 bytes of identification data (ID) contain 1 byte of sector information and 3 bytes of sector number. A sync code is added to the head of every 91 bytes in the recording sector. This forms a physical sector. In all, 52 bytes of sync code is added. The 2048 bytes of user data is thus increased to 2418 bytes.

A Reed-Solomon Product Code (RS-PC) uses a combination of two Reed-Solomon codes (C1 and C2) as a product code. It differs from the CD's CIRC code. The two C1 and C2 product codes are (208,192) and (182,172) in length. Error correction is more challenging on a DVD because the pit size is smaller. In addition, because of the thin substrates, surface defects can more readily obscure the data surface. However, RS-PC is more powerful than the double error correction used in the CD-ROM format and provides improved error protection. RS-PC is also more efficient than CIRC in terms of overhead. In the DVD format, all disc types use the same level of error correction.

Read-only DVDs use EFMPlus modulation. It is an 8/16 RLL code and is similar to the EFM code used in CDs; for example, it uses the same minimum (2) and maximum (10) run length and represents logical 1 channel bits as pit/land or land/pit transitions and logical 0 channel bits as no transition. EFMPlus provides a 6% increase in user storage capacity compared to EFM because its coding is more efficient than EFM. Whereas EFM uses merging bits and a single lookup table and simple concatenation rules to suppress low-frequency content, EFMPlus does not require merging bits and uses a more sophisticated look-up method. The EFMPlus encoder defines four look-up tables each with 351 possible source words. In practice, the source codebook size is 344; seven possible words are discarded to allow for a unique 26-bit sync word. Of these, 256 words are used to code input data. The 88 surplus words are used as alternative channel representations to minimize the running digital sum value (DSV) and thus control low-frequency content.

In a DVD player, data passes through a buffer and is evaluated by a navigator/splitter that separates the bit stream into video, subpicture, audio, and navigational information. The video, subpicture, and audio data is decoded; for example, MPEG-2 video data is decoded as is Dolby Digital audio data.

34.7 DVD-Video

The DVD-Video format provides storage and playback of video content such as motion pictures or concert videos with multichannel soundtracks. The format provides the following: at least 133 minutes of digital video, approaching D1 broadcast picture quality, stereo or multichannel digital audio, multiple aspect ratios, up to 8 language soundtracks, up to 32 subtitles, parental control options, and copy protection.

In the DVD-Video format, data in a video disc is organized using the UDF Bridge file format. A DVD-Video zone and DVD-Other zone are defined under a root directory. In the DVD-Video zone, the VIDEO_TS directory (folder) contains menu and presentation data (video, audio, etc.). A Video Manager defines file types and organization of both video and audio data, and Video Title Set (VTS) subdirectories contain video and audio data files (such as MPEG-2 video and Dolby Digital audio). One Video Manager can contain up to 99 VTS subdirectories.

A DVD-Video VAN disc contains video-audio navigation data in a hybrid video-audio disc. VAN discs are video discs but they contain audio information that can be played by DVD-Audio players. Audio data is contained in an Audio Title Set, and video data in a Video Title Set. The Audio Manager and Video manager define file types and organize both audio and video data; both menu and program data is included.

The DVD-Video standard uses the MPEG-2 data compression algorithm to encode its video program. It employs the MPEG-2

Main Profile at Main Level protocol, also known as MP@ML. This is an intermediate level and below the high level sometimes used for DTV. However, MP@ML yields a high-quality picture that equals that of the professional CCIR-601 standard. The video program is stored as 4:2:0 component video (Y, R-Y, B-Y) with progressive scan and picture resolution of 720 × 480 pixels. The average output bit rate of a DVD-Video player is about 4.7Mbps.

Audio Contents

Both stereo and multichannel soundtracks are accommodated in the audio portion of the DVD-Video standard. There can be 1 to 8 independent channels of linear PCM (LPCM), 1 to 6 channels of 5.1-channel Dolby Digital (AC-3), or 1 to 8 channels (5.1 or 7.1) of MPEG-2 AAC audio. A disc can also optionally employ DTS, SDDS, or other audio coding. Dolby Digital is the coding standard used for multichannel soundtracks in the United States (Region 1). The Dolby Digital sampling frequency is 48kHz, the nominal output bit rate is 384 kbps, and the maximum bit rate is 448 kbps. Optionally, DTS codes multichannel audio data at a nominal bit rate of 1.4Mbps. DTS can optionally be used to code 1 to 8 channels of audio, at sampling frequencies ranging from 8kHz to 192kHz. One DTS layer at a sampling frequency of 44.1kHz can hold up to 74min of 5.1-channel audio. MPEG-1 stereo audio is sampled at 48kHz with a maximum bit rate of 384kbps. MPEG-2 multichannel audio (up to eight channels) is also coded at 48kHz; its maximum bit rate is 912kbps. NTSC titles nominally use Dolby Digital, and PAL titles use MPEG-2 audio coding; however, PAL titles can optionally use Dolby Digital coding.

DVD-Video titles carry a redundant LPCM soundtrack employing sampling rates of either 48 or 96kHz and word lengths of 16, 20, or 24 bits. These LPCM configurations are supported: 16/48 (up to eight channels), 20/48 (up to six channels), 24/48 (up to five channels), 16/96 (up to four channels), 20/96 (up to three channels), and 24/96 (up to two channels). The maximum LPCM bit rate is 6.144Mbps on a DVD-Video. Various contents must be accommodated on a DVD-Video. For example, with an average video bit rate of 3.5Mbps, there might be three audio soundtracks each at 0.384Mbps, and four subtitles each at 0.01Mbps, yielding a total bit rate of 4.692Mbps.

34.8 DVD-Audio

The DVD-Audio specification describes a high-fidelity audio storage medium supporting flexibility in the numbers of channels, sampling frequencies, word lengths, and other features such as video elements. Although the DVD-Video format can provide high-quality audio (such as six channels at 48kHz/20-bit audio), its maximum audio bit rate of 6.144Mbps cannot support the highest audio quality levels. Thus DVD-Audio's maximum bit rate was increased to 9.6Mbps. However, six channels of 96kHz/24-bit audio exceeds the maximum bit rate and high bit rates reduce playing time. Thus the Meridian Lossless Packing (MLP) lossless compression algorithm can be optionally employed to reduce bit rate, providing high fidelity and long playing time. This option allows storage of over 74 minutes of multichannel music on a single data layer. All DVD-Audio discs must contain an uncompressed or MLP-compressed LPCM version of the DVD-Audio portion of the program. For added compatibility with DVD-Video players, DVD-Audio may also include video programs with Dolby Digital, DTS, and/or LPCM tracks.

The DVD-Audio format supports a variety of coding methods and recording parameters. Optional audio coding methods include Dolby Digital, MPEG-1, MPEG-2 with/without extension bit stream, DTS, DSD, SDDS, and MLP. Linear PCM (LPCM) tracks are mandatory on all discs; all DVD-Audio players must support MLP decoding. Unlike some 5.1 channel systems (Dolby Digital, MPEG) the LPCM coding used in DVD-Audio does not band-limit the LFE channel; it is a full-bandwidth channel. DVD-Audio is a scalable format and gives flexibility to content providers. When LPCM coding is used, the number of channels (1 to 6), the word length (16, 20, 24 bit), and the sampling frequency (44.1, 48, 88.2, 96, 176.4, or 192kHz) are all allowed. At the highest sampling frequencies of 176.4kHz and 192kHz, only two-channel playback is possible. The audio coding options and the number of disc layers create a range of playback times. For example, a stereo LPCM program on a data layer might play for 258 minutes or 64 minutes, depending on its recording parameters. Similarly, different configurations of multichannel recordings will yield a range of playing times, as shown in Table 34-1. Use of MLP lossless compression, or lossy compression, increases playing times as well.

Audio channels are placed in two Channel Groups (CG). The grouping hierarchically lists mixes that use the front L and R channels, front L, R and C channels, and the corner L, R, Ls, and Rs channels. The sampling frequency and word length of CG1 is greater than or equal to those of CG2. Generally, CG1 assignments are for front channels, and CG2 assignments are for rear channels. Channels can be assigned as groups of mono to six channels, and different word lengths and front and rear channels can use different sampling frequencies. For example, to reduce storage requirements, front channels could be coded at 24/96 and the rear channels coded at 16/48. The sampling frequencies must be related by a simple integer such as 48/96/192kHz or 44.1/88.2/176.4kHz.

Table 34-1. Examples of Playing Times in DVD-Audio Discs, Not Using MLP Coding.

Audio Contents Combination	Channel Combination	Playback Time per Disc Side			
		12 cm Disc		8cm Disc	
		Single Layer	Dual Layer	Single Layer	Dual Layer
2 channel only	48k/24-bit/2 ch	258min	469min	80min	146min
2 channel only	192k/24-bit/2 ch	64min	117min	20min	36min
2 channel only	96k/24-bit/2 ch	125min	227min	39min	70min
Multichannel Only	96k/24-bit/6 ch	86min	156min	27min	48min
2 ch and multichannel	96k/24-bit/2 ch +96k/24-bit/3 ch and 48k/24-bit/2 ch	76 min each	135min each	23min each	41min each

Meridian Lossless Packing (MLP)

Meridian Lossless Packing (MLP) is an audio coding algorithm used to achieve lossless data compression. It reduces average and peak audio data rates and hence reduces storage capacity requirements. MLP packs audio data more efficiently, reducing file size while preserving data contents. MLP can ensure that the output signal is exactly the same as the input signal by checking the MLP-coded file and confirming its bit accuracy. The compression achieved by MLP depends on the music being coded. Very approximately, it gives a 1.85:1 compression ratio; thus reducing the bit rate by almost 50%, doubling playing time with no loss of audio quality. For example, without compression, 96kHz/24 bit audio requires 2.304Mbps per channel. Thus a six channel recording would require 13.824Mbps, exceeding DVD-Audio's 9.6MHz maximum bit rate; thus LPCM cannot be used in the configuration. In contrast, MLP allows six-channel 96kHz/24-bit recordings; it may achieve bandwidth reduction of 38% to 52%, reducing bandwidth to 6.6 to 8.6Mbps, allowing a playing time of 73 to 80 minutes on a DVD-5 disc. In the two-channel stereo mode of 192kHz/24-bit, MLP provides a playing time of about 117 minutes, versus a playing time of 74 minutes for LPCM coding.

Unlike lossy perceptual coding methods, MLP preserves bit-for-bit content of the audio signal. MLP provides less compression than lossy methods, the degree of compression depends on the audio signal content, and the output bit rate can continually vary according to signal conditions; however, a fixed data rate mode is provided. MLP is a mandatory coding option. Thus, all DVD-Audio players must support MLP decoding, but use of MLP on discs is optional for content providers. MLP may be used on a track-by-track basis. All of the DVD-Audio sampling frequencies are supported by MLP and quantization may be selected for 16 to 24 bits in 1-bit steps. MLP can code both stereo and multichannel signals simultaneously.

34.9 Other DVD Formats

The DVD-Video format is defined in Book B and DVD-Audio is defined in Book C. The DVD family also includes DVD-ROM (Book A), DVD-R (Book D), DVD-RAM (Book E), and DVD-RW (Book F). Books A, B, and C use the UDF Bridge file format while Books D, E, and F use the UDF format. The DVD-ROM, DVD-R, DVD-RAM, and DVD-RW formats are used primarily as computer peripherals or in professional authoring environments.

All DVDs are essentially DVD-ROMs, and all DVDs use the basic

UDF format. Some DVD applications, such as DVD-Video, place specialized material in a specific place such as the DVD-Video zone. Content contained in the DVD-Other zone may be quite varied. DVD-ROM uses that provision for nonspecific storage, acting as a bit bucket formatted with UDF. DVD-ROM are playback-only media used to store data, software, games, etc. With appropriate software, DVD-ROM drives can play DVD-Video and DVD-Audio.

The DVD-R format offers write-once capability to permanently record data. DVD-Rs use a CLV wobbled pregroove to generate a carrier signal used for motor control, tracking and focus. DVD-Rs use pits and lands (known as land prepits) molded into land areas between grooves to encode the time address and other prerecorded signals. A cyanine organic dye recording layer may be used, with a 635 or 650nm laser. The reading laser tracks the pregroove, but the light shines on the prepits peripherally to create a secondary signal that is extracted from the main signal. Discs can use the same reference velocity and track pitch as molded discs to achieve the same unformatted storage capacity. There are two parts to the DVD-R specification: DVD-R General and DVD-R Authoring; both yield discs playable on DVD-Video players.

DVD-R recorders perform an optimum power calibration (OPC) procedure to determine the correct laser writing power for particular discs, using a power calibration area (PCA) on discs to test laser writing power. A recording management area (RMA) saves calibration information, disc contents, and recording locations and remaining capacity information, recorder and disc identifiers for copy protection. The remainder of the disc comprises the information area containing the lead-in, data recordable area, and lead-out. The lead-in contains information on disc format, specification version, physical size and structure, minimum readout rate, recording density, and pointers to the location of the data recordable area where user data is recorded. The lead-out marks the end of the recording area. DVD+R is another write-once format using a dye recording layer and CLV rotation. Capacities of 4.7 and 8.5 (DL) Gbytes are available. DVD+Rs are generally compatible and can be played in many DVD players.

The DVD-RW format allows data rewriting; the specification is an extension to the DVD-R format. Discs use a phase-change recording mechanism and a multilayer disc structure with dielectric layers above and below the recording layer. Data is recorded into a wobbled pregroove with CLV; relatively large data blocks are written. The recording layer may use a silver, indium, antimony, and tellurium compounded layer and allows perhaps 1,000 writing cycles.

DVD-RAM (random access memory) is a true random-access, nonsequential storage format. It uses a phase-change recording mechanism and a wobbled land and groove disc design. Data may be recorded on both planar surfaces of the groove and land; a wider track pitch is employed. This technique doubles disc capacity. Discs contain preembossed pit areas (for every 2k sector) containing addressing header information and zoned constant linear velocity rotational control. A DVD-RAM disc allows perhaps 100,000 rewrite cycles and offers a high degree of stability for archiving integrity.

DVD+RW is a rewritable format that uses phase-change media, a wobbled pregroove, and CAV or CLV rotation, for either raw data transfer or faster data access. Data is recorded in the pregroove, not on the land. Data addresses are represented by modulation of the pregroove; this necessitates somewhat larger writing blocks. Over 100,000 rewrite cycles are possible.

34.10 Blu-ray Disc Format

Blu-ray is a consumer disc format widely used to play motion pictures with high-definition resolution. Its picture quality is greater than that of standard-definition DVD and meets broadcast high-definition DTV standards. Blu-ray is also used to distribute video games, and recordable and rewritable media are available. Blu-ray accommodates a variety of lossy and lossless audio formats. It can reproduce high-quality multichannel sound and also provides long playing times. In addition, a 3D video Blu-ray specification has been introduced.

34.10.1 Overview

As with CD and DVD, Blu-ray disc media are available as prerecorded, recordable and rewritable formats. The specifications are known respectively as BD-ROM, BD-R and BD-RE. The three disc types have the same data capacity. All three types can hold a single data layer or dual data layers. The layers are independent and both layers are read from the same side of the disc. The Blu-ray format uses a 405nm wavelength laser for reading and recording.

A single-layer Blu-ray disc provides a storage capacity that is about 35 times greater than the capacity of a CD disc, and about five times greater than the capacity of a DVD disc. Blu-ray disc diameter is either 12cm or 8cm; these disc dimensions are identical to those of CD and DVD discs. The greater storage capacity is due to several improvements, including a shorter wavelength laser and an objective lens with higher numerical aperture. These permit a narrower track pitch and smaller pit sizes. The three disc formats are compared in Table 34-2.

34.10.2 Disc Capacity

A single-layer Blu-ray BD-ROM disc known as BD-25 can store about 25Gbytes; it is capable of holding at least 2 hours of high-definition video. A dual-layer disc known as BD-50 holds about 50 Gbytes of data. Similarly, BD-27 and BD-54 disc formats are specified. A Mini Blu-ray BD-8 disc with 8-cm diameter and single layer holds about 7.8Gbytes and a Mini Blu-ray BD-16 disc with dual layers holds about 15.6Gbytes. In terms of audio storage, a 50-Gbyte disc can hold over ten hours of 192kHz/24-bit PCM stereo audio, or over 200 hours of 5.1-channel Dolby Digital audio. Table 34-3 compares different Blu-ray disc types in terms of storage capacity and typical playing time.

Table 34-2. Principal Specifications of the CD, DVD and Blu-ray Formats

	CD	**DVD (single layer)**	**Blu-ray (single layer)**
Storage capacity	0.7Gbytes	4.7Gbytes	25Gbytes
Track pitch	1.6µm	0.74µm	0.32µm
Minimum pit length	0.8µm	0.4µm	0.15µm
Storage density	0.41 Gbits/in^2	2.77 Gbits/in^2	14.73 Gbits/in^2

34.10.3 Disc Design

As with CD-ROM and DVD-ROM discs, BD-ROM discs store binary data using pits embossed on a substrate. During playback, the pits yield a change in reflected intensity in the reading laser and those changes are decoded to produce the stored data content. The data layer(s) are placed on the near side of the substrate; data is read from underneath the disc.

The structure of a single-layer Blu-ray BD-ROM disc is shown in Fig. 34-15. Both single- and dual-layer discs use a substrate with a nominal thickness of 1.1mm. In a single-layer disc, the data layer of the substrate is covered by a reflective layer which is topped by a cover layer of 0.1mm thickness. In a dual-layer disc, the substrate is covered by two data layers each separated by a transparent separation layer of about 0.025mm. The inner data layer (L0) is covered by a reflective layer and the outer data layer (L1) is covered by a semireflective (or semitransparent) layer which in turn is covered by a transparent cover layer of 0.075mm. The disc is read through the transmission stack; this is through the cover layer (or through the cover layer, outer data layer and separation layer). The reading laser can be focused on either data layer to permit this.

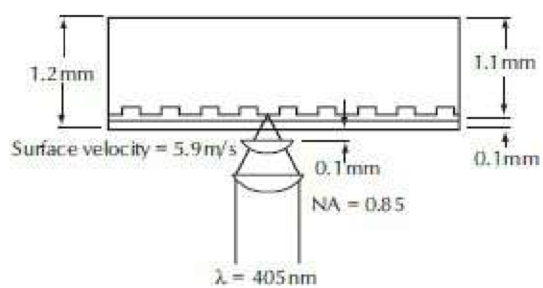


Figure 34-15. Blu-ray discs use a 1.1mm substrate and a 0.1mm protective layer. A single-layer disc is shown, but dual-layer discs can be used.

Unlike CD and DVD discs, in Blu-ray discs, the optical path is through the cover layer, not the substrate. Thus the substrate's optical characteristics are not critical; for example, the substrate can be opaque. Because the objective lens is close to the data layer, optical aberration caused by disc tilt is limited. The cover layer provides scratch resistance and reduces the need for a disc cartridge. The Blu-ray disc specification requires that the cover layer be scratch resistant. A separate protective layer on top of the cover layer is optional.

Table 34-3. Disc Storage Capacities and Typical Playing Times

Disc type	Diameter (cm)	Disc structure	Capacity (billions of bytes) (10^9)[1]	Capacity (Gigabytes) (2^{30})[1]	Typical video playback time (hours)[2]	2-Channel audio-only playback time (hours)[3]	5.1-Channel audio-only playback time (hours)[4]
BD-8	8	1 layer	7.791	7.256	0.7	3.8	27.1
BD-16	8	2 layers	15.582	14.512	1.4	7.5	54.1
BD-25	12	1 layer	25.025	23.306	2.3	12.1	86.9
BD-27	12	1 layer	27.020	25.164	2.5	13.0	93.8
BD-50	12	2 layers	50.050	46.613	4.6	24.1	173.8
BD-54	12	2 layers	54.040	50.329	5.0	26.1	187.6

[1] Capacities can be increased slightly if track pitch is decreased.

[2] High-definition MPEG-2 video with audio tracks, bit rate of 24Mbps.

[3] PCM, bit rate of 4.608Mbps (coded at 24bit/96kHz).

[4] Dolby Digital, bit rate of 0.640Mbps.

CD, DVD and Blu-ray discs all use a spiral pit track. Also, all three formats begin reading from the inner radius and read outward. When a Blu-ray disc has two layers, when the end of the outer layer is reached, the laser can refocus on the inner layer and begin reading inward. This type of tracking is known as opposite track path, or reverse spiral dual layer. A buffer memory is used to allow continuous data output when switching layers. In addition, to effectively extend the buffer size, the data at the switch point can be written with a lower bit rate.

Blu-ray uses a 405nm wavelength blue-violet laser. For example, an indium gallium nitride (InGaN) laser can be used. Numerical aperture is 0.85. Shorter light wavelength, higher NA and a thin cover layer over the data layer allow a (diffraction limited) spot size of 580 nm. Along with efficient data modulation, this permits high data density. Track pitch is 0.32μm. CLV disc rotation is used. Minimum mark length (2T) is 0.149μm for 25- and 50-Gbyte discs. Linear recording density varies for other disc capacities.

Two pit configurations are permitted for Blu-ray discs. As seen from the optical beam, a concave pit is defined as in-pit, and a convex pit is defined as on-pit. Generally, replicated BD-ROM pits use an in-pit configuration. However, for the inner layer of a dual disc, one manufacturing method creates the pits by replicating them in the space layer which yields in-pits. Another method replicates on-pits in the outer layer. Both manufacturing methods yield discs within the Blu-ray specification. In Blu-ray, a pit depth of about λ/4 yields a low jitter figure.

To maintain compatibility, the reflectivity of BD-ROM discs was designed to be similar to that of the previously-released BD-RE specification. As with CD and DVD, aluminum was selected as the preferred metallization material. Two ranges of reflectivity are specified for the BD-ROM data layer: 35 to 70% for a single-layer disc, and 12 to 28% for a dual-layer disc. These figures are somewhat higher than those used in the BD-RE format.

34.10.4 Disc Manufacturing

The manufacture of BD-ROM discs is similar to the manufacture CD and DVD discs. However, many of the tolerances for Blu-ray discs are more critical, and several new manufacturing techniques are used. In most cases, a 1.1mm substrate with pits is formed by injection molding. The reflective layer is deposited by sputtering. In the case of a single-layer disc, the outer cover layer can be applied over the fully reflective layer with a UV-curable resin using spin coating. Alternatively, cover sheets can be punched from a roll, resin is applied to the substrate and the cover sheet is bonded to the substrate by curing the resin with UV irradiation. Alternatively, the cover sheet can be bonded with a pressure-sensitive adhesive (PSA).

The manufacturing process for a dual-layer disc adds several steps. The inner data layer on the substrate is fully metalized by sputtering and a separation layer (a UV-curable adhesive called HPSA) is formed on the substrate by pressure bonding. A stamper is pressed into the HPSA to replicate a second pit surface. UV light irradiates the underside. The stamper is removed and a semireflective layer is sputtered on the HPSA. The cover layer is applied using any of the same methods as for single-layer discs.

34.10.5 Optical Pickup Design

Various pickup designs can be used for Blu-ray playback. In one design, an integrated three-format pickup uses a single aspherical objective lens and a polarized holographic optical element (HOE) device. The HOE uses a birefringent material sandwiched between two substrates. The birefringent material has the same index of refraction as the bonding material for a certain polarization direction, but a different index of refraction for a different perpendicular polarizing direction. The HOE does not affect the 405nm (Blu-ray) wavefront, but at 650nm (DVD) and 780nm (CD) the polarization direction is perpendicular to that at 405 nm resulting in phase distributions. This yields a non-diffracted beam for Blu-ray and diffracted beams for DVD and CD. The HOE is designed to control this phase distribution and also compensates for spherical aberration caused by thickness differences in the CD and DVD substrates.

34.10.6 Audio Codecs

The Blu-ray specification specifies several audio and video codecs to allow flexibility in authoring, and to provide wide compatibility between disc content and player decoders. The BD-ROM specification supports, as either mandatory or optional, seven different audio formats. Players must support PCM, Dolby Digital, and DTS codecs. Optional codecs are: Dolby Digital Plus, Dolby TrueHD lossless, DTS-HD High Resolution Audio and DTS-HD Master Audio lossless.

In terms of mandatory or optional support, several qualifications must be noted. In part, this is because some Dolby and DTS codec formats are extensions to legacy formats and whether they are considered to be mandatory or optional depends on the bitstream configurations. For example, players support Dolby Digital Plus only for more than 5.1 channels; the base 5.1 channels are coded with core Dolby Digital; only higher channels are coded with Dolby Digital. For Dolby TrueHD, it is mandatory that players support the core Dolby Digital bitstream up to 640kbps; support of the lossless bitstream is optional. The core portions of DTS-HD extensions (High Resolution and Master Audio) are mandatorily supported up to 1.509Mbps; support of the HD portion is optional. The Blu-ray specification imposes various technical constraints on audio codecs, for example, in terms of maximum bit rate and number of channels.

The Blu-ray format allows up to 32 primary audio bitstreams (such as main soundtracks) and up to 32 secondary audio bitstreams (such as commentary). It supports up to eight channels coded with PCM, Dolby or DTS formats and may be coded in a variety of channel configurations such as monaural, stereo, 5.1, 7.1, and so on. Sampling frequencies up to 96- and 192kHz are allowed. Sampling frequencies of 44.1 and 88.2kHz are not supported. Some companies have released “high resolution” Blu-ray audio-only discs. For example, some releases contain 24bit/96kHz or 192kHz music files. File encoding may employ LPCM, Dolby TrueHD or DTS Master Audio. Discs will play in all Blu-ray players.

34.11 Blu-ray Recordable Formats

Two recordable Blu-ray disc formats are in common use. The recordable (BD-R) format allows discs to be written to once, and the rewritable (BD-RE) format allows discs to be written to, erased and written to multiple times. Most recordable Blu-ray discs use phase-change technology; either GST (GeSbTe stoi-chiometrical composition) or eutectic phase-change media can be used. BD-R discs may alternately use organic-dye or inorganic-alloy recording. BD-RE discs use an in-groove method while BD-R discs can use in-groove or on-groove recording.

Recordable disc capacities are generally 25 Gbytes for single-layer discs, and 50 Gbytes for dual-layer discs. The BD-RE version 1.0 specification describes disc capacities of 23.3, 25 and 27 Gbytes/layer. The BD5 format stores up to 4.5 Gbytes on a single-layer disc and the BD9 format stores up to 9 Gbytes on a dual-layer disc. Both use inexpensive DVD-type discs. Some Blu-ray players will not play these discs.

There are three versions of BD-RE discs. Version 1.0: unique BD file system and not computer compatible. Version 2.0: UDF 2.5 file system, for computer use, use of AACS. Version 3.0: 8-cm camcorder disc diameter added, backward compatible with Version 2.0. There are also three versions of BD-R discs. Version 1.0: UDF 2.5 file system, for computer use, use of AACS (same as BD-RE Version 2.0). Version 1.2: adds Low to High (BD-R LTH) standard. Version 2.0: 8-cm camcorder disc diameter added, backward compatible with Version 1.0 (same as BD-RE Version 2.0). BD-LTH discs may use different recording media for write-once operation; however, some Blu-ray players cannot play these discs.

In the BD-RE and BD-R formats, a wobbled pre-groove is used for addressing, similar to the method employed in the DVD+RW format. In particular, wobbling addressing is in the radial direction based on minimum shift keying (MSK) modulation and formatted in blocks of 64kbytes. The wobble frequency (1×) is 956.522kHz. An ADIP (address in pregroove) method is used over 56 wobble periods. Nominal wobble length is 5.1405μm with 69 channel bits per wobble. Binary ADIP information is expressed as the position where the sinusoidal wobble is deviated by minimum shift-keying modulation. The shift-keying can be influenced by reading defects; to overcome that, a sawtooth wobble (STW) signal is used to add secondary harmonics to the sinusoidal wobble and binary 0 and 1 correspond to the polarity of the added harmonics.

Bibliography

A. Nishio, et al., “Direct Stream Digital Audio System,” *Audio Engineering Society* preprint 4163, Copenhagen, May 1996.

B. H. Suzuki, “DVD-Audio Specifications,” Proc. Audio Engineering Society 17th International Conference on High-Quality Audio Coding, Florence, Italy, September 2–5, 1999.

Blu-ray Disc Association, “*Blu-ray Disc Format: 1.A Physical Format Specifications for BD-RE*,” 2nd ed., February 2006.

Blu-ray Disc Association, "*Blu-ray Disc: 1.C Physical Format Specifications for BD-ROM*," 5th ed., March 2007.

Blu-ray Disc Association, "*Blu-ray Disc Format: 2.B Audio Visual Application Format Specifications for BD-ROM*," March 2005.

Blu-ray Disc Association, "*Blu-ray Disc Recordable Format: Part 1 Physical Specifications*," February 2006.

Blu-ray Disc Association, "*Blu-ray Disc Rewritable Format: Audio Visual Application Format Specifications for BD-RE*," Version 2.1, March 2008.

Blu-ray Disc Association, "*Blu-ray Disc Rewritable Format: Audio Visual Application Format Specifications for BD-RE*," Version 3.0, March 2008.

Blu-ray Disc Founders, "*Blu-ray Disc Format: General*," August 2004.

Blu-ray Disc Founders, "*Blu-ray Disc Format: 3. File System Specifications for BD-RE, R, ROM*," August 2004.

Blu-ray Disc Founders, "*Blu-ray Disc Format: 4. Key Technologies*," August 2004.

D. Goedhart and R. J. van de Plassche, and E. F. Stikvoort, "Digital-to-Analog Conversion in Playing a CD," *Philips Technical Review*, vol. 40, no. 6, 1982.

D. J. Parker, "Writable DVD," *EMedia Professional*, vol. 12, no. 1, January 1999.

D. R. Guenette and D. J. Parker, "CD, CD-ROM, CD-R, CD-RW, DVD, DVD-R, DVD-RAM: The Family Album," *EMedia Professional*, vol. 10, no. 4, April 1997.

DVD Forum, www.dvdforum.org.

G. Bouwhuis, et al., *Principles of Optical Disc Systems*, Adam Hilger Ltd., 1985.

G. O'Hara, "Understanding the Blu-ray Format," *Mix Mag.*, vol. 33, no. 12, December 2009.

H. Hoeve, J. Timmermans and L. B. Vries, "Error Correction and Concealment in the CD System," *Philips Technical Review*, vol. 40, no. 6, 1982.

IEC 908: *CD Digital Audio System*, 1987.

ISO/IEC 14496-10:2005 "Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding." *International Organization for Standardization*, Geneva, Switzerland (2005).

J. P. J. Heemskerk and K.A.S. Immink, "CD System Aspects and Modulation," *Philips Technical Review*, vol. 40, no. 6, 1982.

J. R. Stuart, "Coding Methods for High-Resolution Recording Systems," *Audio Engineering Society* preprint 4639, New York, September 1997.

J. Taylor, et al, *Blu-ray Disc Demystified*, McGraw-Hill, 2009.

J. Taylor, *DVD Demystified*, third edition, McGraw-Hill, 2006.

J. Verbakel, et al., "Super Audio CD Format," *Audio Engineering Society* preprint 4705, Amsterdam, May 1998.

K. A. S. Immink, "The CD Story," *Journal of the Audio Engineering Society*, vol. 46, no. 5, May 1998.

K. A. S. Immink, "The Digital Versatile Disc (DVD): System Requirements and Channel

Coding," *SMPTE Journal*, vol. 105, no. 8, August 1996.

K. C. Pohlmann, *The Compact Disc Handbook*, Second Edition, Madison, WI: A-R Editions Inc., 1992.

K. C. Pohlmann, *Principles of Digital Audio*, Sixth Edition, New York, NY: McGraw-Hill, 2011.

K. C. Pohlmann, "The CD Formats: Technology and Applications," *Journal of the Audio Engineering Society*, vol. 36, no. 4, April 1988.

L. B. Vries, et al, "The CD Digital Audio System: Modulation and Error Correction," *Audio Engineering Society* preprint 1674, New York, October/November 1980.

M. A. Gerzon, et al., "The MLP Lossless Compression System for PCM Audio," *Journal of the Audio Engineering Society*, vol. 52, no. 3, March 2004.

M. G. Carasso, J. B. H. Peck, and J. P. Sinjou, "The CD Digital Audio System," *Philips Technical Review*, vol. 40, no. 6, 1982.

M. Johnson, C. Crawford, and C. Armbrust, *High-Definition DVD Handbook: Producing for HD DVD and Blu-ray Disc*, McGraw-Hill, 2007.

Moving Pictures Expert Group, *DVD Resources*, www.mpeg.org/MPEG/dvd.html.

M. Zink, P. Starner, and B. Foote, *Programming HD-DVD and Blu-ray Disc: The HD Cookbook*, McGraw-Hill, 2008.

N. Fuchigami and T. Kuroiwa, "An Overview of the DVD-Audio Format," *Audio Engineering Society* preprint H-2, 9th Regional Convention of the Audio Engineering Society, Tokyo, June 23–25, 1999.

N. Fuchigami and T. Kuroiwa, B. H. Suzuki, "DVD-Audio Specifications," *Journal of the Audio Engineering Society*, vol. 48, no. 12, December 2000.

P. G. Craven and M. Gerzon, "Lossless Coding for Audio Discs," *Journal of the Audio Engineering Society*, vol. 44, no. 9, September 1996.

SMPTE, "421M Television – VC-1 Compressed Video Bitstream Format and Decoding Process." *Society of Motion Picture and Television Engineers*.

R. ten Kate, "Disc Technology for Super-Quality Audio Applications," *Audio Engineering Society* preprint 4565, New York, September 1997.

Part 7

Design Applications

Chapter 35

DSP Technology

by Dr. Craig Richardson

35.1 Introduction

Over the past fifty years or so, the field of digital signal processing (DSP) has grown from its origins as a collection of techniques for simulating the behavior of analog systems on digital computers into one of the most widely studied and universally used tools in modern technology. The use of DSP algorithms and implementations has become the rule rather than the exception, with applications in many areas such as high-fidelity audio, communications, radar, sonar, image processing, robotics, seismology, meteorology, applied physics, and more. The remarkable growth of this discipline is largely due to two factors. First, DSP is a powerful problem-solving tool because it exploits the theoretical insights of discrete system theory to describe, analyze, and implement many interesting linear and nonlinear algorithms. Second, and more important, there is a special relationship between VLSI technology and DSP applications. The rapid development of digital integrated circuit technology has continually reduced the cost and increased the speed of the arithmetic operations necessary for DSP applications. In addition, DSP algorithms, which have demanding computational requirements but usually a very regular structure, are very well matched to the capabilities of VLSI. Integrated circuits are making complex DSP applications possible, and DSP applications have become a major motivating factor for building fast, complex integrated circuits. Perhaps the most visible embodiments of this phenomenon are the families of DSP microprocessors commonly called DSP chips. These chips have already had an immense impact on technology and are currently in the process of revolutionizing much of our industrial and technological base.

This chapter will introduce some of the important aspects of DSP technology including the fundamentals of DSP, the sampling process for converting analog signals to digital signals, the algorithm development process, and an introduction to programmable DSP devices. References are provided for finding additional information.

35.2 Digital Signal Processing

DSP is a technology and technique for analyzing and extracting information from signals, synthesizing signals, and manipulating signals. While the acronym DSP typically stands for digital signal processor—the actual microprocessor/computer that is used to implement the system, the term is often used as both a noun and an adjective to describe products such a DSP device. Common applications of DSP include cellular telephones, portable digital audio players, surround sound receivers, high-resolution video conferencing systems, compact disc players, digital cameras, and high-speed modems.

As with many disciplines, there are different perspectives and different layers of abstraction from which to explore DSP. For the purposes of this chapter, DSP will be approached and introduced from the theoretical, physical, and embedded software perspectives.

The theoretical perspective is concerned with the question "is something possible" and is built from fundamentals of DSP theory. This foundation includes linear system theory, complex number theory, and applied mathematics. The theoretical level provides a common language for DSP researchers to study and advance the state of the art.

The physical perspective is concerned with the devices that are used to implement DSP systems. These devices include the programmable digital signal processors that perform mathematical operations at a very high speed, and the details, and best practices of converting an analog signal into a digital signal and then back to an analog signal.

The embedded software perspective is concerned with the actual software and sequence of operations that makes the digital signal processors perform the desired tasks. This software is called *embedded* because it is executed internally on the DSP device and is only user accessible through some user interface or application programming interface, effectively hidden or embedded in the product, hiding the implementation details from the user.

35.3 DSP Signals and Systems Theory

The concepts of signals and systems are critical to an understanding of DSP. Signals can be a function of continuous time (i.e., analog) or of discrete time. Continuous-time signals have a signal value at any given instant of time while discrete-time signals only have a signal value at discrete instants of time. Values of discrete-time signals between the samples, or time intervals, are determined by mathematically interpolating between the known sample values.

Signals represent the data that is to be processed. Examples include an audio file that needs to be compressed for low bit-rate storage or transmission or an image that will be searched for a particular object. A system is a transformation that maps an input signal (or multiple input signals) to an output signal (or multiple output signals), i.e., the black box that maps inputs to outputs. In the music compression example, the output signal could be a smaller file that was created by compressing an input signal while still preserving a high-level of fidelity. In the image example the output signal could simply be a yes/no decision along with positioning information. DSP systems are typically designed from simpler subsystems much like computer software is developed—subroutine by subroutine (one level of abstraction at a time). This section will introduce some fundamental systems and also introduce the useful properties that some systems possess.

35.3.1 Sequences

Discrete-time signals, also called sequences, are most often created by sampling analog, or continuous-time, signals. By sampling a continuous-time signal, a sequence of samples, really a sequence of numbers, can be processed and manipulated in a digital signal processor. Before going further into the sampling process, an introduction to signal and system theory will be presented, starting with discrete-time signals.

Discrete-time signals are represented mathematically as a sequence of numbers. The notation used will denote a sequence, x, as $x = \{x[n]\}$ where n is the index of the n^{th} element in the sequence. In terms of notation, $x[n]$ represents both the nth sample in the sequence and the entire sequence that is a function of n. The index, n, can range over all values from $-\infty$ to $+\infty$.

From a programming perspective, a sequence can be thought of as an infinitely large array of data indexed by an integer variable. In reality, an infinitely long array is not practical, so a sequence is usually represented as a continuous stream of data. Often it is assumed that the sequence starts at time = 0 ($n = 0$) and ends some finite time later ($n = M$).

There are several sequences that are fundamental building blocks of DSP systems. These are the unit impulse, the unit step sequence, and the sinusoid (cosine or sine). The unit impulse is a signal that has a value of 1 at index $n = 0$ and is 0 everywhere else as shown in Fig. 35-1. Mathematically this is denoted by

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \qquad (35\text{-}1)$$

Having defined the unit impulse, it is possible to represent a sequence $x[n]$ as a sum of delayed impulses that have a value of $x[k]$ at $n = k$. Mathematically this is formulated as

$$x[n] = \sum_k x[k]\delta[n-k] \qquad (35\text{-}2)$$

which simply says that the value of $x[n]$ is the collection of its individual samples at time $n = k$.

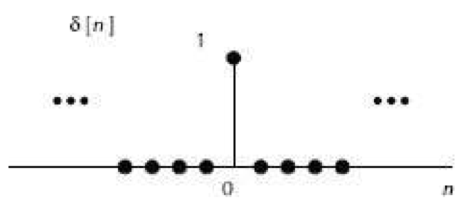


Figure 35-1. Unit impulse sequence has a value of 1 at $n = 0$ and is 0 everywhere else.

The unit step is a signal that starts at index 0 with value 1 and has value 1 for all positive indices as shown in Fig. 35-2. Mathematically, this is denoted by

$$u[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases} \qquad (35\text{-}3)$$

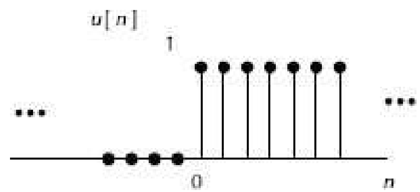


Figure 35-2. The unit step sequence has a value of 1 for $n \geq 0$ and is 0 everywhere else.

The cosine signal is a sinusoid of frequency ω and phase φ. An example of the cosine signal is shown in Fig. 35-3. Mathematically, the cosine signal is denoted by

$$\cos[n] = \cos(\omega n + \phi) \qquad (35\text{-}4)$$

All sequences can also be represented by the numbers that are the sample values $x[n]$. Table 35-1 shows the sample values for the sequence in Fig. 35-4. Only the first sixteen sample values are listed because the sequence repeats itself after the 16th value ($x[15]$).

35.3.2 Systems

Systems transform input signals into output signals. Some commonly used systems include the ideal delay system that delays the output relative to the input and the moving average system that performs some simple low-pass filtering. Systems operate on a signal by operating on each sample individually or groups of samples at a time. For instance, multiplying a sequence by a constant can be implemented by multiplying each sample of the sequence by the constant. Similarly, the addition of two sequences is performed by adding the signals together on a sample-by-sample basis. Other systems, such as an MPEG audio compression system may operate on frames of data that have 1152 samples in each frame or other frame-sizes that may be convenient to the algorithm being implemented. The choice of whether to operate sample-by-sample or frame- by-frame is made by the system designer and algorithm developer.

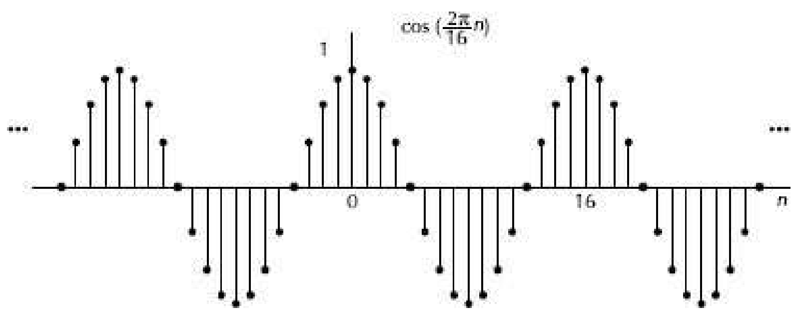


Figure 35-3. A cosine sequence of period 16. This particular cosine sequence is an infinite sequence of values that repeat with a period of 16 samples.

Table 35-1. The Values of the Signal x[*n*] in Fig. 35-4

$x[0]$	1.0000	$x[6]$	−0.7071	$x[12]$	0.0000
$x[1]$	0.9239	$x[7]$	−0.9239	$x[13]$	0.3827
$x[2]$	0.7071	$x[8]$	−1.0000	$x[14]$	0.7071
$x[3]$	0.3827	$x[9]$	−0.9239	$x[15]$	0.9239
$x[4]$	0.0000	$x[10]$	−0.7071	...	
$x[5]$	−0.3827	$x[11]$	−0.3827		

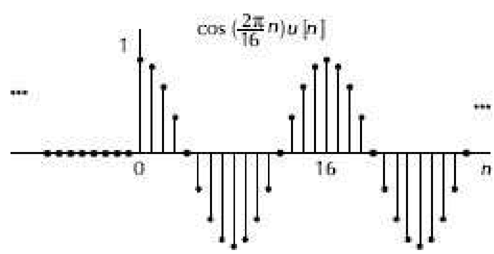


Figure 35-4. The product of the cosine sequence with the unit step sequence, $u[n]$. Notice that all the signal values for $n < 0$ are set to 0.

A fundamental system is the ideal delay. The ideal delay system delays or advances a sequence by the delay amount. This system is defined by the equation

$$y[n] = x[n - n_d], -\infty < n < \infty \qquad (35\text{-}5)$$

where,
n_d is an integer that is the delay of the signal.

The ideal delay system creates an output $y[n]$ by shifting the input signal, x, by n_d samples to the right when n_d is positive. This means that the value of the output signal $y[n]$ at a particular index n is the value of the input signal at index $n - n_d$. For example if the signal is delayed by three samples, then $n_d = 3$ and the output value $y[7]$ is equal to the value of $x[4]$—i.e., the value of $x[k]$ at $k = 4$ now appears at $y[j]$, $j = 7$. The system shifted the input signal three samples to the right as shown in Fig. 35-5.

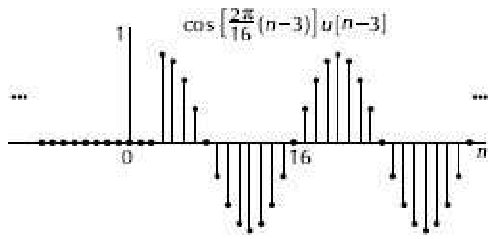


Figure 35-5. The cosine signal from Fig. 1-4 delayed by $n_d = 3$ samples. This delay shifts the sequence to the right by three samples.

The moving average system takes an average of the input signal over some window and then moves to the next sample and takes an average over the new window, etc. The general moving average system is defined by the equation below, where M_1 and M_2 are positive integers. It is called a moving average because to compute each output, $y[n]$, the filter must be moved to the next index and the average recomputed.

$$y[n] = \frac{1}{M_1 + M_2 + 1} \sum_{k=-M_1}^{M_2} x[n-k] \qquad (35\text{-}6)$$

The average sums values together starting M_1 samples forward from the current point and moving M_2 samples back from the current point and divides by the number of points that were summed together to form an average that smooths out the signal. The moving average is a digital filter that removes high frequency information through averaging.

35.3.3 System Properties

System properties are a convenient way to describe broad classes of systems. Important system properties include linearity, shift invariance, causality, and stability. These properties are important because they lead to a representation of systems that can be readily analyzed and implemented.

35.3.3.1 Linearity

A linear system is one where the output of a sum of linearly scaled input signals is equal to the sum of the linearly scaled output signals. Mathematically, a system, $T\{\cdot\}$, is linear when

$$y_1[n] = T\{x_1[n]\}$$

and

$$y_2[n] = T\{x_2[n]\}$$

Then

$$\begin{aligned} T\{ax_1[n] + bx_2[n]\} &= T\{ax_1[n]\} + T\{bx_2[n]\} \\ aT\{x_1[n]\} + bT\{x_2[n]\} &= ay_1[n] + by_2[n] \end{aligned} \qquad (35\text{-}7)$$

This means that when the input to a linear system is a sum of signals, the output is the sum of the signals transformed individually.

As an example, consider a system that performs a scalar multiply $y[n] = \alpha x[n]$ (when $\alpha > 1$, $y[n]$ is a louder version of $x[n]$, and when $\alpha < 1$, $y[n]$ is a quieter version of $x[n]$). This system is linear because

$$\begin{aligned} y[n] &= \alpha(ax_1[n] + bx_2[n]) \\ &= (\alpha ax_1[n] + \alpha bx_2[n]) \end{aligned}$$

An example of a nonlinear system would be a compressor/limiter because the output of a compressor/limiter to a sum of signals is generally not equal to the sum of the compressor/limiters applied to the signals individually since the compressor/limiter system is signal-level dependent and the output from signals below a level can be very different from the output when a sum of signals exceeds the threshold.

35.3.3.2 Time Invariance

A time-invariant system is one where a delay in the input signal causes the output to be delayed by the same amount. Mathematically, a system, T{·}, is time invariant if when $y[n] = T\{x[n]\}$ then

$$T\{x[n-N]\} = y[n-N] \tag{35-8}$$

When the input, $x[n]$, to a linear system is delayed, the output, $y[n]$, is delayed correspondingly. There is no absolute time reference associated with the system. The combination of time invariance and linearity makes the design and analysis of a large class of DSP theory and applications much simpler due to the convolution operation and Fourier analysis tools.[1]

35.3.3.3 Causality

A causal system is one where the output of the system at a given time only depends on the present and past values of the input signal. No future data can be required to produce an output signal at the present time in a causal system. In the moving average system of Eq. 35-6, the system is causal only if $M_1 = 0$.

35.3.3.4 Stability

A system is bounded input/bounded output stable if and only if every bounded input sequence produces a bounded output sequence. A sequence is bounded if each value in the sequence is less than infinity. For real applications, stability is critically important because a system would stop operating properly should it ever become unstable.

35.3.4 Linear Time-Invariant Systems

When the linearity property is combined with the time-invariance property to form a linear time-invariant (LTI) system, then the analysis of systems is very straightforward. Because a sequence can be represented as a sum of weighted delayed impulses as shown in Eq. 35-2, and an LTI system response is the sum of the component responses of the sequence components as shown in Eq. 35-7, the response of an LTI system is completely determined from its response to an impulse. Since an input signal can be represented as a collection of delayed and scaled impulses, the response to the full sequence is known. The response of a system to an impulse is commonly referred to as the *impulse response* of the system. Mathematically

$$x[n] = \sum_k x[k]\delta[n-k]$$

i.e., the sequence $x[n]$ is a sum of scaled and delayed impulses. If $h_k[n] = T\{\delta[n-k]\}$, i.e., the system response to the delayed impulse at $n = k$, then the output $y[n]$ can be formed as

$$\begin{aligned} y[n] &= T\{x[n]\} \\ &= T\left\{\sum_k x[k]\delta[n-k]\right\} \\ &= \sum_k x[k]h_k[n] \end{aligned} \tag{35-9}$$

If the system is also time invariant, then $h_k[n] = h[n-k]$, and the output $y[n]$ is given by

$$\begin{aligned} y[n] &= \sum_k x[k]h[n-k] \\ &= \sum_k h[k]x[n-k] \end{aligned} \tag{35-10}$$

This representation is known as the convolution sum and is commonly written as $y[n] = x[n] * h[n]$. The convolution system takes two sequences, $x[n]$ and $h[n]$, and produces a third sequence $y[n]$. For each value of $y[n]$, the computation requires multiplying $x[k]$ by $h[n-k]$ and summing over all valid indices for k where the signals are non-zero. To compute the output $y[n+1]$, move to the next point, n + 1, and perform the same computation. The convolution is an LTI system and is a building block for many larger systems.

As an example, consider the convolution of the sequences in Fig. 35-6 where $h[n]$ has only three non-zero sample values and $x[n]$ is a cosine sequence that has non-zero sample values for $n \geq 0$.

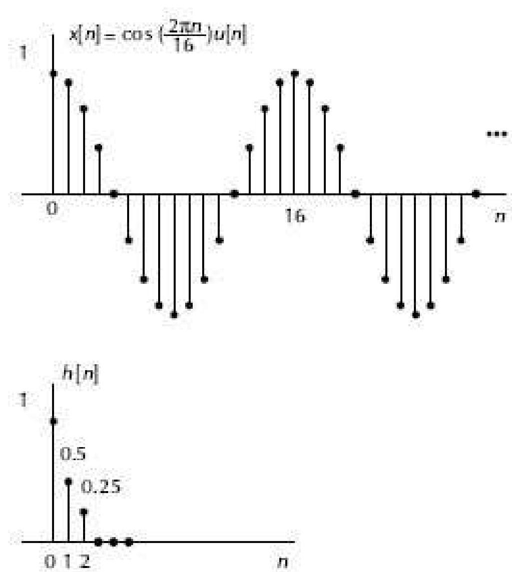


Figure 35-6. A convolution example with two sequences. $x[n]$ is the same signal from Fig. 35-4 with values shown in Table 35-1, and $h[n]$ has the values shown above.

The computation of

$$y[n] = \sum_{k=0}^{2} h[k]x[n-k]$$

is performed as follows. Values of $x[n]$ for $n < 0$ are 0. Only the computation for the first three output samples are shown.

$$
\begin{aligned}
y[0] &= h[0]x[0] + h[1]x[-1] + h[2]x[-2] \\
&= 1.0 \\
y[1] &= h[0]x[1] + h[1]x[0] + h[2]x[-1] \\
&= 1.4239 \\
y[2] &= h[0]x[2] + h[1]x[1] + h[2]x[0] \\
&= 1.4190
\end{aligned}
$$

The result of the convolution is shown in Fig. 35-7 and has the sample values shown in Table 35-2.

Digital filtering with the convolution algorithm described above is a core element of most DSP systems and, for audio products, provides the basis for different styles of audio equalization including parametric equalizers, graphic equalizers, and high-pass, low-pass, and other types of filters, to change how a system sounds in a particular acoustic environment.

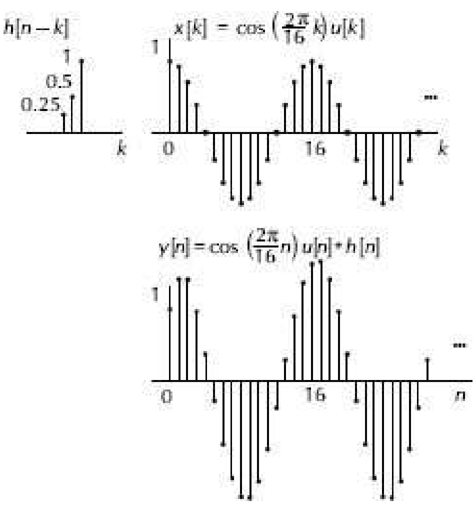


Figure 35-7. The output, $y[n]$, from the convolution of $x[n]$ and $h[n]$ in Fig. 35-6.

Table 35-2 The Result of the Convolution in Fig. 35-7

$y[0]$	1.0000	$y[11]$	−0.9672	$y[22]$	−0.8984
$y[1]$	1.4239	$y[12]$	−0.3681	$y[23]$	−1.3731
$y[2]$	1.4190	$y[13]$	0.2870	$y[24]$	−1.6387
$y[3]$	0.9672	$y[14]$	0.8984	$y[25]$	−1.6548
$y[4]$	0.3681	$y[15]$	1.3731	$y[26]$	−1.4190
$y[5]$	−0.2870	$y[16]$	1.6387	$y[27]$	−0.9672
$y[6]$	−0.8984	$y[17]$	1.6548	$y[28]$	−0.3681
$y[7]$	−1.3731	$y[18]$	1.4190	$y[29]$	0.2870
$y[8]$	−1.6387	$y[19]$	0.9672	$y[30]$	0.8984
$y[9]$	−1.6548	$y[20]$	0.3681	...	
$y[10]$	−1.4190	$y[21]$	−0.2870		

35.4 Frequency Domain Representation

Having defined an LTI system, it is possible to look at the signal from the frequency domain perspective and understand how a system transforms the signals in the frequency domain. The frequency domain represents signals as a combination of various frequencies from low frequency to high frequency. Each time-domain signal has a representation as a collection of frequency components where each frequency component can be thought of as sinusoids or tones. Sinusoids are important because a sinusoidal input to a linear time-invariant system generates an output of the same frequency but with amplitude and phase determined by the system. This very important property makes the representation of signals in terms of sinusoids very useful.

As an example, assume an input signal $x[n]$ is defined as $x[n] = e^{j\omega n}$, i.e., a complex exponential (Euler's relationship from complex number theory that states that $e^{j\omega m} = \cos(\omega n) + j\sin(\omega n)$, where ω is the radian frequency that ranges from $0 \le \omega \le 2\pi$), then using the convolution sum of

$$y[n] = \sum_k h[k]x[n-k]$$

generates

$$y[n] = \sum_k h[k] e^{j\omega(n-k)} \qquad (35\text{-}11)$$

$$y[n] = e^{j\omega n}\left(\sum_k h[k] e^{-j\omega k}\right) \qquad (35\text{-}12)$$

By defining

$$H(e^{j\omega}) = \sum_k h[k] e^{-j\omega k}$$

we have

$$y[n] = H(e^{j\omega}) e^{j\omega n}$$

where,

$H(e^{j\omega})$ represents the phase and amplitude determined by the system.

This shows that a sinusoidal (or, in this case, the complex exponential) input to a linear time invariant system will generate an output that has the same frequency but with an amplitude and phase determined by the system.

$H(e^{j\omega})$ is known as the frequency response of the system and describes how the LTI system will modify the frequency components of an input signal. The transformation

$$H(e^{j\omega}) = \sum_k h[k] e^{-j\omega k}$$

is known as the Fourier transform of the impulse response, $h[n]$. If

H($e^{j\omega}$) is a low-pass filter, then it has a frequency response that attenuates high frequencies but not low frequencies—hence it passes low frequencies. If $H(e^{j\omega})$ is a high-pass filter, then it has a frequency response that attenuates low frequencies but not high frequencies.

In many instances it is more useful to process a signal or analyze a signal from the frequency domain than in the time domain either because the phenomenon of interest is frequency based or our perception of the phenomenon is frequency based. Therefore the frequency domain is an important tool in signal processing.

An example of using the frequency domain is the family of MPEG audio compression standards that exploits the frequency properties of the human auditory system to dramatically reduce the number of bits required to represent the signal without significantly reducing the audio quality.

35.5 The Z-Transform

The Z-transform is a generalization of the Fourier transform that permits the analysis of a larger class of systems than the Fourier transform. In addition, the analysis of systems is easier due to the convenient notation of the Z-transform.[1] The Fourier transform is defined as

$$X(e^{j\omega}) = \sum_k x[k]e^{-j\omega k}$$

while the Z-transform is defined as

$$X(z) = \sum_k x[k]z^{-k}$$

When working with linear time invariant systems, an important relationship is that the Z-transform of the convolution of two sequences is equal to the multiplication of the Z-transforms of the two sequences, i.e., $y[n] = x[n]*h[n] \Leftrightarrow Y(z) = X(z)H(z)$. $H(z)$ is referred to as the system function (a generalization of the transfer function from Fourier analysis).

A common use of the Z domain representation is to analyze a class of systems that are defined as linear constant-coefficient difference equations that have the form of

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{k=0}^{M} b_k x[n-k] \qquad (35\text{-}13)$$

where,

the coefficients a_k and b_k are constant (hence the name constant coefficient).

This general difference equation forms the basis for both finite impulse response (FIR) linear filters, and infinite impulse response (IIR) linear filters. Both FIR and IIR filters are used to implement frequency selective filters (e.g., high-pass, low-pass, bandpass, bandstop, and parametric filters) and other more complicated systems.

FIR filters are a special case of Eq. 35-13, where except for the first coefficient, all the a_k are set to 0, leading to the equation

$$y[n] = \sum_{k=0}^{M} b_k x[n-k] \qquad (35\text{-}14)$$

The important fact to notice is that each output sample $y[n]$ in the FIR filter is formed by multiplying the sequence of coefficients (also known as *filter taps*) by the input sequence values. There is no feedback in an FIR filter—i.e., previous output values are not used to compute new output values. A block diagram of this is shown in Fig. 35-8 where the z^{-1} blocks are used to denote a signal delay of one sample (i.e., the Z-transform of the system $h[n] = \delta[n - 1]$).

An IIR filter contains feedback in the computation of the output $y[n]$—i.e., previous output values are used to create current output values. Because of this feedback, IIR filters can be created that have a better frequency response (i.e., steeper slope for attenuating signals outside the band of interest) than FIR filters for a given amount of computation. However, many DSP architectures are optimized for computing FIR filters—i.e., multiplying and adding signals together continuously—so the choice of which filter style to use will depend on the particular application.

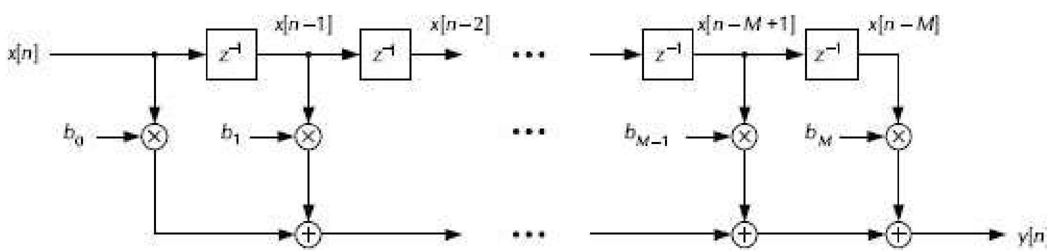


Figure 35-8. A block diagram of an FIR system where the input $x[n]$ is fed into a system that multiplies the delayed input signal with the filter coefficients *bk* and sums the results together to form the output $y[n]$.

35.6 Sampling of Continuous-Time Signals

Understanding digital signal processing fundamentals is the first step. The next step is to create digital sequences from the analog signals around us. The most common way to generate a digital sequence is to start with a continuous-time (analog) signal and create a discrete-time signal. For example, speech signals are continuous-time signals because they are continuous waves of acoustic pressure. A microphone is the transducer that converts the acoustic signal into a continuous-time electric signal. In order to process this signal digitally, it is necessary to convert this signal into the digital domain. Finally, after processing, it is often necessary to convert the discrete-time signal back into a continuous-time signal for playback through a loudspeaker system.

The process of converting an analog signal to a digital signal is often be modeled as a two-step process, as shown in Fig. 35-9, of converting a continuous-time signal to a discrete-time signal (with infinite resolution of the amplitude) and then quantizing the discrete-time signal into finite precision values (creating the digital sequence) that can be processed by a computer.[1] The process of converting the continuous-time signal into a discrete-time signal will be introduced, and then quantization will be reviewed. The quantization step is necessary to create a sample value that has a data word size that is compatible with the arithmetic capabilities of the target DSP, for example 16-bit words, 32-bit words, etc. All real-world analog-to-digital converters (A/Ds) perform both the sampling and quantization process internal to the A/D device, but it is useful to discuss the subsystems separately, not only because it's easier to understand this way, but also because they have different significance and design trade-offs.

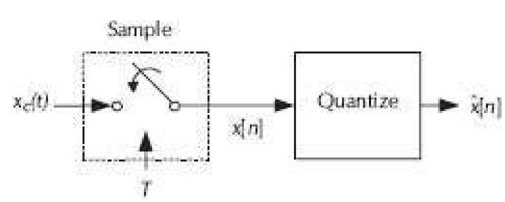


Figure 35-9. Analog-to-digital conversion can be thought of as a two-step process: converting a continuous-time signal to a discrete-time signal, $x[n]$, followed by quantizing the sample to create the digital sequence.

35.6.1 Continuous to Discrete Conversion

The most common method for converting a continuous-time signal, $x_c(t)$, into a discrete-time signal, $x[n]$, is to uniformly sample the signal every T seconds with the equation

$$x[n] = x_c(nT), -\infty < n < \infty \qquad (35\text{-}15)$$

This generates a sequence of samples, $x[n]$, where the value of $x[n]$ is the same as the value of $x_c(t)$ whenever $t = nT$—i.e., at each sampling interval T. $1/T$ is known as the *sampling frequency* and is usually expressed in Hertz or cycles per second.

Mathematically, when a continuous-time signal is sampled, the resulting signal has a frequency response that is related to the underlying continuous-time signal frequency response and the sampling rate. As shown next, this has significant ramifications for how often the signal must be sampled in order for the digital sequence to be reconstructed into an analog signal that accurately represents the original signal.

The sampling process will be analyzed in the frequency domain where it will be assumed that a band limited signal, $x_c(t)$, is to be sampled periodically with sample period T. A band-limited signal is one that has no signal energy higher than a particular frequency, Ω_N, as shown in Fig. 35-10, where Ω represents the frequency axis of the signal. The reason the signal is assumed to be band limited is to prevent frequency aliasing, as will be evident shortly. The assumption of being band limited is significant although generally easily realizable in real-world systems.

The sampling of the continuous-time signal, $x_c(t)$, generates a signal, $x_s(t)$, from equation

$$x_s(t) = \sum_{n=-\infty}^{\infty} x_c(nT)\delta(t-nT) \qquad (35\text{-}16)$$

$x_s(t)$ is the collection of values of $x_c(t)$ at the sampling interval of T. A convenient representation of this signal is as a collection of delayed and weighted impulse functions. The amplitude is the value at the sampling instant and the samples are spaced out by the sampling period T. The process can be analyzed in the frequency domain by first representing the Fourier transform of the impulse sequence as a sequence of impulses in the frequency domain.[2] This means that a sequence of equally spaced impulses in the time domain have a frequency representation that is a sequence of equally spaced impulses in the frequency domain, spaced by the sampling frequency $2\pi/T$. This is shown as

$$S(j\Omega) = \frac{2\pi}{T}\sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s) \qquad (35\text{-}17)$$

where,
$\Omega_s = 2\pi/T$ is the sampling frequency in radians/second.

The Fourier transform of the sampled signal, $x_s(t)$, becomes

$$X_s(j\Omega) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_c(j(\Omega - k\Omega_s)) \qquad (35\text{-}18)$$

Now the frequency response of the sampled continuous-time signal becomes a collection of shifted copies of the original frequency response of the analog signal $X_c(j\Omega)$. Fig. 35-10 shows the frequency response of $X_c(j\Omega)$, the impulse train, $S(j\Omega)$, and the resulting frequency response of the sampled signal, $X_s(j\Omega)$.

This frequency response, $X_S(j\Omega)$, can also be interpreted as the convolution in the frequency domain between the frequency response of the continuous-time signal and the frequency response of the impulse train, $S(j\Omega)$.

$$X_s(j\Omega) = \frac{1}{2\pi} X_c(j\Omega) * S(j\Omega) \qquad (35\text{-}19)$$

From Fig. 35-10 it can be seen that as long as the sampling frequency minus the highest frequency is greater than the highest frequency, $\Omega_S - \Omega_N > \Omega_N$ the frequency copies do not overlap. This condition can be rewritten as $\Omega_S > 2\Omega_N$, which means that the sampling frequency must be at least twice as high as the highest frequency in the signal. If the sampling frequency is less than the highest frequency in the signal, $\Omega_S < 2\Omega_N$, then the frequency copies overlap as shown in Fig. 35-11. This overlap causes the frequencies of the adjacent spectral copies to be added together, which results in the loss of spectral information meaning that the sampled signal is no longer an accurate representation of the original analog signal. It is impossible to remove the effects of aliasing once aliasing has happened. The overlap in the frequency domain is caused because the sampling frequency, Ω_S, is not high enough relative to the highest frequency in the continuous-time signal $X_c(j\Omega)$. As shown above, the sampling frequency must be at least twice as high as the highest frequency in the continuous-time signal in order to prevent this overlap, or aliasing, of frequencies.

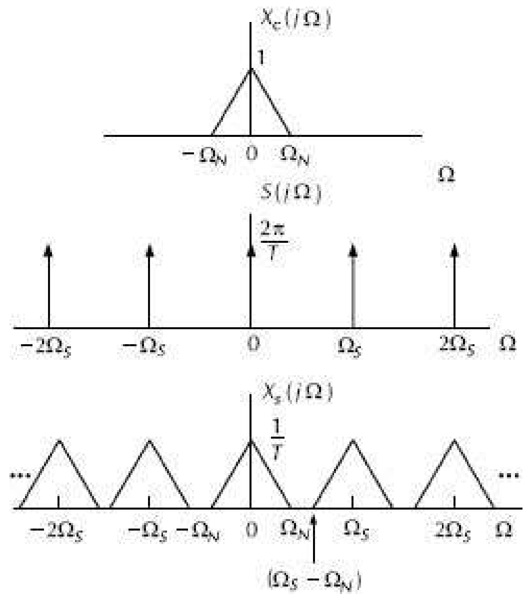


Figure 35-10. The frequency response of the analog signal, $X_c(j\Omega)$, the sampling function, $S(j\Omega)$, and the resulting frequency response of the sampled signal, $X_S(j\Omega)$.

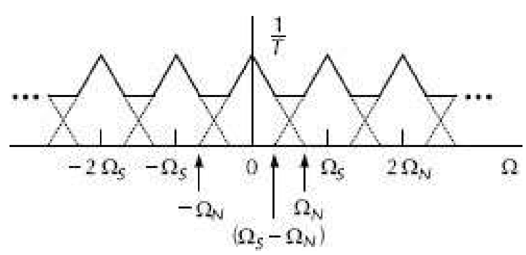


Figure 35-11. Sampling where the sampling frequency, Ω_S, is less than twice the highest frequency, Ω_N.

35.6.2 Reconstructing the Continuous-Time Signal

As seen from sampling a continuous-time signal, if the signal is not sampled fast enough, then the resulting frequency response of the sampled signal will have overlapping copies of the frequency response of the original signal. Assuming the signal is sampled fast enough (at least twice the bandwidth of the signal), the continuous-time signal can be reproduced by simply removing all of the spectral copies except for the desired one. This frequency separation can be performed with an ideal low-pass filter with gain, T, and cut-off frequency, Ω_C, where the cut-off frequency is higher than the highest frequency in the signal as well as the frequency where the first frequency replica starts, i.e., $\Omega_N < \Omega_C < \Omega_S - \Omega_N$. Fig. 35-12 shows the repeated frequency spectrum and the ideal low-pass filter. Fig. 35-13 shows the result of applying the low-pass filter to $X_S(j\Omega)$.

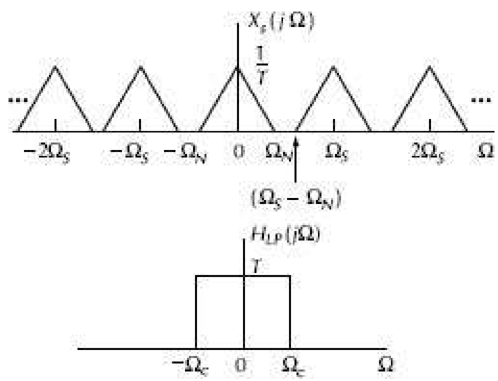


Figure 35-12. The spectrum replicas and the ideal low-pass filter that will remove the copies except for the desired baseband spectrum.

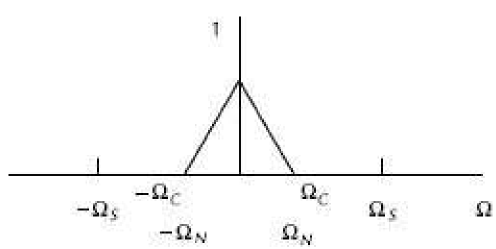


Figure 35-13. The final result of reconstructing the analog signal from the sampled signal.

35.6.3 Sampling Theory

The requirements for sampling are summarized by the Nyquist sampling theorem.[1] Let $x_c(t)$ be a band-limited signal with $X_c(j\Omega) = 0$ for $|\Omega| > \Omega_N$. Then $x_c(t)$ is uniquely determined by its samples, $x[n] = x_c(nT)$, if $\Omega_S = 2\pi/T \geq 2\Omega_N$. The frequency Ω_N is referred to as the *Nyquist frequency*, and the frequency $2\Omega_N$ is referred to as the *Nyquist rate*. This theory is significant because it states that as long as a continuous-time signal is band-limited and sampled at least twice as fast as the highest frequency, then it can be exactly reproduced by the sampled sequence.

The sampling analysis can be extended to the frequency response of the discrete time sequence, $x[n]$, by using the relationships $x[n] = x_c(nT)$ and

$$X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} x[n]e^{-j\omega n}$$

The result is that

$$X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_c\left(j\left(\frac{\omega}{T} - \frac{2\pi k}{T}\right)\right) \qquad (35\text{-}20)$$

$X(e^{j\omega})$ is a frequency-scaled version of the continuous-time frequency response, $X_s(j\Omega)$, with the frequency scale specified by $\omega = \Omega T$. This scaling can also be thought of as normalizing the frequency axis by the sample rate so that frequency components that occurred at the sample rate now occur at 2π. Because the time axis has been normalized by the sampling period T, the frequency axis can be thought of as being normalized by the sampling rate $1/T$

35.6.4 Quantization

The discussion up to this point has been on how to quantify the effects of periodically sampling a continuous-time signal to create a discrete-time version of the signal. As shown in Fig. 35-9, there is a second step—namely, mapping the infinite-resolution discrete-time signal into a finite precision representation (i.e., with a finite number of bits per sample) that can be manipulated in a computer. This second step is known as *quantization*. The quantization process takes the sample from the continuous-to-discrete conversion and finds the closest corresponding finite precision value and represents this level with a bit pattern. This bit pattern code for the sample value is usually a binary two's-complement code so that the sample can be used directly in arithmetic operations without the need to convert to another numerical format (which takes some number of instructions on a DSP processor to perform). In essence, the continuous-time signal must be both quantized in time (i.e., sampled), and then quantized in amplitude before it can be processed by a DSP.

The quantization process is denoted mathematically as

$$x[n] = Q(x[n])$$

where,
$Q(\cdot)$ is the nonlinear quantization operation,
$x[n]$ is the infinite precision sample value.

Quantization is nonlinear because it does not satisfy Eq. 35-7, i.e., the quantization of the sum of two values is not the same as the sum of the quantized values due to how the nearest finite precision value is generated for the infinite-precision value.

To properly quantize a signal, it is required to know the expected range of the signal, i.e., the maximum and minimum signal values. Assuming the signal amplitude is symmetric, the most positive value can be denoted as X_M. The signal then ranges from $+X_M$ to $-X_M$ for a total range of $2X_M$. Quantizing the signal to B bits will decompose the signal into 2^B different values. Each value represents $2X_M/2^B$ in amplitude and is represented as the step size $\Delta = 2X_M 2^{-B} = X_M 2^{-(B-1)}$. As a simplified example of the quantization process, assume that a signal will be quantized into eight different values that can be conveniently represented as a 3-bit value. Fig. 35-14 shows one method of how an input signal, $x[n]$, can be converted into a 3-bit quantized value, $Q(x[n])$. In this figure, values of the input signal between $-\Delta/2$ and $\Delta/2$ are given the value 0. Input signal values between $\Delta/2$ and $3/2$ are represented by their average value Δ, and so forth. The eight output values range from -4Δ to 3Δ for input signals between $-9\Delta/2$ and $7\Delta/2$. Values larger than $7\Delta/2$ are set to 3Δ and values smaller than $-9\Delta/2$ are set to -4Δ—i.e., the numbers saturate (also known as clipping) at the maximum and minimum values, respectively.

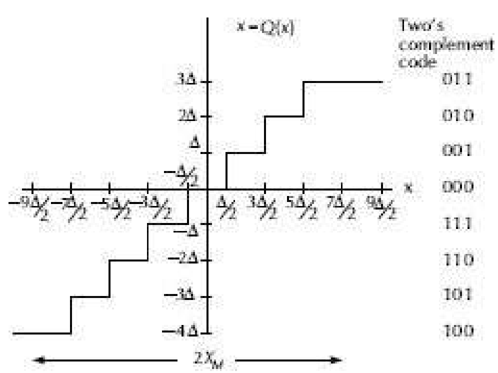


Figure 35-14. The quantization of an input signal, x, into Q(x).

The step size, Δ, has an impact on the resulting quality of the quantization. If Δ is large, fewer bits will be required for each sample to represent the range, $2X_M$, but there will be more quantization errors. If Δ is small, more bits will be required for each sample, although there will be less quantization error. Normally, the system design process determines the value of X_M and the number of bits required in the converter, B. If X_M is chosen too large, then the step size, Δ, will be large and the resulting quantization error will be large. If X_M is chosen too small, then the step size, Δ, will be small, but the signal may clip the A/D converter if the actual range of the signal is larger than X_M.

This loss of information during quantization can be modeled as noise signal that is added to the actual signal as shown in Fig. 35-15. The amount of quantization noise determines the overall quality of the signal. In the audio realm, it is common to sample with 24 bits of resolution on the A/D converter. Assuming a ±15V swing of an analog signal, the granularity of the digitized signal is $30V/2^{24}$, which comes to 1.78µV.

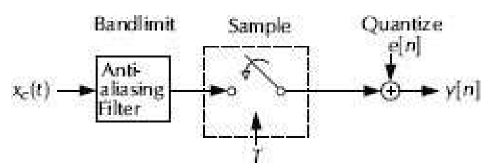


Figure 35-15. The sampling process of Fig. 35-9 with the addition of an antialiasing filter and modeling the quantization process as an additive noise signal.

With certain assumptions about the signal, such as the peak value being about four times the rms signal value, it can be shown that the signal to noise ratio (*SNR*) of the A/D converter is approximately 6dB per bit.[1] Each additional bit in the A/D converter will contribute 6dB to the *SNR*. A large *SNR* is usually desirable, but that must be balanced with overall system requirements, system cost, and possibly other noise issues inherent in a design that would reduce the effectiveness of a high-quality A/D converter in the system. The dynamic range of a signal can be defined as the range of the signal levels over which the *SNR* exceeds a minimum acceptable *SNR*.

There are cost-effective A/D converters that can shape the quantization noise and produce a high-quality signal. Sigma-Delta converters, or noise-shaping converters, use an oversampling technique to reduce the amount of quantization noise in the signal by spreading the fixed quantization noise over a bandwidth much larger than the signal band.[3] The technique of oversampling and noise shaping allows the use of relatively imprecise analog circuits to perform high-resolution conversion. Most digital audio products use these types of converters.

35.6.5 Sample Rate Selection

The sampling rate, *1/T*, plays an important role in determining the bandwidth of the digitized signal. If the analog signal is not sampled often enough, then high-frequency information will be lost. At the other extreme, if the signal is sampled too often, there may be more information than is needed for the application, causing unnecessary computation and adding unnecessary expense to the system.

In audio applications it is common to have a sampling frequency of 48kHz = 48,000Hz, which yields a sampling period of 1/48,000 = 20.83µs. Using a sample rate of 48kHz is why, in many product data sheets, the amount of delay that can be added to a signal is an integer multiple of 20.83µs.

The choice of which sample rate to use depends on the application and the desired system cost. High-quality audio processing would require a high sample rate while low bandwidth telephony applications require a much lower sample rate. A table of common applications and their sample rate and bandwidths are shown in Table 35-3. As shown in the sampling process, the maximum bandwidth will always be less than one-half the sampling frequency. In practice, the antialiasing filter will have some roll-off and will band limit the signal to less than one-half the sample rate. This bandlimiting will further reduce the bandwidth, so the final bandwidth of the audio signal will be a function of the filters implemented in the specific A/D and the sample rate of the system.

Table 35-3 Common Sample Rates Found in Typical Applications and the Practical Bandwidths Realized at Each Sample Rate

Application	Sample Rate	Bandwidth
Telephony applications	8 kHz	3.5kHz
VoIP telephony	16kHz	7kHz
Videoconferencing	16kHz	7kHz
FM radio	32 kHz	15kHz
CD audio	44.1kHz	20kHz
Professional audio	48 kHz	22 kHz
Future audio	96kHz	45 kHz

Recent advances in audio compression technology and algorithms coupled with the availability of cost-effective high-performance DSP's have increased the bandwidth typically used for both audio conferencing and video conferencing to 7kHz and 22kHz, respectively. The wide availability of Voice over IP (VoIP) solutions has greatly improved the audio quality of business telephony solutions. These VoIP solutions support a variety of audio encoders and decoders that are implemented on DSP devices.

35.7 Algorithm Development

Once a signal is digitized, the next step in a DSP system is to process the signal. The system designer will begin the design process with some goal in mind and will use the algorithm development phase to develop the necessary steps (i.e., the algorithm) for achieving the goal.

The design cycle for a DSP system generally has three distinct phases as shown in Fig. 35-16: an abstract algorithm conceptualization phase, in which various mathematical algorithms and systems are explored; an algorithm development phase, where the algorithms are tested on large amounts of data; and a system implementation phase, where specific hardware is used to realize the system.

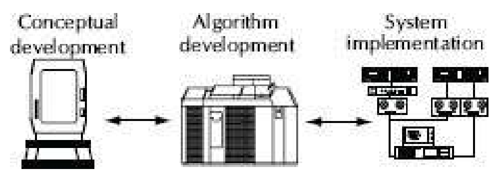


Figure 35-16. The three phases of DSP application development.

Traditionally, the three phases in the DSP design cycle have been performed by three entirely different groups of engineers using three entirely different classes of tools, although this process has converged as development tools have improved. The algorithm conceptualization phase is most often performed by researchers in a laboratory environment using highly interactive, graphically oriented DSP simulation and analysis tools. In this phase, the researcher begins with the concept of what to accomplish and creates the simulation environment that will enable changes and reformulations of the approach to the problem. No consideration is given, at this point, to the computational performance issues. The focus is on proof of concept issues—proving that the approach can solve the problem (or a temporarily simplified version of the problem).

In the algorithm development phase, the algorithms are fine-tuned by applying them to large data sets of signals, often using high-speed computers or workstations to achieve the required throughput. During this step it is often necessary to refine the high-level conceptualization in order to address issues that arose while running data through the system. Simulations are characterized by having many probes on the algorithm to show intermediate signal values, states, and any other useful information to aid in troubleshooting both the algorithm and the implementation of the simulation.

Once the simulation performs as desired, the next step is to create a real-time implementation of the simulation. The purpose of the real-time implementation is to better simulate the final target product, to begin to understand what the real-time memory and computational requirements will be, and to run real-time data through the system. There is no substitute for running real-time data through the system because real-time data typically exhibits characteristics that were either not anticipated or have unintended consequences in the simulation environment, for example background noise, reverberation, male and female talkers, different languages, or signal level variation may be significantly different from the initial assumptions while in the algorithm development phase. Because of this real-time data is generally more stressful to an algorithm than simulated, or non-real-time, data.

Often, with the introduction of real-time data, it may be necessary to go to the conceptual level again and further refine the algorithm.

Although advanced development tools and high-speed processors have blurred the distinction between simulation and real-time implementation, the goal of the real-time implementation is to "squeeze as much algorithm as possible" into the target processor (or processors). Squeezing more into the target processor is a desirable goal because it is usually much less expensive to use a single signal processor than to use multiple processors.

35.8 Digital Signal Processors

Programmable digital signal processors are microprocessors with particular features suited to performing arithmetic operations such as multiplication and addition very efficiently.[4], [5] While traditionally, these enhancements have improved the performance of the processor at the expense of ease of programmability, development tools have improved significantly and higher-level languages, using optimized libraries to fully take advantage of the DSP processing capabilities, are commonly used.

A typical microprocessor will have an arithmetic and logic unit for performing arithmetic operations, a memory space, I/O pins, and possible other peripherals such as serial ports and timers. A DSP processor will often have fewer peripherals, but will include a hardware multiplier, often a high-speed internal memory space, more memory addressing modes, an instruction cache and a pipeline, and even a separation of the program and data memory spaces to help speed program execution. The hardware multiplier allows the DSP processor to perform a multiplication in a single clock cycle while microprocessors typically take multiple clock cycles to perform this task. With clock cycles easily exceeding 400MHz, up to 400 million multiples can occur every second. At this rate, more than 8000 multiplies can occur in the time span required to collect one sample of data at a 48kHz sample rate (400M/48,000).

A high-speed internal memory bank can be used to speed the access to the data and/or program memory space. By making the memory high speed, the memory can be accessed twice within a single clock cycle, allowing the processor to run at maximum performance. This means that proper use of internal memory enables more processing to take place within a given speed processor when compared to using external memory.

The instruction cache is also used to keep the processor running more efficiently because it stores recently used instructions in a special place in the processor where they can be accessed quickly, such as when looping program instructions over signal data.

The pipeline is a sequential set of steps that allow the processor to fetch an instruction from memory, decode the instruction, and execute the instruction. By running these subsystems in parallel, it is possible for the processor to be executing one instruction while it is decoding the next one and fetching the instruction after that. This streamlines the execution of instructions.

35.8.1 DSP Arithmetic

Programmable DSPs offer either fixed-point or floating-point arithmetic. Although floating-point processors are typically more expensive and offer less performance than fixed-point processors, VLSI hardware advances are minimizing the differences although there are typically some costs advantages to the fixed-point processors due to the lower complexity, smaller amount of silicon required, and typically higher quantities in which they are used. The main advantage of a floating-point processor is the ability to be free of numerical scaling issues, simplifying the algorithm development and implementation process.

When most people think of floating-point numbers, they think in terms of fractions and decimal points. Typically, floating-point DSPs can represent very large and very small numbers and use 32-bit (or longer) words composed of a 24-bit mantissa and an 8-bit exponent, which together provide a dynamic range from 2^{-127} to 2^{128}. This vast range in floating-point devices means that the system developer does not need to spend much time worrying about numerical issues such as overflow (a number too large to be represented) or underflow (a number too small to be represented). In a complicated system, there is enough complexity already without having to worry about numerical issues as well.

Fixed-point arithmetic is called fixed-point because it has a fixed decimal point position and because the numbers have an implicit scale, depending on the range that must be represented. This scale must be tracked by the programmer when performing arithmetic on fixed-point numbers. Most DSPs use the fixed-point 2s-complement format, in which a positive number is represented as a simple binary value and a negative value is represented by inverting all the bits of the corresponding positive value and then adding 1. Assuming a 16-bit word, there are $2^{16} = 65{,}536$ possible combinations or values that can be represented which allows the representation of numbers ranging from the largest positive number of $2^{15} - 1 = 32{,}767$ to the smallest negative (e.g., most negative) number of $-2^{15} = -32{,}768$.

There are many times when it is important to represent fractions in addition to integer numbers. To represent fractions, the implied position of the decimal point must be moved. When using 16-bit arithmetic to represent fractions only, with no integer component, a Q15 arithmetic format with an implied decimal point and 15 bits of fraction data to the right of the decimal point could be used. In this case, the largest number that can be represented is still $2^{15} - 1$, but now this number represents 32,767/32,768 = 0.999969482, and the smallest negative number is still -2^{15}, but this number represents −32,768/32,768 = −1. Using Q15 arithmetic, it is possible to represent numbers between 0.999969482 and −1. As another example, representing numbers that range between 16 and −16 would require Q11 arithmetic (4 bits before the implied decimal point). An implementation may use different implied decimal positions for different variables in a system and it is the task of the programmer to keep track of it all.

Because of the smaller word size and simpler arithmetic operations when compared to floating-point processors, fixed-point DSPs typically use less silicon area than their floating-point counterparts, which translates into lower prices and less power consumption. The trade-off is that, due to the limited dynamic range and the rules of fixed-point arithmetic, an algorithm designer must play a more active role in the development of a fixed-point DSP system. The designer has to decide whether the given word width (typically 16 or 24 bits) will be interpreted as integers or fractions, apply scale factors if required, and protect against possible register overflows at potentially many different places in the code. Overflow occurs in two ways in a fixed-point DSP.[4] Either a register overflows when too many numbers are added to it or the program attempts to store N bits from the accumulator and the discarded bits are important. A complete solution to the overflow problem requires the system designer to be aware of the scaling of all the variables so that overflow is sufficiently unlikely. An underflow occurs if a number is smaller than the smallest number that can be represented. Floating-point arithmetic keeps track of the scaling automatically simplifying the programmer's job. The exponent keeps track of where the decimal point should be. Checking for overflow/underflow with fixed-point architectures and preventing these conditions makes changing a DSP algorithm more difficult because, not only are algorithmic changes required, there are also numeric issues to contend with. Usually, once an implementation for a particular application has matured past the development stage, the code (which may have begun as floating-point code) may be ported to a fixed-point processor to allow the cost of the product to be reduced. The decision factors for porting floating-point code to a fixed-point processor include the price and anticipated volumes of the resulting product and also the cost of the development and maintenance effort to port a fixed-point system.

The dynamic range supported in a fixed-point processor is a function of the number of bits in the processor's data registers. As with A/D conversion, each bit adds 6dB to the *SNR*. A 24-bit DSP has 48dB more dynamic range than a 16-bit DSP.

35.8.2 Implementation Issues

The implementation of an algorithm into a real system is often much more complicated than using a high-level language compiler to automatically optimize the code for maximum performance. Real-time systems have constraints such as limited memory, limited computational performance, and most importantly, need to handle the real-time data that is continuously sent from the A/D converter to the DSP and the real-time data that must be sent from the DSP back to the D/A converter. Interruptions in this real-time data are typically not acceptable because, for example, in an audio application, these interruptions will cause audible pops and clicks in the audio signal.

Real-time programming requires that all of the computation required to produce the output signal must happen within the amount of time it takes to acquire the input signal from the A/D converter. In other words, each time an input sample is acquired, an output sample must be produced. If the processing takes too long to produce the output, then, at some point, incoming data from the A/D will not be able to be processed, and input samples will be lost. As an example, assume a system samples at 48kHz and performs parametric equalization on a signal. Assuming that each band of parametric equalization requires 5 multiplies and 4 adds, which can be implemented in 9 clock cycles, then a 400MHz DSP has 8000 instructions that can be executed in the time between samples. These instructions would allow a maximum of 888 bands of parametric equalization (8000/9 = 888) which may be applied to multiple input within an audio processing device. Now, realistically, the system is performing other tasks such as collecting data from the A/D converter, sending data to the D/A converter, handling overhead from calling subroutines and returning from subroutines, and is possibly responding to interrupts from other subsystems. So the actual number of bands of equalization could be significantly less than the theoretical maximum of 888 bands.

DSPs will have a fixed amount of internal memory and a fixed amount of external memory that can addressed. Depending on the system to be designed, it can be advantageous to minimize the amount of external memory that is required in a system because that can lead to reduced parts costs, reduced manufacturing expense, and higher reliability. However, there is usually a trade-off between computational requirements and memory usage. Often, it is possible to trade memory space for increased computational power and vice versa. A simple example of this would be the creation of a sine wave. The DSP can either compute the samples of a sine wave, or look-up the values in a table. Either method will produce the appropriate sine wave, but the former will require less memory and more CPU while the latter will require more memory and less CPU. The system designer usually makes a conscious decision regarding which trade-off is more important.

35.8.3 System Delay

Depending on the application, one of the most important issues in an implementation is the amount of delay or latency that is introduced into the system by the sampling and processing. Fig. 35-17 shows the typical digital system. The analog signal comes into the A/D converter that digitizes and quantizes the signal. Once digitized, the signal is typically stored in some data buffers or arrays of data. The data buffers could be one sample long or could be longer depending on whether the algorithm operates on a sample-by-sample basis or requires a buffer of data to perform its processing. The system buffers are usually configured in a ping-pong fashion so that while one buffer is filling up with new data from the A/D, the other is being emptied by the DSP as it pulls data from the buffer to process the data.

Following the system buffer may be a data conversion block that converts the data from a fixed-point integer format provided by the A/D to either some other fixed-point format or a floating-point processor, depending on the DSP and the numerical issues. Following this, there may be some application buffers that store buffers of data to give the DSP some flexibility in how much time it takes to process a single block of data. The application buffers can be viewed as a rubber band that allows the DSP to use more time for some frames of data and less time for other frames of data. As long as the average amount of time required to process a buffer of data is less than the amount of time required to acquire that buffer of data, the DSP will make real-time. If the amount of time required to process a buffer takes longer than the time to acquire the buffer, then the system will be unable to process all buffers and will have to drop buffers because there will not be any processing time left over to collect the next buffer from the A/D converter. In this case the system will not make real time and the missing buffers will produce audible pops in an audio signal. The application buffers can be used to compensate for some frames that may require more processing (more CPU time) than others. By providing more frames over which to average the computation, the DSP will more likely make real time. Of course, if the DSP cannot perform the required amount of computation on average during the time that a buffer of data is acquired, then averaging over more and more frames will not help. The system will eventually miss real time and have to drop samples with audible consequences. Having too many data buffers has the downside that the system latency increases. The maximum latency that may be acceptable will depend on the application.

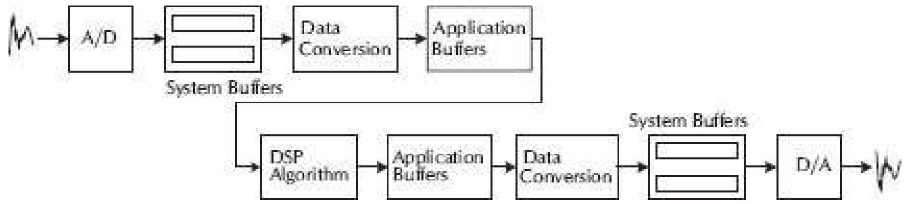


Figure 35-17. A block diagram of the typical DSP system.

After the application buffers, the DSP algorithm performs the operations that are desired and then passes the data to possibly another set of application buffers that in turn can be converted from the numerical format of the DSP to the format required by the D/A converter. Finally the data will be sent to the D/A converter and converted back into an analog signal.

An accounting of the delay of the system should include all delays beginning when the analog signal comes in contact with the A/D converter to when the analog signal leaves the D/A converter. Table 35-4 shows the potential delays in each of the blocks of Fig. 35-17. For this exercise, it is assumed that a frame of data consists of N samples, where $N \geq 1$. Each frame of delay adds $N \cdot 1/T$ seconds of delay to the system. For example, a delay of 16 samples at 48kHz corresponds to 16/48,000 = 333.3μs.

Further complicating the delay measurements is the possible requirement of sending information to an external system. This could be in the form of sending a bitstream to a remote decoder, receiving a bitstream from a remote encoder, and also any error detection and/or correction on a bitstream that may be required.

35.8.4 Choosing a DSP

The choice of which DSP to use for a particular application depends on a collection of factors including:

Cost. DSPs range in price from several dollars to hundreds of dollars. Low-cost DSP processors are typically 16-bit fixed-point devices with limited amounts of internal memory and few peripherals. Low-cost DSPs are typically suited for extremely high volume applications, where the exact capabilities required, and no more, are built into the chip.

High-cost DSPs typically are newer processors that have a great deal of internal memory or other architectural features including floating-point arithmetic, multiple processing cores, and high speed communication ports.

Computational Power:MHz, MIPs, MFLOPs. Computational power is measured in several different ways including processor speed (MHz), millions of instructions per second (MIPS), and millions of floating-point operations per second (MFLOPS). The computational power of a processor is usually directly related to cost. An MIP means that one million instructions can be executed per second. The instructions that can be executed could include memory loads and stores or perhaps arithmetic operations. An MFLOP means one million floating-point operations can be executed per second. A floating-point operation includes multiplies and/or adds.

Table 35-4. A Summary of Delay Issues in a Typical DSP System

Block	Delay	Description
A/D	From 1 to 16 samples	Most A/D converters have some amount of delay built in due to the processing that is done. Oversampling A/Ds in particular have more delay than other types of A/Ds.
System Buffers	Adds at least 1	In the ping-pong buffer scheme, the system is always processing the last frame of data while

	frame of delay	the A/D is supplying the data from the next frame of data.
Data conversion	Possibly none	The conversion of the data format may be lumped with the algorithm processing delay.
Application buffers	Adds M-1 frames of delay for M buffers	Generalizing the ping-pong buffer scheme to M buffers, the system is always processing the oldest buffer, which is M-1 buffers behind the most recent buffer.
DSP algorithm	Variable, although usually at least 1 frame	There are two primary ways a DSP algorithm adds delay. One is processing delay and the other is algorithmic delay. Processing delay occurs because the processor is not infinitely fast, so it takes some amount of time to perform all of the computation. If the DSP has no extra CPU cycles after performing the computation, then the processing time adds a full frame of delay to the system. If it takes more than a frame of delay to perform the computation, then the system will not make real time. The algorithmic delay comes from any requirement to use data from future frames of data (i.e., buffer the data) in order to make decisions about the current frames of data and other delays inherent in the algorithm process.
D/A	From 1 to 16 samples	As with the A/D converter there is some delay associated with converting a digital signal into an analog signal. Current converters typically have no more than 16 samples of delay.

Often the architecture of the DSP allows the DSP to execute two (or more) floating-point operations per instruction. In this case the MFLOPs would be twice (or more) the MIPs rating of the processor.

Higher-speed processors allow the user to pack more features into a DSP product, but with a higher cost.

Power Consumption. Depending on the application, low power may be important for long battery life or low heat dissipation. DSPs will have a power rating and, often, a watt/MIP rating to estimate power consumption.

Architecture. Different manufacturers' DSPs have different features and trade-offs. Some processors may allow extremely high-speed computational rates but at the expense of being difficult to program. Some may offer ease of multiprocessing, multiple arithmetic processors, or other features.

Arithmetic Precision. The use of floating-point arithmetic simplifies arithmetic operations. Fixed-point processors often have lower cost but often require additional instructions to maintain the level of numerical accuracy that is often required. The final production volume of the end product often dictates whether the added development time is worth the cost savings.

Peripherals. Certain features of processors such as the ability to share processor resources among linked processors or access to external memory/devices can have a significant impact on which processor to use for a particular application. Integrated timers, serial ports, network ports, and other features can reduce the number of additional parts required in a design.

Code Development. The amount of code already developed for a particular processor family may dictate the choice of processors. Real-time code development takes significant time and the investment can be substantial. The ability to reuse existing code is a significant time saver in getting products to market.

Development Tools. The development tools are critical to the timely implementation of an algorithm on a particular processor. If the tools are not available or are not functional, the development process will most likely be extended beyond any reasonable time estimate.

Third Party Support. DSP processor manufacturers have a network of companies that provide tools, algorithm implementations, and hardware solutions for particular problems. It is possible that some company has already implemented, and makes a living out of implementing, the type of solution that is required for a given application.

35.9 Programming a DSP

DSPs, like many other processors, are only useful if they can easily input and output data. The software system used to input and output data is called an I/O system. As shown in Fig. 35-17, a DSP application program typically processes an input stream of data to produce some output data. The processing of this data is performed under the direction of the application program, which usually includes one or more algorithms programmed on the DSP. The DSP application program consists of acquiring the input stream data, using the algorithms to process the data, and then outputting the processed data to the output data stream. An example of this is a speech data compression system where the input stream is a data stream representing uncompressed speech. The output stream, in this case, is the compressed speech data and the application consists of getting the uncompressed input speech data, compressing the data, and then sending the compressed data to the output stream.

One of the most important factors that a DSP I/O system must address is the idea of real-time. An extremely important aspect of these real time A/D and D/A systems is that the samples must be produced and consumed at a fixed rate in order for the system to work in real-time. Although an A/D or D/A converter is a common example of a real-time device, other devices not directly related to real-time data acquisition can also have real time constraints. This is particularly true if they are being used to supply, collect, or transfer real-time information from devices such as disk drives and interprocessor communication links. In the speech compression example, the output stream might be connected to a modem that would transmit the compressed speech to another DSP system that would uncompress the speech. The I/O system should be designed to interface to these devices (or any other) as well.

Another important aspect of a real-time I/O system is the amount of delay (also known as latency) imposed from input to output. For instance, when DSPs are used for in-room sound reinforcement or two-way speech communication (i.e., telecommunications), the delay must be minimized. If the DSP system causes a noticeable delay, the conversation would be awkward and the system would be considered unacceptable. Therefore, the DSP I/O system should be capable of minimizing I/O delay to a reasonable value, on the order of 20–40ms or less.

```c
#include <stdio.h>
#include <custom_io.h>
#include <malloc.h>

#define LEN 800

void main(argc,argp)
char **argp;
int argc;
{
   SIG_Stream input, output;
   SIG_Attrs sig_attrs;
   BUF_Buffer buffer;

   buffer = BUF_create(SEG_DRAM,LEN,0);

   input = SIG_open(argp[1],SIG_READ,buffer,0);

   SIG_getattrs(input,&sig_attrs);

        output =
           SIG_open(argp[2],SIG_WRITE,buffer,&sig_attrs);

  while (SIG_get(input,buffer))
              {
       /* data processing of buffer */
       my_DSP_algorithm(buffer);

       SIG_put(output,buffer);
              }
       return(0);
}
```

Figure 35-18. An example C program for collecting data from an A/D using an input signal stream created with SIG_open and sending data to the D/A using the output signal stream and processing the data with the function my_DSP_algorithm().

Programming a DSP is usually accomplished in a combination of higher-level languages such as C and assembly languages. The higher-level language provides a portable implementation that can potentially be run on multiple different platforms. Assembly language allows for a more computationally efficient implementation at the expense of increased development time and decreased portability. By starting in a higher-level language, the developer can incrementally optimize the implementation by benchmarking which subroutines are taking the most time, optimizing these routines, and then finding the next subroutine to optimize.

A typical C code shell for implementing a DSP algorithm is shown in Fig. 35-18. Here, the C code calls subroutines to allocate some buffer memory to store signal data, open an I/O signal stream, and then get data, processes the data, and then send the data to the output stream. The input and output streams typically have lower level device drivers for talking directly to the A/D and D/A converters, respectively.

35.10 Conclusion

This chapter has introduced the fundamentals of DSP from a theoretical perspective (signal and system theory), and a practical perspective (implementation and latency). The concepts of real-time systems, data acquisition, and selecting digital signal processors have also been introduced. DSP is a large and encompassing subject and the interested reader is encouraged to learn more through the exhaustive treatment given to this material in the references.[1,6]

References

1. Alan V. Oppenheim, Ronald W. Schafer, with John R. Buck, *Discrete-Time Signal Processing*, 2nd Edition, Upper Saddle River, NJ:Prentice Hall, 1999.

2. Alan V. Oppenheim, A. S. Willsky, with S. H. Nawab, *Signals and Systems*, 2nd Edition, Upper Saddle River, NJ: Prentice Hall, 1997.

3. Pervez M. Aziz, Henrik V. Sorensen, and Jan Van Der Spiegel, “An Overview of Sigma-Delta Converters,” *IEEE Signal Processing Magazine*, Vol. 13, no. 1, January 1996, pp. 61–84.

4. Edward A. Lee, “Programmable DSP Architectures: Part I,” *IEEE ASSP Magazine*, Vol. 5, no. 4, October 1988, pp. 4–19.

5. Edward A. Lee, "Programmable DSP Architectures: Part II," *IEEE ASSP Magazine*, Vol. 6, no. 1, January 1989, pp. 4–14.

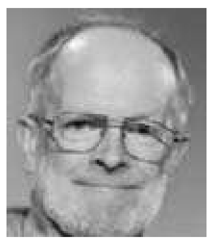
6. James H. McClellan, Ronald W. Schafer, Mark Yoder, *DSP First: A Multimedia Approach*, Upper Saddle River, NJ:Prentice Hall 1999.

Chapter 36

Grounding and Interfacing

by Bill Whitlock

Dedication



Neil Muncy, 1938-2012. I knew Neil Muncy for nearly 20 years. We worked together in the Audio Engineering Society standards working group on grounding and EMC practices, which started many lively, and sometimes contentious, technical conversations. In one of those, he asked me "Why is it that transformers can solve noise problems nothing else can?" Discovering the answer kick-started my quest to understand grounding and signal interfaces as completely as I could. Neil was both an inspiring colleague and a kind, generous friend. I dedicate this chapter to his memory.

Bill Whitlock

36.1 Introduction

Many audio professionals think of system grounding as a black art. How many times have you heard someone say that a cable is picking up noise, presumably from the air like a radio receiver? Even equipment manufacturers often don't have a clue what's really going on when there's a problem. The most basic rules of physics are routinely overlooked, ignored, or forgotten. As a result, myth and misinformation have become epidemic! This chapter is intended to enable sound engineers to understand and either avoid or solve real-world noise problems. The electronic system engineering joke that cables are sources of potential trouble connecting two other sources of potential trouble contains more truth than humor. Because equipment ground connections have profound effects on noise coupling at signal interfaces, we must appreciate how interfaces actually work as well as when, why, and how equipment is grounded. Although the subject can't be reduced to just a few simple rules, it doesn't involve rocket science or complex math either.

For convenience in this chapter, we'll use the term *noise* to mean to signal artifacts that originate from sources *external* to the signal path. This includes hum, buzz, clicks, or pops originating from the power line and interference originating from radio-frequency devices. A predictable amount of white noise is inherent in all electronic devices and must be expected. This random noise, heard as hiss, will also limit the usable dynamic range of any audio system, but this is not the subject of this chapter!

Any signal accumulates noise as it flows through the equipment and cables in a system. Once it contaminates a signal, noise is essentially impossible to remove without altering or degrading the signal. Therefore, noise and interference must be prevented along the entire signal path. It might seem trivial to transfer signal from the output of one audio device to the input of another but, in terms of noise and interference, signal interfaces are truly the danger zone! Let’s start with some basic electronics that apply to interfaces.

36.2 Basic Electronics

Fields can exert invisible forces on objects within them. In electronics, we’re concerned with electric and magnetic fields. Almost everyone has seen a demonstration of iron filings sprinkled on paper used to visualize the magnetic field between the north and south poles of a small magnet. A similar electric field exists between two points having a constant voltage difference between them. Fields like these, which neither move nor change in intensity, are called *static* fields.

If a field, either magnetic or electric, moves in space or fluctuates in intensity, the other kind of field will be generated. In other words, a changing electric field will set up a changing magnetic field or a changing magnetic field will set up a changing electric field.

This interrelationship gives rise to *electromagnetic waves*, in which energy is alternately exchanged between electric and magnetic fields as they travel through space at the speed of light.

Everything physical is made of atoms whose outermost components are electrons. An electron carries a negative electric charge and is the smallest quantity of electricity that can exist. Some materials, called *conductors* and most commonly metals, allow their outer electrons to move freely from atom to atom. Other materials, called *insulators* and most commonly air, plastic, or glass, are highly resistant to such movement. This movement of electrons is called current flow. Current will flow only in a complete circuit consisting of a connected source and load. **Regardless of how complex the path becomes, all current leaving a source must return to it!**

36.2.1 Circuit Theory

An electric potential or *voltage*, sometimes called *emf* for electromotive force, is required to cause current flow. It is commonly denoted E (from emf) in equations and its unit of measure is the *volt*, abbreviated V. The resulting rate of current flow is commonly denoted I (from intensity) in equations and its unit of measure is the ampere, abbreviated A. How much current will flow for a given applied voltage is determined by circuit resistance. Resistance is denoted R in equations and its unit of measure is the ohm, symbolized Ω.

Ohm’s Law defines the quantitative relationship between basic units of voltage, current, and resistance:

$$E = I \times R$$

which can be rearranged as

$$R = \frac{E}{I}$$

$$I = \frac{E}{R}$$

For example, a voltage E of 12V applied across a resistance R of 6Ω will cause a current flow I of 2A.

Circuit elements may be connected in parallel, series, or combinations of both, Figs. 36-1 and 36-2. Although the resistance of wires that interconnect circuit elements is generally assumed to be negligible, we will discuss this later.

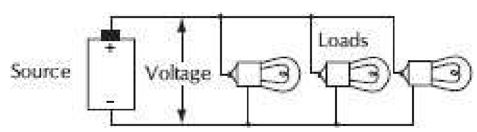


Figure 36-1. The voltage is the same across all elements in a parallel circuit.

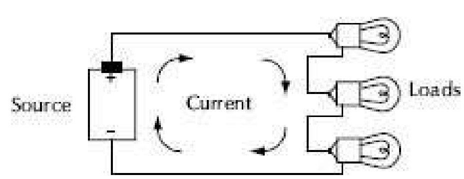


Figure 36-2. The current is the same through all elements in a series circuit.

In a parallel circuit, the total source current is the sum of the currents through each circuit element. The highest current will flow in the lowest resistance, according to Ohm's Law. The equivalent single resistance seen by the source is always lower than the lowest resistance element and is calculated as

$$R_{EQ} = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R} \ldots + \frac{1}{n}} \qquad (36\text{-}1)$$

In a series circuit, the total source voltage is the sum of the voltages across each circuit element. The highest voltage will appear across the highest resistance, according to Ohm's Law. The equivalent single resistance seen by the source is always higher than the highest resistance element and is calculated as

$$R_{EQ} = R1 + R2 + R3 \ldots + Rn \qquad (36\text{-}2)$$

Voltages or currents whose value (magnitude) and direction (polarity) are steady over time are generally referred to as *dc*. A battery is a good example of a dc voltage source.

36.2.2 Ac Circuits

A voltage or current that changes value and direction over time is generally referred to as *ac*. Consider the voltage at an ordinary 120V, 60Hz ac receptacle.

Since it varies over time according to a mathematical sine function, it is called a *sine wave*. Fig. 36-3 shows how it would appear on an oscilloscope where time is the horizontal scale and instantaneous voltage is the vertical scale with zero in the center. The instantaneous voltage swings between peak voltages of +170V and – 170V. A cycle is a complete range of voltage or current values that repeat themselves periodically (in this case every 16.67ms). Phase divides each cycle into 360° and is used mainly to describe instantaneous relationships between two or more ac waveforms. Frequency indicates how many cycles occur per second of time. Frequency is usually denoted *f* in equations, and its unit of measure is the hertz, abbreviated Hz. Audio signals rarely consist of a single sine wave. Most often they are complex waveforms consisting of many simultaneous sine waves of various amplitudes and frequencies in the 20Hz to 20,000Hz (20kHz) range.

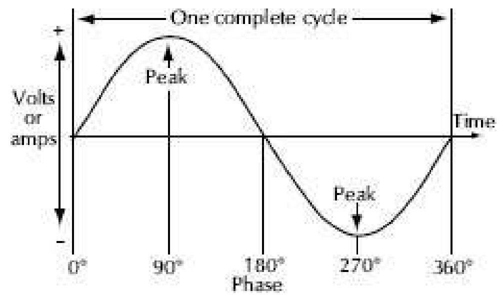


Figure 36-3. Sine wave as displayed on an oscilloscope.

36.2.3 Capacitance, Inductance, and Impedance

An electrostatic field exists between any two conductors having a voltage difference between them. Capacitance is the property that tends to oppose any change in the strength or charge of the field. In general, capacitance is increased by larger conductor surface areas and smaller spacing between them. Electronic components expressly designed to have high capacitance are called *capacitors*. Capacitance is denoted *C* in equations and its unit of measure is the Farad, abbreviated F. It's very important to remember that unintentional or parasitic capacitances exist virtually everywhere. As we will see, these parasitic capacitances can be particularly significant in cables and transformers!

Current must flow in a capacitor to change its voltage. Higher current is required to change the voltage rapidly and no current will flow if the voltage is held constant. Since capacitors must be alternately charged and discharged in ac circuits, they exhibit an apparent ac resistance called *capacitive reactance*. Capacitive reactance is inversely proportional to both capacitance and frequency since an increase in either causes an increase in current, corresponding to a decrease in reactance.

$$X_C = \frac{1}{2\pi fC} \tag{36-3}$$

where,
X_C is capacitive reactance in Ω,
f is frequency in Hz,
C is capacitance in F.

In general, capacitors behave as open circuits at dc and gradually become short circuits, passing more and more current, as frequency increases.

As shown in Fig. 36-4, a magnetic field exists around any conductor carrying current at right angles to the axis of flow. The strength of the field is directly proportional to current. The direction, or polarity, of the magnetic field depends on the direction of current flow. Inductance is the property that tends to oppose any change in the strength or polarity of the field. Note that the fields around the upper and lower conductors have opposite polarity. The fields inside the loop point in the same direction, concentrating the field and increasing inductance. An electronic component called an *inductor* (or *choke*) is most often made of a wire coil with many turns to further increase inductance. Inductance is denoted L in equations and its unit of measure is the henry, abbreviated H. Again, remember that unintentional or parasitic inductances are important, especially in wires!

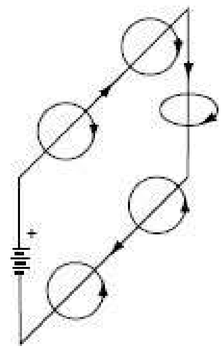


Figure 36-4. Magnetic field surrounding conductor.

If we abruptly apply a dc voltage to an inductor, a magnetic field is generated within the wire and moves outward as current begins to flow. But, in accordance with the law of induction, the rising field strength will induce a voltage, called *back emf*, in the wire which works to oppose current flow. The faster the field increases its strength, the more back emf will be induced to oppose current flow. The net result is to slow the buildup of current as it approaches its final value, which is limited by the applied voltage and circuit resistance. In ac circuits, for a constant applied voltage, this slowing reduces current flow as frequency increases because less time is available each cycle for current to rise. This apparent increase in ac resistance is called *inductive reactance*. Inductive reactance increases in direct proportion to both inductance and frequency.

$$X_L = 2\pi f L \qquad (36\text{-}4)$$

where,
X_L is inductive reactance in Ω,
f is frequency in Hz,
L is inductance in H.

In summary, inductors behave as short circuits at dc and gradually become open circuits, passing less and less current, as frequency increases.

Impedance is the combined effect of both resistance and reactance for circuits that contain resistance, capacitance, and inductance, which is the case with virtually all real-world circuits. Impedance is represented by the letter *Z* and is measured in ohms. Impedance can be substituted for *R* in the Ohm's Law equations. Impedance is a more general term than either resistance or reactance and, for ac circuits is the functional equivalent of resistance.

36.2.4 Single Wires

The electrical properties of wire are often overlooked. Consider a 10ft length of #12 AWG solid copper wire.

1. The resistance of a wire is directly proportional to its length, inversely proportional to its diameter, and depends strongly on the material. From standard wire tables, we find the dc resistance of #12AWG annealed copper wire is 1.59Ω/1000ft or 0.0159Ω for a 10ft length. At frequencies below about 500Hz, this resistance largely sets the impedance.
2. The inductance of a straight wire is nearly independent of its diameter but is directly proportional to its length. From the formula for the inductance of a straight round wire,[1] we find its inductance is 4.8μH. As shown in Fig. 36-5, this causes a rise in impedance beginning at about 500Hz, reaching 30Ω at 1MHz (AM radio). Replacing the wire with a massive ½in diameter copper rod would reduce impedance only slightly to 23Ω.
3. Electromagnetic waves travel through space or air at the speed of light. The physical distance traveled by a wave during one

cycle is called *wavelength*. The equation is

$$M = \frac{984}{f} \qquad (36\text{-}5)$$

where,
M is wavelength in ft,
f is frequency in MHz.

For 1MHz AM radio, 100MHz FM radio, and 2GHz cell phone signals, wavelengths are about 1000ft, 10ft, and 6in, respectively.

4. Any wire will behave as an antenna at frequencies where its physical length is a quarter-wavelength or multiples thereof. This is responsible for the impedance peaks and dips seen at 25MHz intervals in Fig. 36-5.

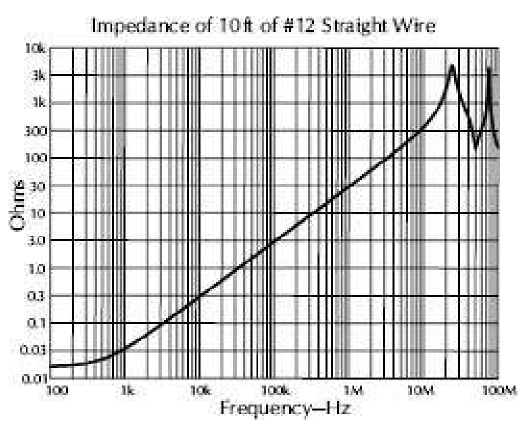


Figure 36-5. Wire is low-impedance current path only at low frequencies.

36.2.5 Cables and Transmission Lines

A cable consists of two or more conductors that are kept in close proximity over their length. Cables, such as those for ac power and loudspeakers, are generally used to convey power to a load. In a pair of such conductors, because the same current flows to and from the load in opposite directions, the magnetic fields have the same intensity but are of opposite polarity as shown in Fig. 36-6. In theory, there would be zero external field, and zero net inductance, if the two conductors could occupy the same space. The cancellation of round trip inductance due to magnetic coupling varies with cable construction, with typical values of 50% for zip cord, 70% for a twisted pair, and 100% for coaxial construction.

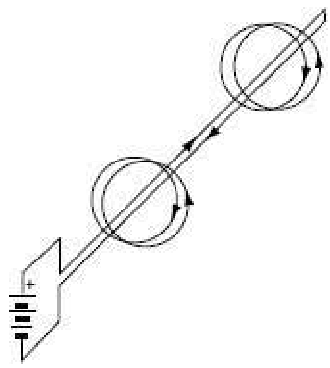


Figure 36-6. Cancellation of field in circuit pair.

At very high frequencies, a cable exhibits very different characteristics than it does at, say, 60Hz power frequencies. This is caused by the finite speed, called *propagation velocity*, at which electrical energy travels in wires. It is about 70% of the speed of light for typical cables making wavelengths in cable correspondingly shorter. A cable is called electrically short when its physical length is under 10% of a wavelength at the highest frequency of interest. Wavelength at 60Hz for typical cable is about 2200miles (mi), making any power cable less than 220mi long electrically short. Likewise, the wavelength at 20kHz for typical cable is about 34,500ft, making any audio cable less than about 3500ft long electrically short. Essentially identical instantaneous voltage and current exists at all points on an electrically short cable and its signal coupling behavior can be represented by lumped resistance, capacitance, and magnetically coupled inductance as shown in Fig. 36-7. Its equivalent circuit can then be analyzed by normal network theory.

When a cable is longer than 10% of a wavelength, signals must be considered to propagate as electromagnetic waves and the cable can properly be called a *transmission line*. This includes typical cables longer than 7ft for 10MHz video, 8in for 100MHz FM radio, and 0.8in for 1000MHz CATV signals. Significantly different instantaneous voltages exist along the length of a transmission line. For all practical purposes, its electrical equivalent is a distributed circuit consisting of a large number of small inductors and resistors in series and capacitors in parallel. If an electrical impulse were applied to one end of an infinitely long cable, it would appear to have a purely resistive impedance. This *characteristic impedance* of the cable is a result of its inductance and capacitance per unit length, which is determined by its physical construction. Theoretically, the electrical impulse or wave would ripple down the infinite length of the cable forever. But actual transmission lines always have a far end. If the far end is left open or shorted, none of the wave's energy can be absorbed and it will reflect back toward the source. However, if the far end of the line is terminated with a resistor of the same value as the line's characteristic impedance, the wave energy will be completely absorbed. To the wave, the termination appears to be simply more cable. A properly terminated transmission line is often said to be matched. Generally, impedances of both the driving source and the receiving load are matched to the characteristic impedance of the line. In a mismatched line, the interaction between outgoing and reflected waves causes a phenomenon called *standing waves*. A measurement called *standing-wave ratio* (SWR) indicates mismatch, with an SWR of 1.00 meaning a perfect match.

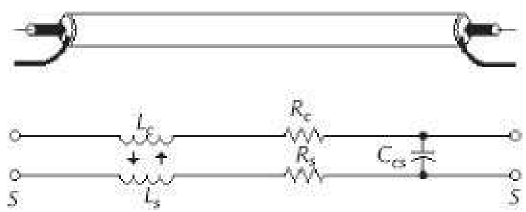


Figure 36-7. Lumped-circuit model of electrically short coaxial cable.

36.3 Electronics of Interfaces

36.3.1 Balanced and Unbalanced Interfaces

An *interface* is a signal transport subsystem consisting of three components: a driver (one device's output), a line (interconnecting cable), and a receiver (another device's input). These components are connected to form a complete circuit for signal current, which requires a line having two signal conductors. The impedances of the signal conductors, usually with respect to ground, are what determine whether an interface is balanced or unbalanced. A concise definition of a balanced circuit is:

> *A balanced circuit is a two-conductor circuit in which both conductors and all circuits connected to them have the same impedance with respect to ground and to all other conductors. The purpose of balancing is to make the noise pickup equal in both conductors, in which case it will be a common-mode signal that can be made to cancel out in the*

load.[2]

The use of balanced interfaces is an extremely potent technique to prevent noise coupling into signal circuits. It is so powerful that many systems, including telephone systems, use it in place of shielding as the main noise reduction technique!

Theoretically, a balanced interface can reject any interference, whether due to ground voltage differences, magnetic fields, or electric fields, as long as it produces identical voltages on each of the signal lines and the resulting peak voltages don't exceed the capabilities of the receiver.

A simplified balanced interface is shown in Fig. 36-8. Any voltage that appears on both inputs, since it is common to the inputs, is called a *commonmode* voltage. A balanced receiver uses a differential device, either a specialized amplifier or a transformer, that inherently responds only to the difference in voltage between its inputs. By definition, such a device will reject, i.e., have no response to, common-mode voltages. The ratio of differential gain to common-mode gain of this device is its *common-mode rejection ratio*, or CMRR. It's usually expressed in dB, and higher numbers mean more rejection. Section 36.5.1 will describe how CMRR often degrades in real-world systems and how it has traditionally been measured in ways that have no relevance to real-world system performance.

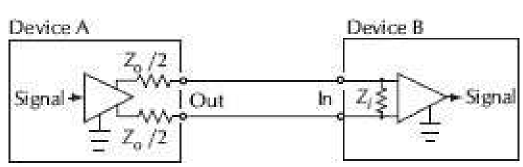


Figure 36-8. Basic balanced interface.

Two signal voltages have symmetry when they have equal magnitudes but opposite polarities. Symmetry of the desired signal has advantages, but they concern head room and crosstalk, not noise or interference rejection. The noise or interference rejection property is independent of the presence of a desired differential signal. Therefore, it can make no difference whether the desired signal exists entirely on one line, as a greater voltage on one line than the other, or as equal voltages on both of them. However, the symmetry myth is widespread. A typical example is: Each conductor is always equal in voltage but opposite in polarity to the other. The circuit that receives this signal in the mixer is called a *differential amplifier* and this opposing polarity of the conductors is essential for its operation.[3] Like many others, it describes a balanced interface in terms of signal symmetry but never mentions impedances! Even the BBC test for output balance is actually a test for signal symmetry.[4] The idea that balanced interface is somehow defined by signal symmetry is simply wrong! It has apparently led some designers, mostly of exotic audiophile gear, to dispense with a differential amplifier input stage in their push-pull amplifiers. They simply amplify the (assumed) symmetrical input signals in two identical, ground-referenced amplifier chains. No mechanism exists to reject common-mode voltage (noise and interference) and it is actually amplified along with the signal, creating potentially serious problems. Rejection of common-mode voltages is the single most important function of a balanced receiver.

In an unbalanced circuit, one signal conductor is grounded (near-zero impedance) and the other has some higher impedance. As we will discuss in section 36.5.4, the fact that not only signal but ground noise currents flow and cause voltage drops in the grounded conductor makes an unbalanced interface inherently susceptible to a variety of noise problems.

36.3.2 Voltage Dividers and Impedance Matching

Every driver has an internal impedance, measured in ohms, called its *output impedance*. Actual output impedance is important, as we discuss below, but often absent from equipment specifications. Sometimes, especially for consumer gear, the only impedance associated with an output is listed as recommended load impedance. While useful if listed in addition to output impedance, it is not what we need to know! A perfect driver would have a zero output impedance but, in practical circuit designs, it's neither possible nor necessary. Every receiver has an internal impedance, measured in ohms, called its *input impedance*. A perfect receiver would have an infinite input impedance but again, in practical circuit designs, it's neither possible nor necessary.

Figs. 36-8 and 36-9 illustrate ideal interfaces. The triangles represent ideal amplifiers having infinite impedance input, i.e., draw no current, and zero impedance output, i.e., deliver unlimited current, and the line conductors have no resistance, capacitance, or inductance. The signal voltage from the driver amplifier causes current flow through the driver output impedance(s) Z_o, the line, and receiver input impedance Z_i. Note that the output impedance of the balanced driver is split into two equal parts. Because current is the same in all parts of a series circuit and voltage drops are proportional to impedances, this circuit is called a *voltage divider*.

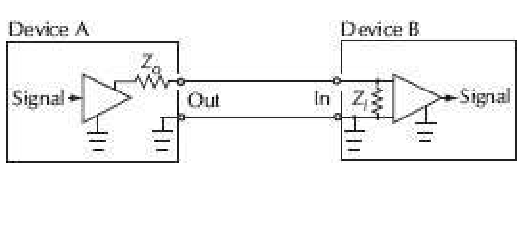


Figure 36-9. Basic unbalanced interface.

The goal of an interface is, with rare exception, to deliver maximum signal voltage from the output of one device to the input of another. Making Z_i much larger than Z_o assures that most of the signal voltage is delivered to the receiver and very little is lost in the driver. In typical devices, Z_o ranges from 30Ω to 1KΩ and Z_i ranges from 10kΩ to 100kЦ which transfers 90–99.9% of the available, i.e., unloaded or open circuit, signal voltage.

Matching is a term that often causes confusion. A little math and Ohm's Law will prove that when Z_o and Z_i are equal, maximum power is transferred from source to load, although half the signal voltage is lost. If transmission line effects apply, Z_o and Z_i must terminate or match the characteristic impedance of the line to prevent reflection artifacts. Although modern audio systems seldom use cables long enough for transmission line effects to apply or benefit from maximum power transfer, early telephone systems did both. Telephone systems began by using miles of existing open wire telegraph lines that, due to their wire size and spacing, had a characteristic impedance of 600Ω. Since amplifiers didn't exist, the system was entirely passive and needed to transfer maximum power from one phone to another. Therefore, transformers, filters, and other components were designed for 600Ω impedances to match the lines. These components were eventually incorporated into early sound reinforcement, radio, and recording systems. And the 600Ω legacy still lives on, even though modern requirements for it are all but extinct.

Sometimes, instead of meaning equal, matching is used to mean optimizing some aspect of circuit performance. For example, the output transformer in a vacuum-tube power amplifier is used to optimize power output by converting or impedance matching the low-impedance loudspeaker to a higher impedance that suits the characteristics of the tubes. Similarly, the modern technique of making Z_i much larger than Z_o to transfer maximum voltage in signal interfaces is often referred to as *voltage matching*. It uses 10kΩ or higher input impedances, called bridging because many inputs can be paralleled across the same output line with negligible drop in level. About 60Ω has been suggested as the optimum Z_o for driving up to 2000ft of typical shielded twisted pair cable in these balanced interfaces.[5]

36.3.3 Line Drivers and Cable Capacitance

A line driver and cable interact in two important ways. First, output impedance Z_o and the cable capacitance form a low-pass filter that will cause high-frequency roll-off. A typical capacitance for either unbalanced or balanced shielded audio cable might be about 50pF/ft. If output impedance were 1kΩ (not uncommon in unbalanced consumer gear), response at 20kHz would be −0.5dB for 50ft, −1.5dB for 100ft, and −4dB for 200ft of cable. If the output impedance were 100Ω (common in balanced pro gear), the effects would be negligible for the same cable lengths. Low-output impedance is especially important when cable runs are long. Also be aware that some exotic audio cables have extraordinarily high capacitance.

Second, cable capacitance requires additional high-frequency current from the driver. The current required to change the voltage on a capacitance is directly proportional to the rate of change or *slew rate* of the voltage. For a sine wave,

$$SR = 2\pi f V_p \tag{36-6}$$

where,
SR is slew rate in V/s,
f is frequency in Hz,
V_p is peak voltage.

$$I = SR \times C \tag{36-7}$$

where,
I is current in A,
SR is slew rate in V/μs,
C is capacitance in μF.

For example, the slew rate is 1V/μs at 20kHz for a sine-wave of $8V_p$ or $5.6V_{rms}$, which is also +17dBu or +15dBV. For a cable of 100ft at 50pF/ft, C would be 5000pF or 0.005μF. Therefore, peak currents of 5mA are required to drive just the cable capacitance to +17 dBu at 20kHz. Obviously, increasing level, frequency, cable capacitance, or cable length will increase the current required. Under the previous conditions, a cable of 1000ft would require peak currents of 50mA. Such peak currents may cause protective current limiting or clipping in the op-amps used in some line drivers. Since it occurs only at high levels and high frequencies, the audible effects may be subtle.

Of course, the load at the receiver also requires current. At a +17dBu level, a normal 10kΩ balanced input requires a peak current of only 0.8mA. However, a 600Ω termination at the input requires 13mA. Matching 600Ω sources and loads not only places a current burden on the driver but, because 6dB (half) of signal voltage is lost, the driver must generate +23dBu to deliver +17dBu to the input.

Unnecessary termination wastes driver current and unnecessary matching of source and load impedances wastes head room!

36.3.4 Capacitive Coupling and Shielding

Capacitances exist between any two conductive objects, even over a relatively large distance. As we mentioned earlier, the value of this capacitance depends on the surface areas of the objects and the distance. When there are ac voltage differences between the objects, these capacitances cause small but significant currents to flow from one object to another by means of the changing electric field (widely referred to as *electrostatic* fields although technically a misnomer since static means unchanging).

Strong electric fields radiate from any conductor operating at a high ac voltage and, in general, weaken rapidly with distance. Factors that increase coupling include increasing frequency, decreasing spacing of the wires, increasing length of their common run, increasing impedance of the victim circuit, and increasing distance from a ground plane. For some of these factors, there is a point of diminishing returns. For example, for parallel 22-gauge wires, there is no significant reduction in coupling for spacing over about 1in.[6] Capacitive coupling originates from the voltage at the source. Therefore, coupling from a power circuit, for example, will exist whenever voltage is applied to the circuit regardless of whether load current is flowing.

Capacitive coupling can be prevented by placing electrically conductive material called a *shield* between the two circuits so that the electric field, and the resulting current flow, linking them is diverted. A shield is connected to a point in the circuit where the offending current will be harmlessly returned to its source, usually called *ground*—more about ground later. For example, capacitive coupling between a sensitive printed wiring board and nearby ac power wiring could be prevented by locating a grounded metal plate (shield) between them, by completely enclosing the board in a thin metal box, or by enclosing the ac power wiring in a thin metal box.

Similarly, as shown in Fig. 36-10, shielding can prevent capacitive coupling to or from signal conductors in a cable. Solid shields, such as conduit or overlapped foil, are said to have 100% coverage. Braided shields, because of the tiny holes, offer from 70% to 98% coverage. At very high frequencies, where the hole size becomes significant compared with interference wavelength, cables with combination foil/braid or multiple braided shields are sometimes used.

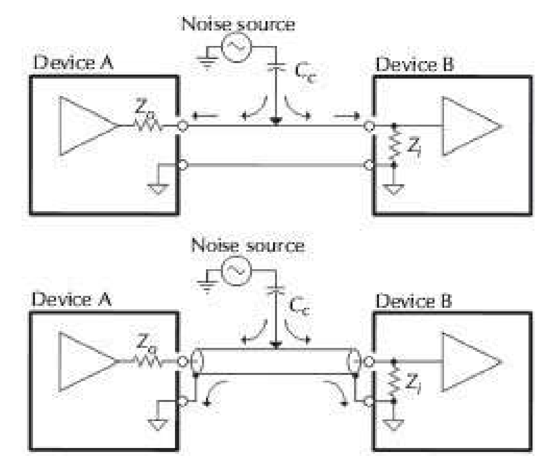


Figure 36-10. Capacitive noise coupling.

36.3.5 Inductive Coupling and Shielding

When any conductor cuts magnetic lines of force, in accordance with the law of induction, a voltage is induced in it. If an alternating current flows in the conductor, as shown at the left in Fig. 36-11, the magnetic field also alternates, varying in intensity and polarity. We can visualize the magnetic field, represented by the concentric circles, as expanding and collapsing periodically. Because the conductor at the right cuts the magnetic lines of force as they move across it, an ac voltage is induced over its length. This is the essential principle of a transformer. Therefore, current flowing in a wire in one circuit can induce a noise voltage in another wire in a different circuit. Because the magnetic field is developed only when current flows in the source circuit, noise coupling from an ac power circuit, for example, will exist only when load current actually flows.

If two identical conductors are exposed to identical ac magnetic fields, they will have identical voltages induced in them. If they are series connected as shown in Fig. 36-12, their identical induced voltages tend to cancel. In theory, there would be zero output if the two conductors could occupy the same space.

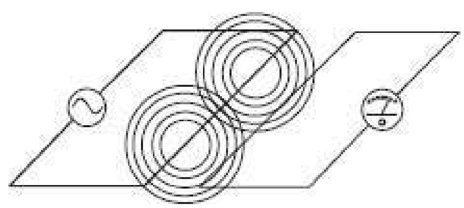


Figure 36-11. Inductive coupling between wires.

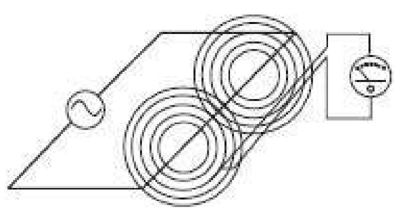


Figure 36-12. Coupling cancellation in loop.

Magnetic fields become weaker rapidly as distance from the source increases, usually as the square of the distance. Therefore, cancellation depends critically on the two conductors being at precisely the same distance from the magnetic field source.

Twisting essentially places each conductor at the same average distance from the source. So-called *star quad* cable uses four conductors with those opposing each other connected in parallel at each cable end. The effective magnetic center for each of these pairs is their center line and the two sets of pairs now have coincident center lines reducing the loop area to zero. Star quad cable has approximately 100 times (40dB) better immunity to power-frequency magnetic fields than standard twisted pair. The shield of a coaxial cable also has an average location coincident with the center conductor. These construction techniques are widely used to reduce susceptibility of balanced signal cables to magnetic fields. In general, a smaller physical area inside the loop results in less magnetic radiation as well as less magnetic induction.

Another way to reduce magnetic induction effects is shown in Fig. 36-13. If two conductors are oriented at a 90° (right angle), the second doesn't cut the magnetic lines produced by the first and will have zero induced voltage. Therefore, cables crossing at right angles have minimum coupling and those running parallel have maximum coupling. The same principles also apply to circuit board traces and internal wiring of electronic equipment.

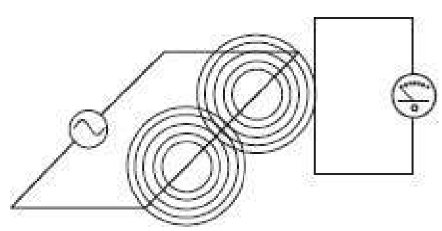


Figure 36-13. Zero coupling at right angles.

Magnetic circuits are similar to electric circuits. Magnetic lines of force always follow a closed path or circuit, from one magnetic pole to the opposite pole, always following the path of least resistance or highest conductivity. The magnetic equivalent of electric current and conductivity are flux density and permeability. High-permeability materials have the ability to concentrate the magnetic force lines or flux. The permeability of air and other nonmagnetic materials such as aluminum, plastic, or wood is 1.00. The permeability of common ferromagnetic materials is about 400 for machine steel, up to 7000 for common 4% silicon transformer steel, and up to 100,000 for special nickel alloys. The permeability of magnetic materials varies with flux density. When magnetic fields become very intense, the material can become saturated, essentially losing its ability to offer an easy path for any additional flux lines. Higher permeability materials also tend to saturate at a lower flux density and to permanently lose their magnetic properties if mechanically stressed.

The basic strategy in magnetic shielding is to give the flux lines a much easier path to divert them around a sensitive conductor, circuit, or device. In general, this means that the shield must be a complete enclosure with a high magnetic permeability. The choice of the most effective shielding material depends on frequency. At low frequencies, below say 100kHz, high-permeability magnetic materials are most effective. We can calculate how effective a conduit or cable shield will be at low frequencies:

$$SE = 20\log\left(1 + \frac{\mu t}{d}\right) \qquad (36\text{-}8)$$

where,

SE is shielding effect in dB,

μ is permeability of shield material,

t and *d* are the thickness and diameter (in the same units) of the conduit or shield.[7]

Thus, standard 1 inch EMT, made of mild steel with a low-frequency permeability of 300, will provide about 24dB of magnetic shielding at low frequencies, but this will diminish to zero around 100kHz. Fortunately, only low-frequency magnetic fields are generally a problem. In severe cases, nesting one magnetic shield inside another may be necessary.

Typical copper braid or aluminum foil cable shielding has little effect on magnetic fields at audio frequencies. If a shield is grounded at both ends, it behaves somewhat like a shorted turn to shield the inner conductors from magnetic fields.[8] Depending on the external impedance between the grounded ends of a cable shield, it may begin to become effective against magnetic fields somewhere in the 10kHz to 100kHz range. Box shields of aluminum or copper are widely used to enclose RF circuits because they impede magnetic fields through this eddy current action and are excellent shielding for electric fields as well. There is an excellent explanation of this high-frequency shielding in reference 9. However, copper or aluminum shielding is rarely an effective way to prevent noise coupling from audio-frequency magnetic fields.

36.4 Grounding

Historically, grounding became necessary for protection from lightning strokes and industrially-generated static electricity, i.e., belts in a flour mill. As utility power systems developed, grounding became standard practice to protect people and equipment. As electronics developed, the common return paths of various circuits were referred to as ground, regardless of whether or not they were eventually connected to earth. Thus, the very term *ground* has become vague, ambiguous, and often fanciful. Broadly, the purpose of grounding is to electrically interconnect conductive objects, such as equipment, in order to minimize voltage differences between them. An excellent general definition is that a *ground is simply a return path for current*, which will always return to its source. The path may be intentional or accidental—electrons don't care and don't read schematics![10]

Grounding-related noise can be the most serious problem in any audio system. Common symptoms include hum, buzz, pops, clicks, and other noises. Because equipment manufacturers so often try to explain away these problems with the nebulous term *bad grounding*, most system installers and technicians feel that the entire subject is an incomprehensible black art. Adding to the confusion are contradictory rules proposed by various experts. Ironically, most universities teach very little about the real-world aspects of grounding. Graduates take with them the grounding fantasy that all grounds are equipotential, that is, have the same voltage. The fantasy certainly allows them to avoid complicated real-world interpretation of all those ground symbols on a schematic diagram, but the same fantasy can lead to noise disaster in their audio equipment and system designs.

Grounding has several important purposes and most often a single ground circuit serves, intentionally or accidentally, more than one purpose. We must understand how these ground circuits work and how noise can couple into signal circuits if we expect to control or eliminate noise in audio systems.

36.4.1 Earth Grounding

An *earth ground* is one actually connected to the earth via a low-impedance path. In general, earth grounds are necessary only to protect people from lightning. Before modern standards such as the National Electrical Code (NEC or just Code) were developed, lightning that struck a power line was often effectively routed directly into buildings, starting a fire or killing someone. Lightning strikes are the discharge of giant capacitors formed by the earth and clouds. Strikes involve millions of volts and tens of thousands of amperes, producing brief bursts of incredible power in the form of heat, light, and electromagnetic fields. Electrically, lightning is a high-frequency event, with most of its energy concentrated in frequencies over 300kHz! That's why, as we discussed in section 36.2.4, wiring to ground rods should be as short and free of sharp bends as possible. The most destructive effects of a strike can be avoided by simply giving the current an easy, low-impedance path to earth before it enters a building. Because overhead power lines are frequent targets of lightning, virtually all modern electric power is distributed on lines having one conductor that is connected to earth ground frequently along its length.

Fig. 36-14 shows how ac power is supplied through a three-wire split single-phase service to outlets on a typical 120Vac branch circuit in a building. One of the service wires, which is often uninsulated, is the grounded neutral conductor. Note that both the white neutral and the green safety ground wires of each branch circuit are tied or bonded to each other and an earth ground rod (or its equivalent grounding electrode system) at the service entrance as required by Code. *This earth ground, along with those at neighboring buildings and at the utility poles, provide the easy paths for lightning to reach earth.*

Telephone, CATV, and satellite TV cables are also required to divert or arrest lightning energy before it enters a building. The telco-supplied gray box or NIU provides this protection for phone lines as × grounding blocks do for CATV and satellite dishes. NEC Articles 800, 810, and 820 describe requirements for telephone, satellite/TV antennas, and CATV, respectively. All protective ground connections should be made to the same ground rod used for the utility power, if the ground wire is 20ft or less in length. If longer, separate ground rods must be used, and they must be bonded to the main utility power grounding electrode with a #6 AWG wire.[11] Otherwise, because of considerable soil resistance between separate ground rods, thousands of volts could exist between them when lightning events occur or downed power lines energize the signal lines. Without the bond such events could seriously damage a computer modem, for example, that straddles a computer grounded to one rod via its power cord and a telephone line protectively grounded to another.[12]

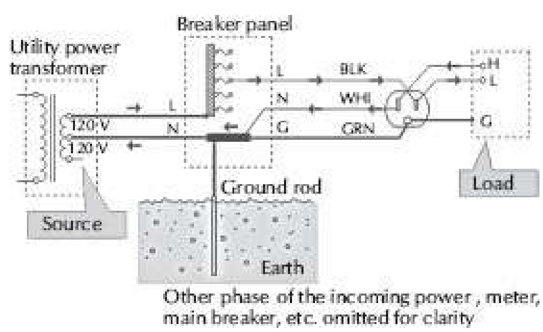


Figure 36-14. Simplified residential ac power from feeder to outlet.

36.4.2 Fault or Safety Grounding

Any ac line powered device having conductive exposed parts (which includes signal connectors) can become a shock or electrocution hazard if it develops certain internal defects. Insulation is used in power transformers, switches, motors, and other internal parts to keep the electricity where it belongs. But, for various reasons, the insulation may fail and effectively connect live power to exposed metal. This kind of defect is called a *fault*. A washing machine, for example, could electrocute someone who happened to touch the machine and a water faucet (assumed grounded via buried metal pipes) at the same time.

NEC requires that 120Vac power distribution in homes and buildings use a three-wire system as shown in Fig. 36-15. To prevent electrocution, most devices have a third wire connecting exposed metal to the safety ground pin of these outlets. The outlet safety ground is routed, through either the green wire or metallic conduit, to the neutral conductor and earth ground at the main breaker panel. The connection to neutral allows high fault current to flow, quickly tripping the circuit breaker, while the earth ground connection minimizes any voltage that might exist between equipment and other earth-grounded objects, such as water pipes, during the fault event. Power engineers refer to voltage differences created by these fault events as *step* or *touch* potentials. The neutral (white) and line (black) wires are part of the normal load circuit that connects the voltage source to the load. The green wire or conduit is intended to carry fault currents only.

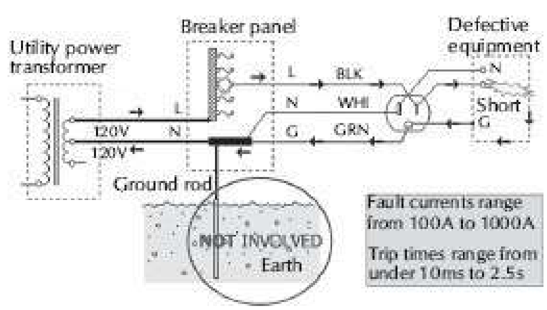


Figure 36-15. Fault protection is provided by safety ground to neutral bond.

NEC also requires safety grounding of wiring raceways and equipment cabinets, including rack cabinets. Per Article 250-95, safety grounding wires, which may be bare or insulated, must have a minimum size of #14 copper for a 15A or #12 copper for a 20A branch circuit to assure rapid circuit breaker action. This grounding path must be bonded to the safety grounding system, not to building steel or a separate earth ground system! *Separate earth grounds cannot provide safety grounding*!! As shown in Fig. 36-16, soil resistance is far too high to guarantee tripping of a circuit breaker under fault conditions.[13] With safety grounds in place, potentially deadly equipment faults simply cause high currents from power line hot to safety ground, quickly tripping circuit breakers and removing power from those branch circuits. Safety grounding in many residential and commercial buildings is provided through metal conduit, metallic J-boxes, and saddle-grounded or SG outlets. Technical or isolated grounding will be discussed in section 36.7.

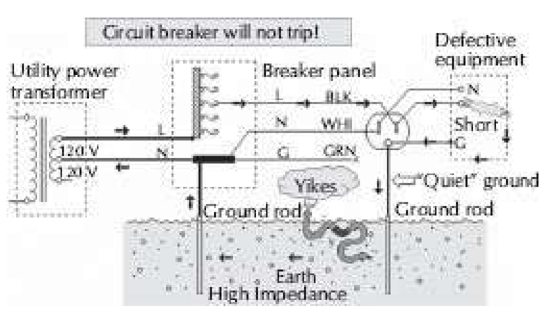


Figure 36-16. Fault protection is not provided by an earth ground connection!

When trying to track down and correct system noise problems, it easy to assume that power outlets are wired correctly. Low-cost outlet testers, which generally cost less than $10.00, will find dangerous problems such as hot-neutral or hot-ground reversals and open connections. Because they check for correct voltages between the pins, and both neutral and ground are normally at 0V, they cannot detect a neutral-ground reversal. This insidious wiring error can create nightmarish noise problems in an audio system. Finding the error by visual inspection of outlets is one possibility, but this could get labor intensive if the number of outlets is large. For large systems, and even those that can't be powered down, a sensitive, noncontact, clamp-on current probe can help identify the forks in the road when troubleshooting.[14] Code requires that neutral and safety ground be bonded only at the power service disconnecting means that is generally at the main breaker panel. Serious system noise problems can also occur when an extraneous neutral-to-ground connection exists elsewhere in the building wiring. A special test procedure can be used to determine this condition.[15]

NEVER, NEVER use devices such as three-prong-to two-prong ac plug adapters—a.k.a. ground lifters—to solve a noise problem! Such an adapter is intended to provide a safety ground (via the cover plate screw to a grounded saddle outlet and J-box) in cases where three-prong plugs must be connected to two-prong receptacles in pre-1960 buildings, Fig. 36-17.

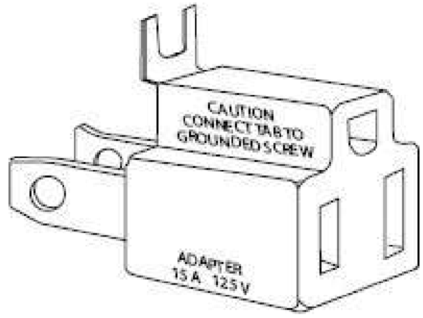


Figure 36-17. This is intended to provide a safety ground.

Consider two devices with grounding ac plugs that are connected by a signal cable. One device has a ground lifter on its plug and the other doesn't. If a fault occurs in the lifted device, the fault current flows through the signal cable to get to the grounded device. It's very likely that the cable will melt and catch fire!

Also consider that consumer audio and video equipment is responsible for about ten electrocutions every year in the United States. In a typical year, this equipment causes some 2000 residential fires that result in 100 civilian injuries, 20 deaths, and over $30 million in property losses, Fig. 36-18.[16,17]

Figure 36-18. Interconnect cables can carry lethal voltages throughout a system if just one ground lifted device fails.

Some small appliances, power tools, and consumer electronics are supplied with two-prong (ungrounded) ac plugs. Sometimes called *double insulated*, these devices are specially designed to meet strict UL and other requirements to remain safe even if one of their two insulation systems fails. Often there is a one-shot thermal cutoff switch inside the power transformer or motor windings to prevent overheating and subsequent insulation breakdown. *Only devices that carry a UL-listed label and originally supplied with ungrounded ac plugs should ever be operated without safety grounding*. Devices originally supplied with grounding three-prong plugs must always be operated with the safety ground properly connected!

36.4.3 Signal Grounding and EMC

EMC stands for electromagnetic compatibility, which is a field concerned with interference from electronic devices and their susceptibility to the interference created by other devices. As the world becomes increasingly wireless and digital, the general electromagnetic environment is becoming increasingly hostile. Engineers working in other disciplines, most notably information technology or IT—where signal/data frequencies are very high and narrowband—tend to minimize our difficulties in making audio systems robust against hostile electrical environments. In fact, high-quality audio systems are unique among electronic systems in two ways:

1. The signals cover a very broad, nearly 5 decade, range of frequencies.
2. The signals can require a very wide, currently over 120dB, dynamic range.

Adding to the difficulty is the fact that ac power frequencies and their harmonics also fall within the system's working frequency range. As you might suspect, grounding plays a pivotal role in controlling both emissions and susceptibility in both electronic devices and systems. In general, the same principles and techniques that reduce emissions will also reduce susceptibility. Grounding schemes generally fall into one of three categories:

1. Single point or star grounding.
2. Multipoint or mesh grounding.
3. Frequency selective transitional or hybrid grounding.

At frequencies below about 1MHz (which includes audio), virtually all experts agree that star grounding works best because system wiring is electrically short compared to the wavelengths involved. At these low frequencies, the dominant noise coupling problems arise from the simple lumped parameter behavior of wiring and electronic components. This includes the resistance and inductance of wires, the noise currents resulting from capacitances between utility power and system grounds, and magnetic and capacitive coupling effects.

On the other hand, at higher frequencies, system wiring can become electrically long and transmission line effects, such as standing waves and resonances, become the dominant problems. For example, because a 25ft (7.5m) audio cable is a quarter wavelength long at 10MHz, it becomes an antenna. At frequencies of 100MHz or higher, even a 12in (30cm) wire can no longer be considered low impedance path. To be effective at these frequencies, therefore, grounding schemes must emulate a flat metal sheet having extremely low inductance called a *ground plane*.

In practice, this can usually only be approximated with a multipoint ground system using wires. The wire lengths between points must remain well under a quarter-wavelength so, as frequency increases, larger numbers of increasingly shorter wires must be used to create the mesh. Ultimately, only a real ground plane can produce low-impedance ground connections at very high frequencies. Even a ground plane is not a perfect or equipotential ground, i.e., zero volts at all points. Because it has finite resistance, significant voltage differences can be developed between connection points to it.[18] Therefore, it should come as no surprise that IT and RF engineers prefer mesh grounding techniques while audio engineers prefer star grounding techniques.

At power and audio frequencies, a so-called *ground loop* allows noise and signal currents to mix in a common wire. Single-point grounding avoids this by steering signal currents and noise currents in independent paths. But at ultrasonic and radio frequencies, noise currents tend to bypass wires because they look like inductors and tend to flow instead in unintended paths consisting of parasitic capacitances. This makes star grounding essentially useless in controlling high-frequency interference in practical systems. Mesh grounding does a better job of controlling high-frequency interference, but since many ground loops are formed, low-frequency noise can easily contaminate signals. For audio systems, sometimes even inside audio equipment, there is clearly a conflict.

This conflict can be resolved by the hybrid grounding scheme. Capacitors can be used to create multiple high-frequency ground connections while allowing audio-frequency currents to take a path determined by the directly wired connection. Thus, the ground system behaves as a star system at low frequencies and a mesh system at high frequencies.[19] This technique of combining ground plane and star grounding is quite practical at the physical dimensions of a circuit board or an entire piece of equipment. At the system level the same conflict exists regarding grounding of audio cable shields. Ideally, at low frequencies, a shield should be grounded at one end only, but for maximum immunity to RF interference it should be grounded at both ends (and even intermediate points, if possible). This situation can be resolved by grounding one end directly and the other end through a small capacitor.[20] The shield grounding issue will be discussed further in section 36.5.2.

36.4.4 Grounding and System Noise

Most real-world systems consist of at least two devices that are powered by utility ac power. These power line connections unavoidably cause significant currents to flow in ground conductors and signal interconnect cables throughout a system. Properly wired, fully Code-compliant premises ac wiring generates small ground voltage differences and leakage currents. They are harmless from a safety viewpoint but potentially disastrous from a system noise viewpoint. Some engineers have a strong urge to reduce these unwanted voltage differences by shorting them out with a large conductor. The results are most often disappointing.[21] Other engineers think that system noise can be improved experimentally by simply finding a better or quieter ground. They hold a fanciful notion that noise current can somehow be skillfully directed to an earth ground, where it will disappear forever![22] In reality, since the earth has resistance just like any other conductor, earth ground connections are not at zero volts with respect to each other or any other mystical or absolute reference point.

36.4.4.1 Power Line Noise

The power line normally consists of a broad spectrum of harmonics and noise in addition to the pure 60Hz sine wave voltage. The noise is created by power supplies in electronic equipment, fluorescent lights, light dimmers, and intermittent or sparking loads such as switches, relays, or brush-type motors (i.e., blenders, shavers, etc.). Any device that causes a sudden change in load current, such as the notorious light dimmer, will generate high-frequency noise. Even an ordinary light switch will briefly arc internally as it is switched off and its contacts open, generating a burst of noise containing significant energy to at least 1MHz that is launched into the power wiring. The wiring behaves like a complex set of misterminated transmission lines gone berserk, causing the energy to reflect back and forth throughout the premises wiring until it is eventually absorbed or radiated. Power line noise can couple into signal paths in several ways, usually depending on whether the equipment uses two-prong or three-prong (grounding) ac power connections.

36.4.4.2 Parasitic Capacitances and Leakage Current Noise

In every ac-powered device, parasitic capacitances (never shown in schematic diagrams!) always exist between the power line and the internal circuit ground and/or chassis because of the unavoidable interwinding capacitances of power transformers and other line connected components. Especially if the device contains anything digital, there may also be intentional capacitances in the form of power line interference filters. These capacitances cause small but significant 60Hz leakage currents to flow between power line and chassis or circuit ground in each device. Because the coupling is capacitive, current flow increases at higher noise frequencies. Fig. 36-19 shows the frequency spectrum of current flow in 3nF of capacitance connected between line and safety ground at an ac outlet in a typical office.

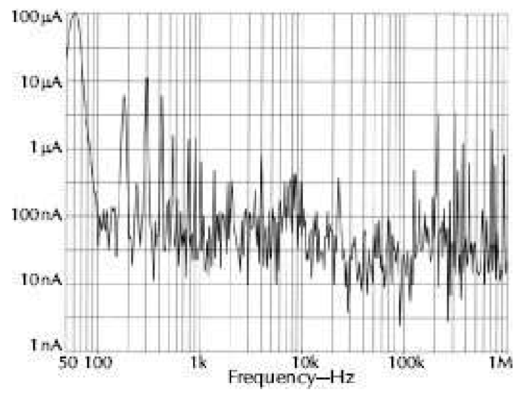


Figure 36-19. Typical leakage current from line to safety ground coupled via 3000µF capacitance into a 75Ω spectrum analyzer input.

This tiny current, although it poses no shock hazard, causes hum, buzz, pops, clicks, and other symptoms when it couples into the audio signal path. This capacitive coupling favors higher frequencies, making buzz a more common symptom than pure hum. We must accept noisy leakage currents as a fact of life.

36.4.4.3 Parasitic Transformers and Inter-Outlet Ground Voltage Noise

Substantial voltages are magnetically induced in premises safety ground wiring when load current flows in the circuit conductors as shown in Fig. 36-20. The magnetic fields that surround the line and neutral conductors, which carry load current, magnetically induce a small voltage over the length of the safety ground conductor, effectively forming a parasitic transformer. The closer the safety ground conductor is to either the line or neutral conductor, the higher the induced voltage. Because, at any instant in time, line and neutral currents are equal but flow in opposite directions, there is a plane of zero magnetic field exactly midway between the line and neutral conductors as shown in Fig. 36-21. Therefore, Romex® and similar bonded cables generally generate significantly lower induced voltages than individual wires in conduit, where the relative positioning of the wires is uncontrolled.

Fig. 36-22 shows the results of an experiment that was the subject of an AES paper.[23] The experiments had two goals:

1. Create a carefully-controlled "reference" conductor geometry to confirm the predictions of theory.
2. Compare several widely-used conductor configurations used in premises ac power wiring over a frequency range that simulate harmonic currents drawn by equipment power supplies. The experiment also tested a little-known and somewhat radical idea to further reduce ground voltage induction or GVI in the "conduit transformer."

The voltage induced in this transformer is directly proportional to the rate of change of load current in the circuit. This results in the rising 20dB/decade slope of all the plots. The reason ordinary phase-control light dimmers create so much interference is that they switch current very quickly. Fig. 23A shows the current in a circuit consisting of an inexpensive light dimmer driving six 100-watt incandescent lamps and set for 50% brightness (the worst case for interference). Fig. 23B shows the rise-time on an expanded time scale—the transition from zero to 8A takes only about 5μs and contains substantial energy up to 70kHz. Since the magnetic induction into safety ground favors high frequencies, noise susceptibility problems in a system will likely become most evident when a light dimmer is involved.

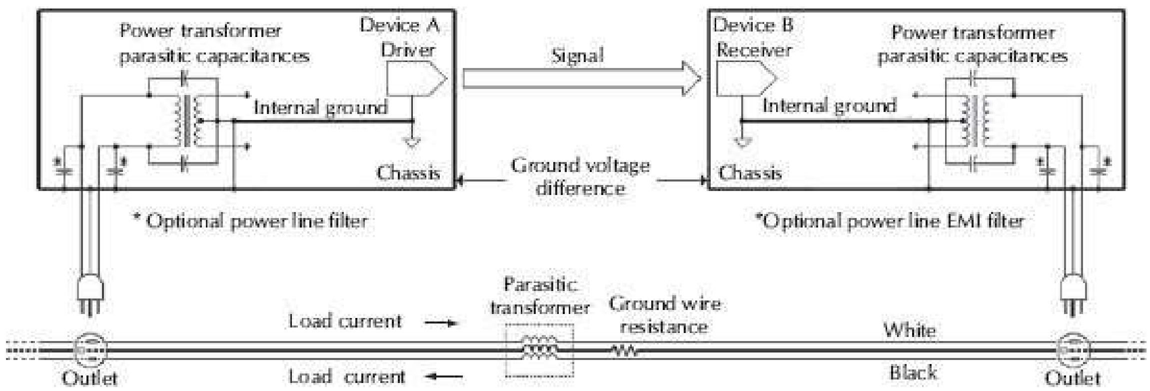


Figure 36-20. Voltage difference is magnetically induced over length of safety-ground premises wiring.

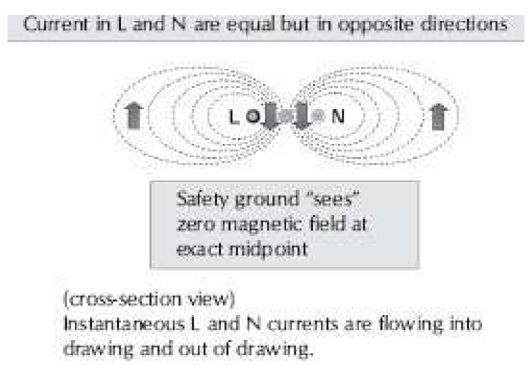


Figure 36-21. Magnetic fields surrounding line and neutral can induce voltage into safety ground.

The voltage induced in any transformer is directly proportional to the rate of change of load current in the circuit. With an ordinary phase-control light dimmer the peak voltages induced can become quite high. When the dimmer triggers current on 120 times per second, it switches on very quickly (a few microseconds) as shown in Fig. 36-23. Since the magnetic induction into safety ground favors high frequencies, noise coupling problems in a system will likely become most evident when a light dimmer is involved. The problems are usually worst at about half-brightness setting of the dimmer.

This parasitic transformer action generates small ground voltage differences, generally under 1V, between ac outlets. The voltage differences tend to be higher between two outlets on different branch circuits, and higher still if a device on the branch circuit is also connected to a remote or alien ground such as a CATV feed, satellite dish, or an interbuilding tie line. *We must accept interoutlet ground noise voltage as a fact of life.*

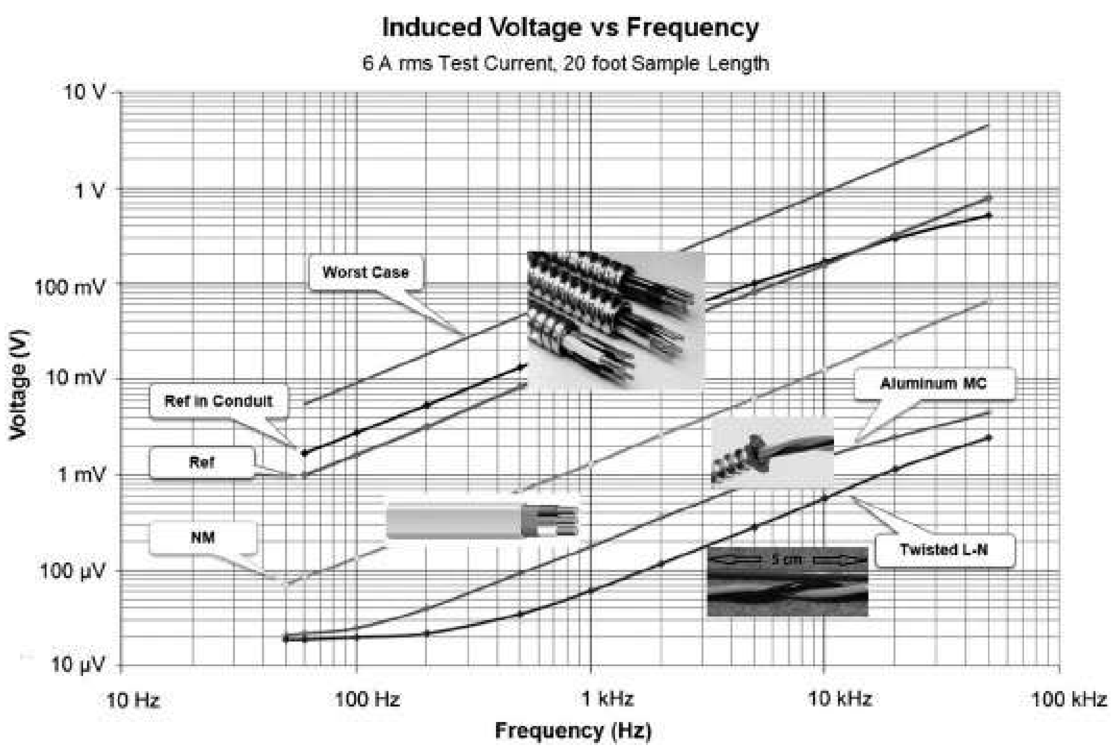


Figure 36-22. Gound voltage induction test results for various conductor configurations in premises ac power wiring.

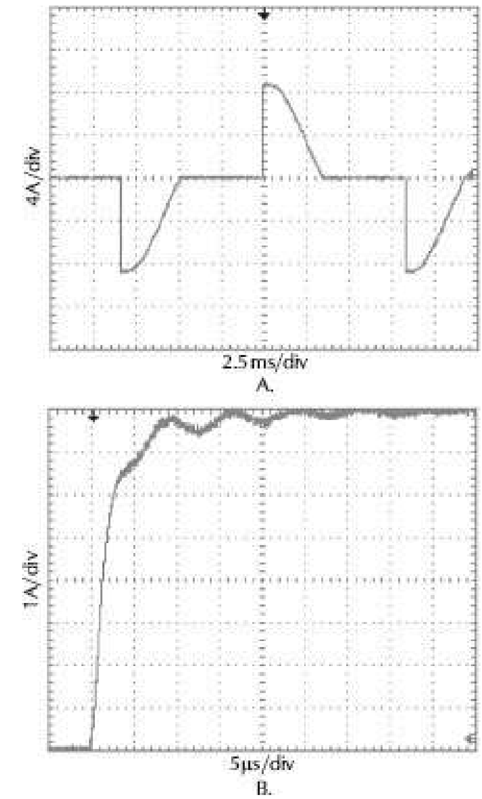


Figure 36-23. A. Phase controlled light dimmer current. B. Rise time expanded to show switching speed.

36.4.4.4 Ground Loops

For our purposes, a ground loop is formed when a signal cable connects two pieces of equipment whose connections to the power line or other equipment causes a power-line-derived current to flow in the signal cable.

The first, and usually worst, kind of ground loop occurs between *grounded* devices—those with three-prong ac plugs. Current flow in signal cables, as shown in Fig. 36-24, can easily reach 100mA or more.

The second kind of ground loop occurs between floating devices—those with two-prong ac plugs. Each pair of capacitances CF (for EMI filter) and CP (for power transformer parasitic) in the schematic form a capacitive voltage divider between line and neutral, causing some fraction of 120Vac to appear between chassis and ground. For UL-listed ungrounded equipment, this leakage current must be under 0.75mA (0.5mA for office equipment). This small current can cause an unpleasant, but harmless, tingling sensation as it flows through a person's body. More relevant is the fact that these noisy leakage currents will flow in any wire connecting such a floating device to safety ground, or connecting two floating devices to each other as shown in Fig. 36-25.

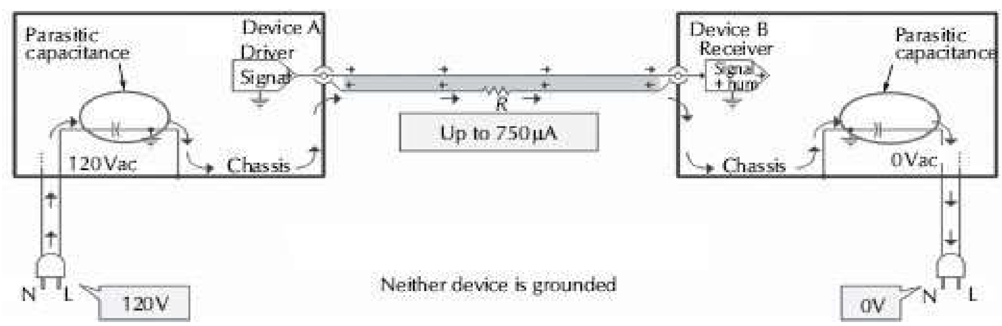


Figure 36-25. For ungrounded equipment, interconnect cables complete a capacitive loop.

36.5 Interface Problems in Systems

If properly designed balanced interfaces were used throughout an audio system, it would theoretically be noise-free. Until about 1970, equipment designs allowed real-world system to come very close to this ideal. But since then, balanced interfaces have fallen victim to two major design problems—and both can properly be blamed on equipment manufacturers. Even careful examination of manufacturers' specifications and data sheets will not reveal either problem—the devil is in the details. These problems are effectively concealed because the marketing departments of most manufacturers have succeeded in dumbing down their so-called specifications over the same time period.

First is degraded noise rejection, which appeared when solid-state differential amplifiers started replacing input transformers. Second is the pin 1 problem that appeared in large numbers when PC boards and plastic connectors replaced their metal counterparts. Both problems can be avoided through proper design, of course, but in this author’s opinion, part of the problem is that the number of analog design engineers who truly understand the underlying issues is dwindling and engineering schools are steering most students into the digital future where analog issues are largely neglected. Other less serious problems with balanced interfaces are caused by balanced cable construction and choices of cable shield connections.

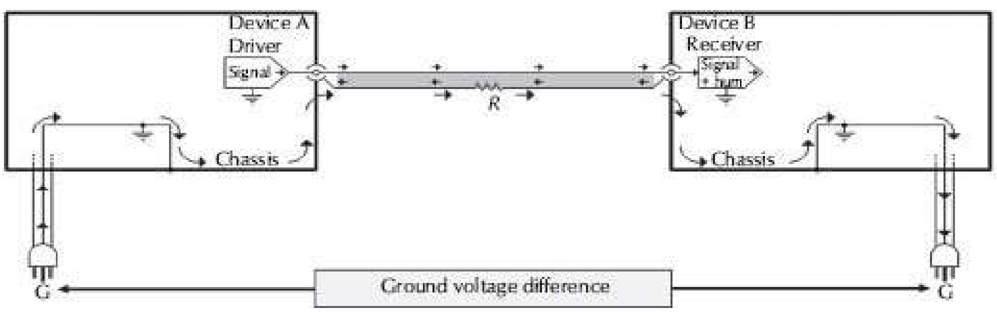


Figure 36-24. For grounded equipment, interconnect cables complete a wired loop.

On the other hand, unbalanced interfaces have an intrinsic problem that effectively limits their use to only the most electrically benign environments. Of course, even this problem can be solved by adding external ground-isolation devices, but the best advice is to avoid unbalanced interfaces whenever possible in professional systems!

36.5.1 Degraded Common-Mode Rejection

Balanced interfaces have traditionally been the hallmark of professional sound equipment. In theory, systems comprised of such equipment are completely noise-free. However, an often overlooked fact is that the common-mode rejection of a complete signal interface does not depend solely on the receiver, but on how the receiver interacts with the driver and line—performing together as a subsystem.

In the basic balanced interface of Fig. 36-26, the output impedances of the driver $Z_o/2$ and the input impedances of the receiver Z_{cm} effectively form the Wheatstone bridge shown in Fig. 36-27. If the bridge is not balanced or nulled, a portion of the ground noise V_{cm} will be converted to a differential signal on the line. This nulling of the common-mode voltage is critically dependent on the ratio matching of the pairs of driver/receiver common-mode impedances R_{cm} in the – and + circuit branches. The balancing or nulling is unaffected by impedance across the two lines, such as the signal input impedance Z_i in Fig. 36-8 or the signal output impedance of the driver. It is the common-mode impedances that matter!

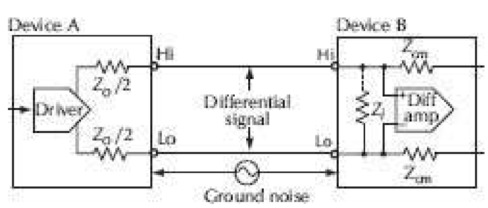


Figure 36-26. Simplified balanced interface.

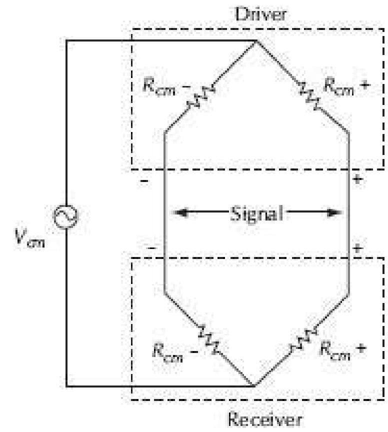


Figure 36-27. The balanced interface is a Wheatstone bridge.

The bridge is most sensitive to small fractional impedance changes in one of its arms when all arms have the same impedance.[24] It is least sensitive when the upper and lower arms have widely differing impedances, e.g., when upper arms are very low and lower arms are very high, or vice versa. Therefore, we can minimize the sensitivity of a balanced system (bridge) to impedance imbalances by making common-mode impedances very low at one end of the line and very high at the other. This condition is consistent with the requirements for voltage matching discussed in section 36.3.2.

Most active line receivers, including the basic differential amplifier of Fig. 36-28, have common-mode input impedances in the 5kΩ to 50kΩ range, which is inadequate to maintain high CMRR with real-world sources. With common-mode input impedances of 5kΩ, a source imbalance of only 1Ω, which could arise from normal contact and wire resistance variations, can degrade CMRR by 50dB. Under the same conditions, the CMRR of a good input transformer would be unaffected because of its 50MΩ common-mode input impedances. Fig. 36-29 shows computed CMRR versus source imbalance for different receiver common-mode input impedances. Thermal noise and other limitations place a practical limit of about 130dB on most actual CMRR measurements.

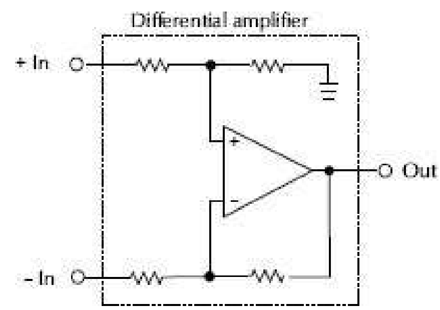


Figure 36-28. Basic differential amplifier.

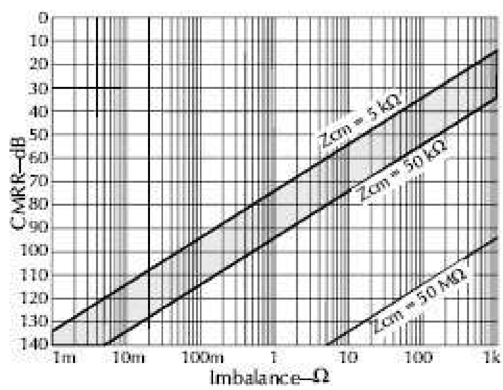


Figure 36-29. Noise rejection versus source impedance/imbalance.

How much imbalance is there in real-world signal sources? Internal resistors and capacitors determine the output impedance of a driver. In typical equipment, $Z_0/2$ may range from 25 to 300Ω. Since the resistors are commonly ±5% tolerance and the coupling capacitors are ±20% at best, impedance imbalances up to about 20Ω should be routinely expected. This defines a real-world source. In a previous paper, this author has examined balanced audio interfaces in some detail, including performance comparisons of various receiver types.[25] It was concluded that, regardless of their circuit topology, popular active receivers can have very poor CMRR when driven from such real-world sources. The poor performance of these receivers is a direct result of their low common-mode input impedances. If common-mode input impedances are raised to about 50MΩ, 94dB of ground noise rejection is attained from a completely unbalanced 1kΩ source, which is typical of consumer outputs. When common-mode input impedances are sufficiently high, an input can be considered truly universal, suitable for any source—balanced or unbalanced. A receiver using either a good input transformer or the InGenius® integrated circuit[26] will routinely achieve 90–100dB of CMRR and remain unaffected by typical real-world output imbalances.

The theory underlying balanced interfaces is widely misunderstood by audio equipment designers. Pervasive use of the simple differential amplifier as a balanced line receiver is evidence of this. And, as if this weren't bad enough, some have attempted to improve it. Measuring input X and Y input impedances of the simple differential amplifier individually leads some designers to alter its equal resistor values. However, as shown in Fig. 36-30, if the impedances are properly measured simultaneously, it becomes clear that nothing is wrong. The fix grossly unbalances the common-mode impedances, which destroys the interface CMRR for any real-world source. This and other misguided *improvements* completely ignore the importance of common-mode input impedances.

The same misconceptions have also led to some CMRR tests whose results give little or no indication of how the tested device will actually behave in a real-world system. Apparently, large numbers of designers test the CMRR of receivers with the inputs either shorted to each other or driven by a laboratory precision signal source. The test result is both unrealistic and misleading. Inputs rated at 80dB of CMRR could easily deliver as little as 20dB or 30dB when used in a real system. Regarding their previous test, the IEC had recognized *that test is not an adequate assurance of the performance of certain electronically balanced amplifier input circuits*. The old method simply didn't account for the fact that source impedances are rarely perfectly balanced. To correct this, this author was instrumental in revising IEC Standard: 60268-3 Sound System Equipment – Part 3: Amplifiers. The new method, as shown in Fig. 36-31, uses typical ±10Ω source impedance imbalances and clearly reveals the superiority of input transformers and some new active input stages that imitate them. The new standard was published August 30, 2000. The Audio Precision APx520 and APx525, introduced in 2008, are the first audio instruments to offer the new CMRR test.

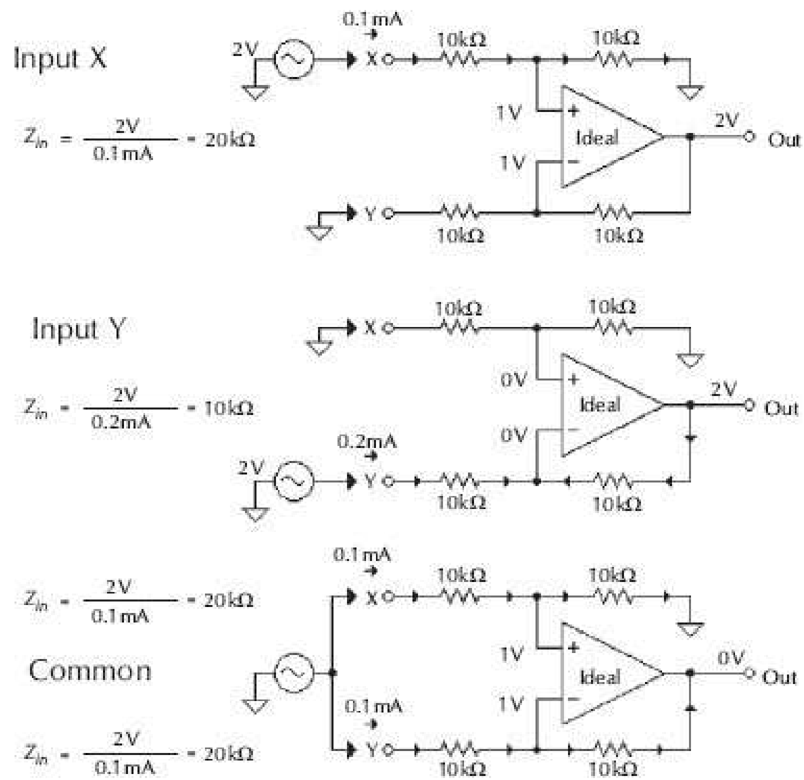


Figure 36-30. Common-mode impedances apply to a voltage applied to both inputs simultaneously.

36.5.2 The Pin 1 Problem

In his now famous paper in the 1995 *AES Journal*, the late Neil Muncy says:

> *This paper specifically addresses the problem of noise coupling into balanced line-level signal interfaces used in many professional applications, due to the unappreciated consequences of a popular and widespread audio equipment design practice which is virtually without precedent in any other field of electronic systems.*[27]

Common impedance coupling occurs whenever two currents flow in a shared or common impedance. A noise coupling problem is created when one of the currents is ground noise and the other is signal. The common impedance is usually a wire or circuit board trace having a very low impedance, usually well under an ohm. Unfortunately, common impedance coupling has been designed into audio equipment from many manufacturers. The noise current enters the equipment via a terminal, at a device input or output, to which the cable shield is connected via a mating connector. For XLR connectors, it's pin 1 (hence the name); for ¼ inch connectors, it's the sleeve; and for RCA/IHF connectors, it's the shell.

To the user, symptoms are indistinguishable from many other noise coupling problems such as poor CMRR. To quote Neil again,

> *Balancing is thus acquiring a tarnished reputation, which it does not deserve. This is indeed a curious situation. Balanced line-level interconnections are supposed to ensure noise-free system performance, but often they do not.*

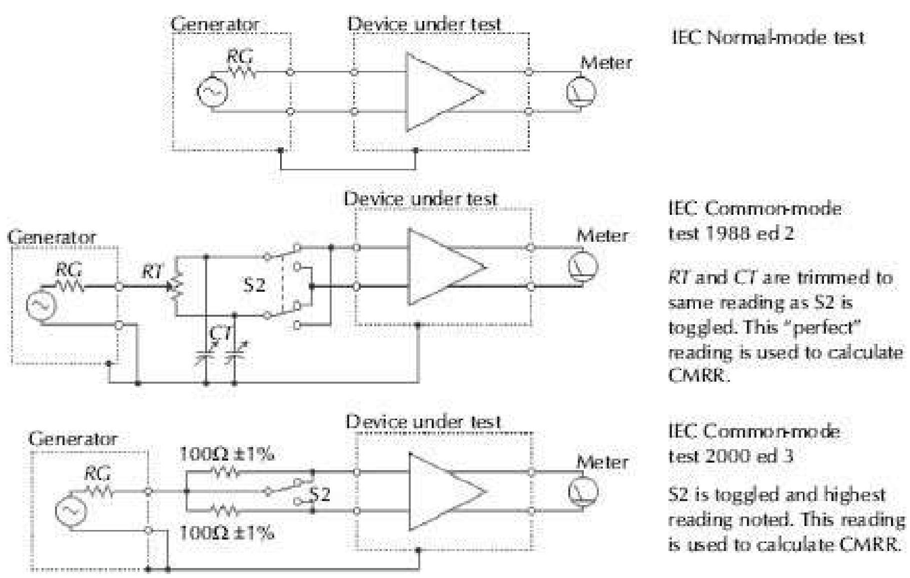


Figure 36-31. Old and new IEC tests for CMRR compared.

In balanced interconnections, it occurs at line inputs and outputs where interconnecting cables routinely have their shields grounded at both ends. Of course, grounding at both ends is required for unbalanced interfaces.

Fig. 36-32 illustrates several examples of common impedance coupling. When noise currents flow in signal reference wiring or circuit board traces, tiny voltage drops are created. These voltages can couple into the signal path, often into very high gain circuitry, producing hum or other noise at the output. In the first two devices, pin 1 current is allowed to flow in internal signal reference wiring. In the second and third devices, power line noise current (coupled through the parasitic capacitances in the power transformer) is also allowed to flow in signal reference wiring to reach the chassis/safety ground. This so-called *sensitive equipment* will produce additional noise independent of the pin 1 problem. For the second device, even disconnecting its safety ground (not recommended) won't stop current flow through it between input and output pin 1 shield connections.

Fig. 36-33 shows three devices whose design does not allow shield current to flow in signal reference conductors. The first uses a star connection of input pin 1, output pin 1, power cord safety ground, and power supply common. This technique is the most effective prevention. Noise currents still flow, but not through internal signal reference conductors. Before there were printed circuit boards, a metal chassis served as a very low-impedance connection (effectively a ground plane) connecting all pins 1 to each other and to safety ground. Pin 1 problems were virtually unknown in those vintage designs. Modern printed circuit board–mounted connectors demand that proper attention be paid to the routes taken by ground noise currents. Of course, this same kind of problem can and does exist with RCA connectors in unbalanced consumer equipment, too.

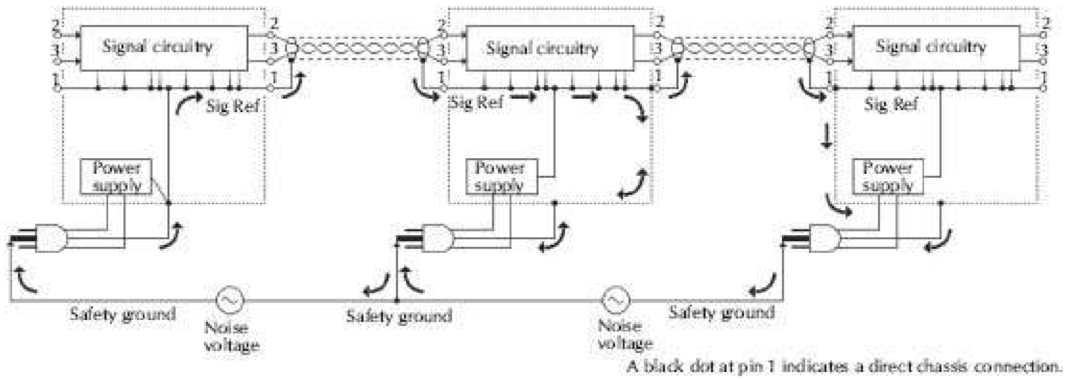


Figure 36-32. How poor routing of shield currents produces the pin 1 problem.

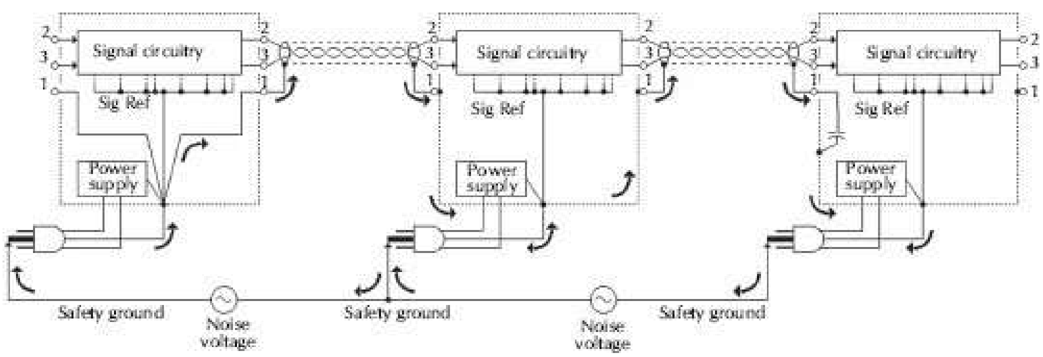


Figure 36-33. Equipment with proper internal routing of shield currents.

Fortunately, tests to reveal such common impedance coupling problems are not complex. Comprehensive tests using lab equipment covering a wide frequency range have been described by Cal Perkins[32] and simple tests using an inexpensively built tester called the *hummer* have been described by John Windt.[33] Jensen Transformers, Inc. variant of the Hummer is shown in Fig. 36-34. It passes a rectified ac current of 60–80mA through the potentially troublesome shield connections in the device under test to determine if they cause the coupling.

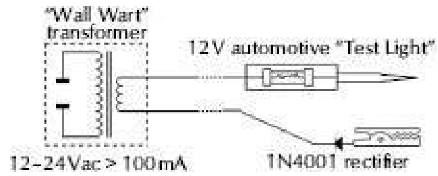


Figure 36-34. The Hummer II. Courtesy Jensen Transformers, Inc.

The glow of the automotive test lamp shows that a good connection has been made and that test current is indeed flowing. The procedure:

1. Disconnect all input and output cables, except the output to be monitored, as well as any chassis connections (due to rack mounting, for example) from the device under test.
2. Power up the device.
3. Meter and, if possible, listen to the device output. Hopefully, the output will simply be random noise. Try various settings of operator controls to familiarize yourself with the noise characteristics of the device under test without the hummer connected.
4. Connect the hummer clip lead to the device chassis and touch the probe tip to pin 1 of each input or output connector. If the device is properly designed, there will be no output hum or change in the noise floor.
5. Test other potentially troublesome paths, such as from an input pin 1 to an output pin 1 or from the safety ground pin of the power cord to the chassis (a three-to-two-prong ac adapter is handy to make this connection).

Note: Pin 1 might not be connected directly to ground in some equipment—hopefully, this will be at inputs only! In this case, the hummer's lamp may not glow—this is OK.

36.5.3 Balanced Cable Issues

At audio frequencies, even up to about 1MHz, cable shields should be grounded at one end only, where the signal is ground referenced. At higher frequencies, where typical system cables become a small fraction of a wavelength, it's necessary to ground it at more than one point to keep it at ground potential and guard against RF interference.[28,29] Based on my own work, there are two additional reasons that there should always be a shield ground at the driver end of the cable, whether the receiver end is grounded or not, see Figs. 36-35 and 36-36. The first reason involves the cable capacitances between each signal conductor and shield, which are mismatched by 4% in typical cable. If the shield is grounded at the receiver end, these capacitances and driver common-mode output impedances, often mismatched by 5% or more, form a pair of low-pass filters for common-mode noise. The mismatch in the filters converts a portion of common-mode noise to differential signal. If the shield is connected only at the driver, this mechanism does not exist. The second reason involves the same capacitances working in concert with signal asymmetry. If signals were perfectly symmetrical and capacitances perfectly matched, the capacitively coupled signal current in the shield would be zero through cancellation. Imperfect symmetry and/or capacitances will cause signal current in the shield. This current should be returned directly to the driver from which it came. If the shield is grounded at the receiver, all or part of this current will return via an undefined path that can induce crosstalk, distortion, or oscillation.[30]

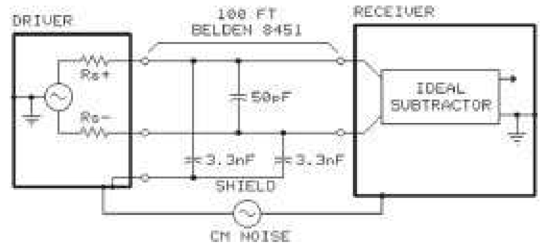


Figure 36-35. Shield grounded only at driver—Good.

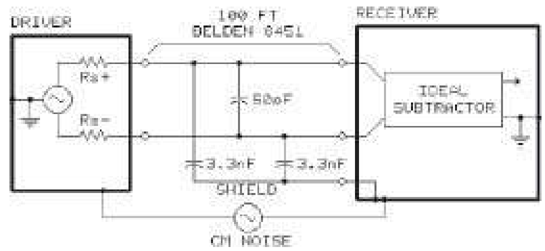


Figure 36-36. Shield grounded only at receiver—Risky.

With cables, too, there is a conflict between the star and mesh grounding methods as discussed in section 36.4.3. But this low-frequency versus high-frequency conflict can be substantially resolved with a hybrid approach involving grounding the receive end of cables through an appropriate capacitance (shown in the third device of Fig. 36-33).[28,29] Capacitor values in the range of 10nF to 100nF are most appropriate for the purpose. Such capacitance has been integrated into the Neutrik EMC series connectors. The merits of this scheme were the subject of several years of debate in the Audio Engineering Society Standards Committee working group that developed AES48.

As discussed in section 36.2.5, twisting essentially places each conductor at the same average distance from the source of a magnetic field and greatly reduces differential pickup. Star quad cable reduces pickup even further, typically by about 40dB. But the downside is that its capacitance is approximately double that of standard shielded twisted pair.

SCIN, or shield-current induced noise, may be one consequence of connecting a shield at both ends. Think of a shielded twisted pair as a transformer with the shield acting as primary and each inner conductor acting as a secondary winding, as shown in the cable model of Fig. 36-37. Current flow in the shield produces a magnetic field which then induces a voltage in each of the inner conductors. If these voltages are identical, and the interface is properly impedance balanced, only a common-mode voltage is produced that can be rejected by the line receiver. However, subtle variations in physical construction of the cable can produce unequal coupling in the two signal conductors. The difference voltage, since it appears as signal to the receiver, results in noise coupling. Test results on six commercial cable types appear in reference.[31] In general, braided shields perform better than foil shields and drain wires.

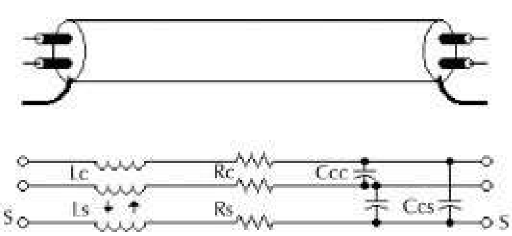


Figure 36-37. Shield of a shielded twisted pair cable is magnetically coupled to inner conductors.

And, to make matters even worse, grounding the shield of balanced interconnect cables at both ends also excites the pin 1 problem if it exists. Although it might appear that there's little to recommend grounding at both ends, it is a widely accepted practice. As you can see, noise rejection in a real-world balanced interface can be degraded by a number of subtle problems and imperfections. But, as discussed in section 36.6.4, it is virtually always superior to an unbalanced interface!

36.5.4 Coupling in Unbalanced Cables

The overwhelming majority of consumer as well as high-end audiophile equipment still uses an audio interface system introduced over 60 years ago and intended to carry signals from chassis to chassis inside the earliest RCA TV receivers! The ubiquitous RCA cable and connector form an unbalanced interface that is extremely susceptible to common impedance noise coupling.

As shown in Fig. 36-38, noise current flow between the two device grounds or chassis is through the shield conductor of the cable. This causes a small but significant noise voltage to appear across the length of the cable. Because the interface is unbalanced, this noise voltage will be directly added to the signal at the receiver.[34] In this case, the impedance of the shield conductor is responsible for the *common impedance coupling*. This coupling causes hum, buzz, and other noises in audio systems. It's also responsible for slow-moving hum bars in video interfaces and glitches, lock-ups, or crashes in unbalanced, e.g., RS-232, data interfaces.

Consider a 25ft interconnect cable with foil shield and a #26AWG drain wire. From standard wire tables or actual measurement, its shield resistance is found to be 1.0Ω. If the 60Hz leakage current is 300μA, the hum voltage will be 300 μV. Since the consumer audio reference level is about −10dBV or 300mV, the 60Hz hum will be only 20log(300μV/300mV) = −60dB relative to the signal. For most systems, this is a very poor signal-to-noise ratio! For equipment with two-prong plugs, the 60Hz harmonics and other high-frequency power-line noise (refer to Fig. 36-19) will be capacitively coupled and result in a harmonic-rich buzz.

Because the output impedance of device A and the input impedance of device B are in series with the inner conductor of the cable, its impedance has an insignificant effect on the coupling and is not represented here. Common-impedance coupling can become extremely severe between two grounded devices, since the voltage drop in the safety ground wiring between the two devices is effectively parallel connected across the length of the cable shield. This generally results in a fundamental-rich hum that may actually be larger than the reference signal!

Coaxial cables, which include the vast majority of unbalanced audio cables, have an interesting and underappreciated quality regarding common-impedance coupling at high frequencies, Fig. 36-39. Any voltage appearing across the ends of the shield will divide itself between shield inductance L_S and resistance R_S according to frequency. At some frequency, the voltages across each will be equal (when reactance of L_S equals R_S). For typical cables, this frequency is in the 2 to 5kHz range. At frequencies below this transition frequency, most of the ground noise will appear across R_S and be coupled into the audio signal as explained earlier. However, at frequencies above the transition frequency, most of the ground noise will appear across L_S. Since L_S is magnetically coupled to the inner conductor, a replica of the ground noise is induced over its length. This induced voltage is then subtracted from the signal on the inner conductor, reducing noise coupling into the signal. At frequencies ten times the transition frequency, there is virtually no noise coupling at all—common-impedance coupling has disappeared. Therefore, common-impedance coupling in coaxial cables ceases to be a noise issue at frequencies over about 50kHz. Remember this as we discuss claims made for power line filters that typically remove noise only above about 50kHz.

Unbalanced interface cables, regardless of construction, are also susceptible to magnetically induced noise caused by nearby low-frequency ac magnetic fields. Unlike balanced interconnections, such noise pickup is not nullified by the receiver.

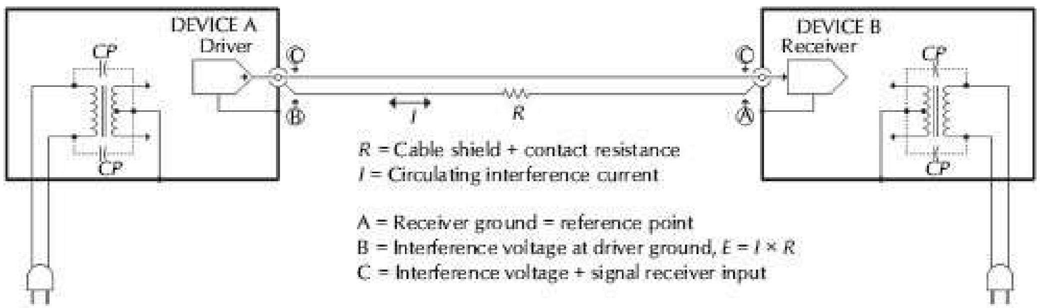


Figure 36-38. Common impedance coupling in an unbalanced audio, video, or data interface.

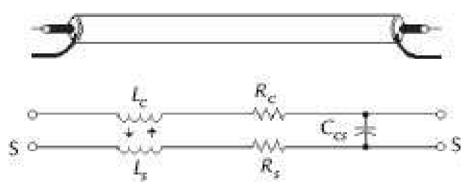


Figure 36-39. Magnetic coupling between shield and center conductor is 100%.

36.5.5 Bandwidth and RF Interference

RF interference isn't hard to find—it's actually very difficult to avoid, especially in urban areas. It can be radiated through the air and/or be conducted through any cables connected to equipment. Common sources of radiated RF include AM, shortwave, FM, and TV broadcasts; ham, CB, remote control, wireless phone, cellular phone, and a myriad of commercial two-way radio and radar transmitters; and medical and industrial RF devices. Devices that create electrical sparks, including welders, brush-type motors, relays, and switches can be potent wideband radiators. Less obvious sources include arcing or corona discharge from power line insulators (common in seashore areas or under humid conditions) or malfunctioning fluorescent, HID, or neon lights. Of course, lightning, the ultimate spark, is a well-known radiator of momentary interference to virtually anything electronic.

Interference can also be conducted via any wire coming into the building. Because power and telephone lines also behave as huge outdoor antennas, they are often teeming with AM radio signals and other interference. But the most troublesome sources are often inside the building and the energy delivered through ac power wiring. The offending source may be in the same room as your system or, worse yet, it may actually be a part of your system! The most common offenders are inexpensive light dimmers, fluorescent lights, CRT displays, digital signal processors, or any device using a switching power supply.

Although cable shielding is a first line of defense against RF interference, its effectiveness depends critically on the shield connection at each piece of equipment. Because substantial inductance is added to this connection by traditional XLR connectors and grounding pigtails, the shield becomes useless at high radio frequencies. Common-mode RF interference simply appears on all the input leads.[35] Because the wire limitations discussed in section 36.2.4 apply to grounding systems, contrary to widespread belief, grounding is not an effective way to deal with RF interference. To quote the late Neil Muncy:

> *Costly technical grounding schemes involving various and often bizarre combinations of massive copper conductors, earth electrodes, and other arcane hardware are installed. When these schemes fail to provide expected results, their*

proponents are usually at a loss to explain why.[36]

The wider you open the window, the more dirt flies in. One simple, but often overlooked, method of minimizing noise in a system is to limit the system bandwidth to that required by the signal.[37] In an ideal world, every signal-processing device in a system would contain a filter at each input and output connector to appropriately limit bandwidth and prevent out-of-band energy from ever reaching active circuitry. This RF energy becomes an audio noise problem because the RF is demodulated or detected by active circuitry in various ways, acting like a radio receiver that adds its output to the audio signal. Symptoms can range from actual reception of radio signals or a 59.94Hz buzz from TV signals or various tones from cell phone signals to much subtler distortions, often described as a veiled or grainy audio quality.[38] The filters necessary to prevent these problems vary widely in effectiveness and, in some equipment, may not be present at all. Sadly, the performance of most commercial equipment will degrade when such interference is coupled to its input.[39]

36.6 Solving Real-World System Problems

How much noise and interference are acceptable depends on what the system is and how it will be used. Obviously, sound systems in a recording studio need to be much more immune to noise and interference than paging systems for construction sites.

36.6.1 Noise Perspective

The decibel is widely used to express audio-related measurements. For *power* ratios,

$$dB = 10\log\frac{P_1}{P_2} \qquad (36\text{-}9)$$

For *voltage* or *current* ratios, because power is proportional to the square of voltage or current:

$$dB = 20\log\frac{E_1}{E_2}$$
$$dB = 20\log\frac{I_1}{I_2} \qquad (36\text{-}10)$$

Most listeners describe 10dB level decreases or increases as halving or doubling loudness, respectively, and 2dB or 3dB changes as just noticeable. Under laboratory conditions, well-trained listeners can usually identify level changes of 1dB or less. The dynamic range of an electronic system is the ratio of its maximum undistorted signal output to its residual noise output or noise floor. Up to 120dB of dynamic range may be required in high-end audiophile sound systems installed in typical homes.[40]

36.6.2 Troubleshooting

Under certain conditions, many systems will be acceptably noise-free in spite of poor grounding and interfacing techniques. People often get away with doing the wrong things! But, notwithstanding anecdotal evidence to the contrary, logic and physics will ultimately rule.

Troubleshooting noise problems can be a frustrating, time-consuming experience but the method described in section 36.6.2.2 can relieve the pain. It requires no electronic instruments and is very simple to perform. Even the underlying theory is not difficult. The tests will reveal not only what the coupling mechanism is but also where it is.

36.6.2.1 Observations, Clues, and Diagrams

A significant part of troubleshooting involves how you think about the problem. First, don't assume anything! For example, don't fall into the trap of thinking, just because you've done something a particular way many times before, it simply can't be the problem. Remember, even things that can't go wrong, do! Resist the temptation to engage in guesswork or use a shotgun approach. If you change more than one thing at a time, you may never know what actually fixed the problem.

Second, ask questions and gather clues! If you have enough clues, many problems will reveal themselves before you start testing. Be sure to write everything down—imperfect recall can waste a lot of time! Troubleshooting guru Bob Pease[41] suggested these basic questions:

1. Did it ever work right?
2. What are the symptoms that tell you it's not working right?
3. When did it start working badly or stop working?
4. What other symptoms showed up just before, just after, or at the same time as the failure?

Operation of the equipment controls, and some elementary logic, can provide very valuable clues. For example, if a noise is unaffected by the setting of a gain control or selector, logic dictates that it must be entering the signal path after that control. If the noise can be eliminated by turning the gain down or selecting another input, it must be entering the signal path before that control.

Third, sketch a block diagram of the system. Fig. 36-40 is an example diagram of a simple home theater system. Show all interconnecting cables and indicate approximate length. Mark any balanced inputs or outputs. Generally, stereo pairs can be indicated with a single line. Note any device that is grounded via a three-prong ac plug. Note any other ground connections such as equipment racks, cable TV connections, etc.

36.6.2.2 The Ground Dummy Procedure

An easily constructed adapter or ground dummy is the key element in this procedure. By temporarily placing the dummy at strategic locations in the interfaces, precise information about the nature and location of the problem is revealed. The tests can specifically identify:

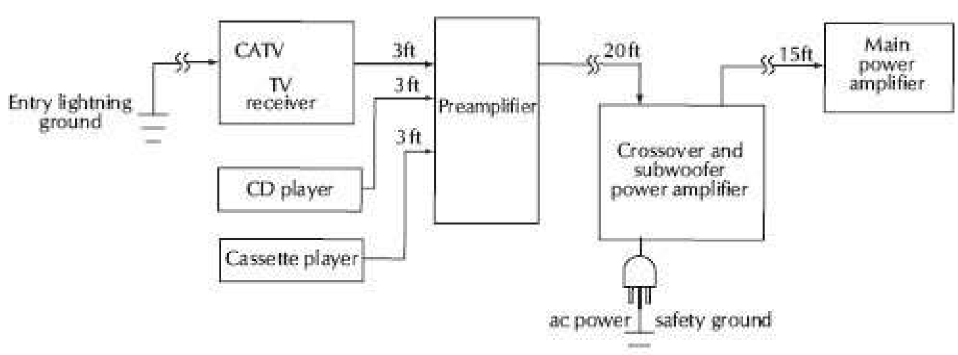


Figure 36-40. Block diagram of example system.

1. Common-impedance coupling in unbalanced cables.
2. Shield current-induced coupling in balanced cables.
3. Magnetic or electrostatic pickup of nearby magnetic or electrostatic fields.
4. Common-impedance coupling (the pin 1 problem) inside defective devices.

5. Inadequate CMRR of the balanced input.

The ground dummy can be made from standard connector wired as shown in Figs. 36-41 and 36-42. Since a dummy does not pass signal, mark it clearly to help prevent it, being accidentally left in a system.

Each signal interface is tested in four steps. As a general rule, always start at the inputs to the power amplifiers and work backward toward the signal sources. *Be very careful when performing the tests not to damage loudspeakers or ears!* The surest way to avoid possible damage is to turn off the power amplifier(s) before reconfiguring cables for each test step.

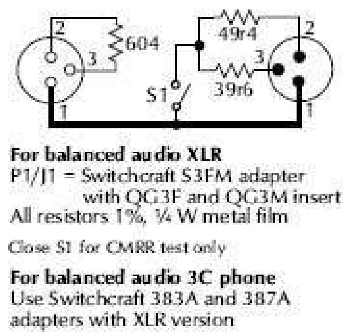


Figure 36-41. Balanced ground dummy.

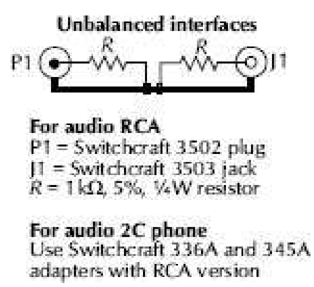
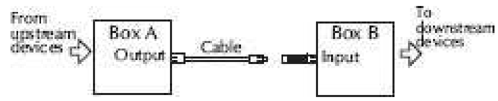


Figure 36-42. Unbalanced ground dummy.

36.6.2.2.1 For Unbalanced Interfaces

STEP 1: Unplug the cable from the input of Box B and plug in only the dummy as shown below.

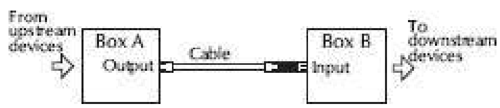


♦ Output quiet?

No—The problem is either in Box B or farther downstream.

Yes—Go to next step.

STEP 2: Leaving the dummy in place at the input of Box B, plug the cable into the dummy as shown below.

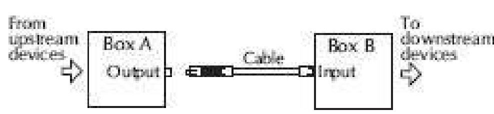


♦ Output quiet?

No—Box B has a pin 1 problem (see section 4.3 to confirm this).

Yes—Go to next step.

STEP 3: Remove the dummy and plug the cable directly into the input of Box B. Unplug the other end of the cable from the output of Box A and plug it into the dummy as shown below. Do not plug the dummy into Box A or let it touch anything conductive.

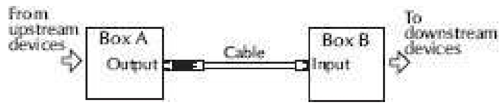


♦ Output quiet?

No–Noise is being induced in the cable itself. Reroute the cable to avoid interfering fields (see section 36.4.2 or 36.4.4).

Yes—Go to next step.

STEP 4: Leaving the dummy in place on the cable, plug the dummy into the output of Box A as shown below.



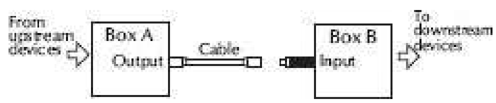
♦ Output quiet?

No—The problem is common-impedance coupling (see section 36.4.4). Install a ground isolator at the input of Box B.

Yes—The noise is coming from (or through) the output of Box A. Perform the same test sequence on the cable(s) connecting Box A to upstream devices.

36.6.2.2.2 For Balanced Interfaces

STEP 1: Unplug the cable from the input of Box B and plug in only the dummy (switch open or NORM) as shown below.

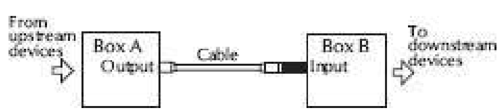


♦ Output quiet?

No—The problem is either in Box B or farther downstream.

Yes—Go to next step.

STEP 2: Leaving the dummy in place at the input of Box B, plug the cable into the dummy (switch open or NORM) as shown below.

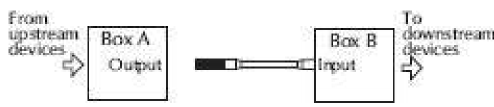


♦ Output quiet?

No—Box B has a Pin 1 problem (see hummer test, section 36.4.2, to confirm this).

Yes—Go to next step.

STEP 3: Remove the dummy and plug the cable directly into the input of Box B. Unplug the other end of the cable from the output of Box A and plug it into the dummy (switch open or NORM) as shown below. *Do not plug the dummy into Box A or let it touch anything conductive.*

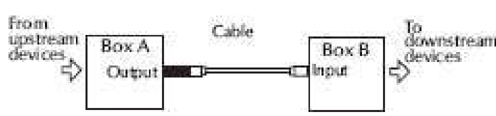


♦ Output quiet?

No—Noise is being induced in the cable itself by an electric or magnetic field. Check the cable for an open shield connection, reroute the cable to avoid the interfering field, or replace the cable with a starquad type (see sections 36.2.5 and 36.4.3).

Yes—Go to next step.

STEP 4: Leaving the dummy in place on the cable, plug the dummy (switch open or NORM) into the output of Box A as shown below.



♦ Output quiet?

No—The problem is shield-current-induced noise (see section 36.4.3). Replace the cable with a different type (without a drain wire) or take steps to reduce current in the shield.

Yes—Go to next step.

STEP 5: Leave the dummy and cable as for step 4, but move the dummy switch to the CMRR (closed) position.

♦ Output quiet?

No—The problem is likely inadequate common-mode rejection of the input stage of Box B. This test is based on the IEC commonmode rejection test but uses the actual common-mode voltage present in the system. The nominal 10Ω imbalance may not simulate the actual imbalance at the output of Box A, but the test will reveal input stages whose CMRR is sensitive to source imbalances. Most often, adding a transformer-based ground isolator at the input of Box B will cure the problem.

Yes—The noise must be coming from (or through) the output of Box A. Perform the same test sequence on the cable(s) connecting Box A to upstream devices.

36.6.3 Solving Interface Problems

36.6.3.1 Ground Isolators

A device called a *ground isolator* solves the inherent common-impedance coupling problem in unbalanced interfaces. Broadly defined, a ground isolator is a differential responding device with high common-mode rejection. It is not a filter that can selectively remove hum, buzz, or other noises when simply placed anywhere in the signal path. To do its job, it must be installed where the noise coupling would otherwise occur.

A transformer is a passive device that fits the definition of a ground isolator. Transformers transfer a voltage from one circuit to another without any electrical connections between the two circuits. It converts an ac signal voltage on its primary winding into a fluctuating magnetic field that is then converted back to an ac signal voltage on its secondary winding (discussed in detail in Chapter 15

Audio Transformers).

As shown in Fig. 36-43, when a transformer is inserted into an unbalanced signal path, the connection between device grounds via the cable shield is broken. This stops the noise current flow in the shield conductor that causes the noise coupling, as discussed in section 36.5.4. As discussed in Chapter 15 *Audio Transformers*, the highest noise rejection is achieved with input-type transformers containing Faraday shields. A transformer-based isolator for consumer audio signals using such transformers, the ISO-MAX® model CI-2RR, is shown in Fig. 36-44. To avoid bandwidth loss, such isolators must be located at the receive end of interconnections, using minimum-length cables between isolator outputs and equipment inputs. Conversely, isolators using output-type transformers, such as most other commercial isolators, may be freely located but will achieve significantly less noise rejection.

Ground isolators can also solve most of the problems associated with balanced interfaces. The ISO-MAX® Pro model PI-2XX shown in Fig. 36-45 often improves CMRR by 40dB to 60dB and provides excellent CMRR even if the signal source is unbalanced. Because it also features DIP switches to reconfigure cable shield ground connections, it can also solve pin 1 problems. Because it uses input-type transformers, it attenuates RF interference such as AM radio by over 20dB. Again, to avoid bandwidth loss, it must be located at the receive end of long cable runs, using minimum-length cables between isolator outputs and equipment inputs. Other models are available for microphone signals and other applications. The vast majority of commercial hum eliminators and a few special-purpose ISO-MAX® models use output-type transformers, which may be freely located but offer significantly less CMRR improvement and have essentially no RF attenuation.

Several manufacturers make active (i.e., powered) ground isolators using some form of the simple differential amplifier shown in Fig. 36-28. Unfortunately, these circuits are exquisitely sensitive to the impedance of the driving source. Fig. 36-46 compares the measured 60Hz (hum) rejection of a typical active isolator to a transformer-based isolator. Over the typical range of consumer output impedances, 100Ω to 1kΩ the transformer has about 80dB more rejection!

Passive isolators based on input-type transformers have other advantages, too. They require no power, they inherently suppress RF interference, and they're immune to most overvoltages that can be sudden death to active circuitry.

36.6.3.2 Multiple Grounding

When a system contains two or more grounded devices, such as the TV receiver and the subwoofer power amplifier in our example home theater system, a wired ground loop is formed as shown in Fig. 36-47.

As discussed in sections 36.5.3 and 36.5.4, noise current flowing in the shaded path can couple noise into the signal as it flows in unbalanced cables or through the equipment's internal the ground path. This system would likely exhibit a loud hum regardless of the input selected or the setting of the volume control because of noise current flow in the 20ft cable. You might be tempted to break this ground loop by lifting the safety ground at the subwoofer. Reread section 36.4.2 and don't do it!

One safe solution is to break the ground loop by installing a ground isolator in the audio path from preamp to subwoofer as shown in Fig. 36-48. This isolator could also be installed in the path from TV receiver to preamp, but it is generally best to isolate the longest lines since they are more prone to coupling than shorter ones.

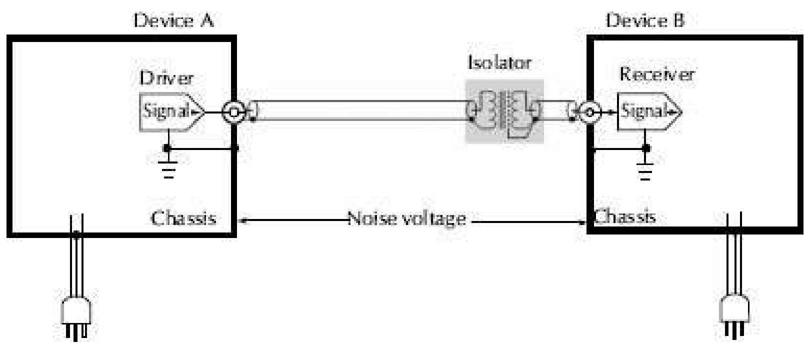


Figure 36-43. Ground isolator stops noise current in shield of unbalanced cable.

Figure 36-44. Stereo unbalanced audio isolator. Courtesy of Jensen Transformers, Inc.

Figure 36-45. Stereo balanced audio isolator. Courtesy of Jensen Transformers, Inc.

Another safe solution is to break the ground loop by installing a ground isolator in the CATV signal path at the TV receiver as shown in Fig. 36-49. These RF isolators generally should be installed where the cable connects to the local system, usually at a VCR or TV input. If an RF isolator is used at the input to a splitter, ground loops may still exist between systems served by the splitter outputs since the splitter provides no ground isolation. Although it can be used with a conventional TV or FM antenna, never install an RF isolator between the CATV drop or antenna and its lightning ground connection (see section 36.4.1). Isolators will *not* pass dc operating power to the dish in DBS TV systems.

Since most unbalanced interfaces are made to consumer devices that have two-prong ac plugs, isolating the signal interfaces may leave one or more pieces of equipment with no ground reference whatsoever. This could allow the voltage between an isolator's input and output to reach 50Vac or more. While this isn't dangerous (leakage current is limited in UL-listed devices), it would require unrealistically high (CMRR over 140dB) performance by the isolator to reject it! The problem is solved by grounding any floating gear as shown in Fig. 36-50. This is best done by replacing the two-prong ac plug with a three-prong type and adding a wire (green preferred) wire connected between the safety ground pin of the new ac plug and a chassis ground point.

A screw may be convenient as the chassis ground point. Use an ohmmeter to check for continuity between the screw and the outer contact of an RCA connector, which itself can be used if no other point is available. Although, in the example above, an added ground at either the preamp or the power amp would suffice, grounding the device with the highest leakage current—usually those devices with the highest ac power consumption rating—will generally result in the lowest noise floor.

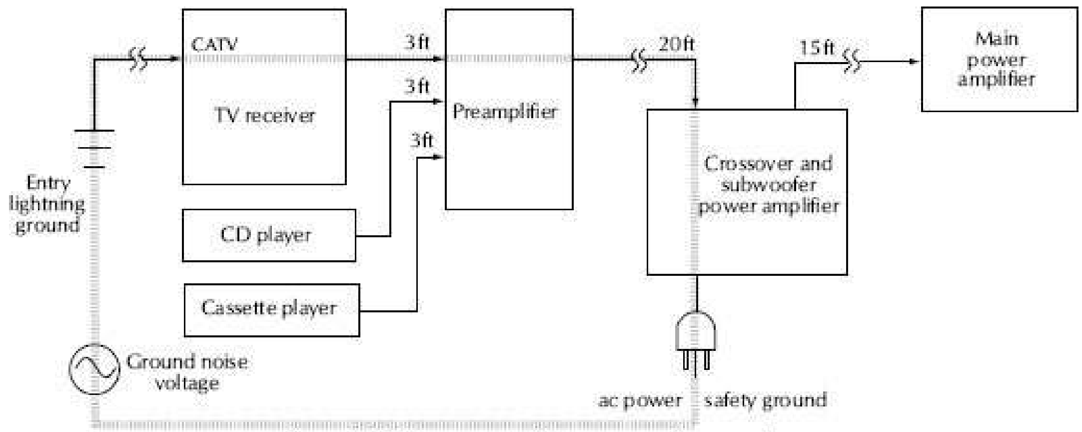


Figure 36-46. Loop created by two ground connections.

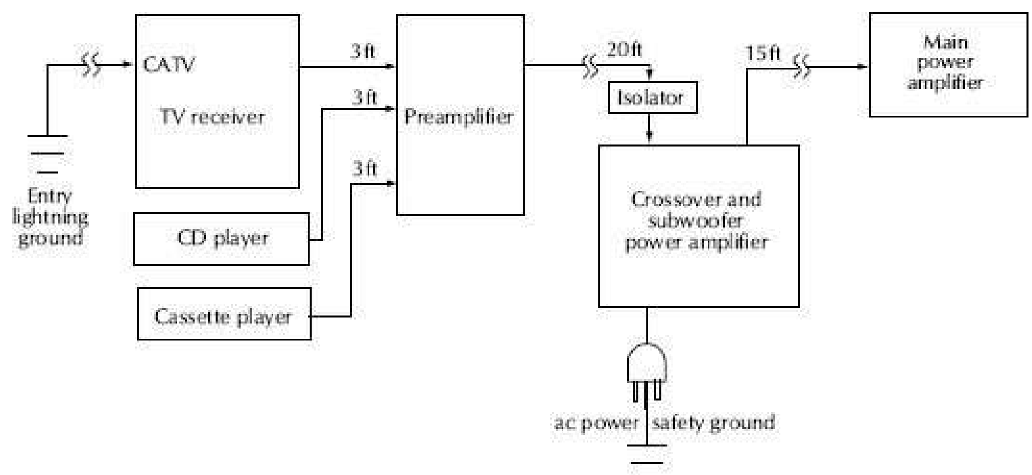


Figure 36-47. Using an audio ground isolator to break the loop.

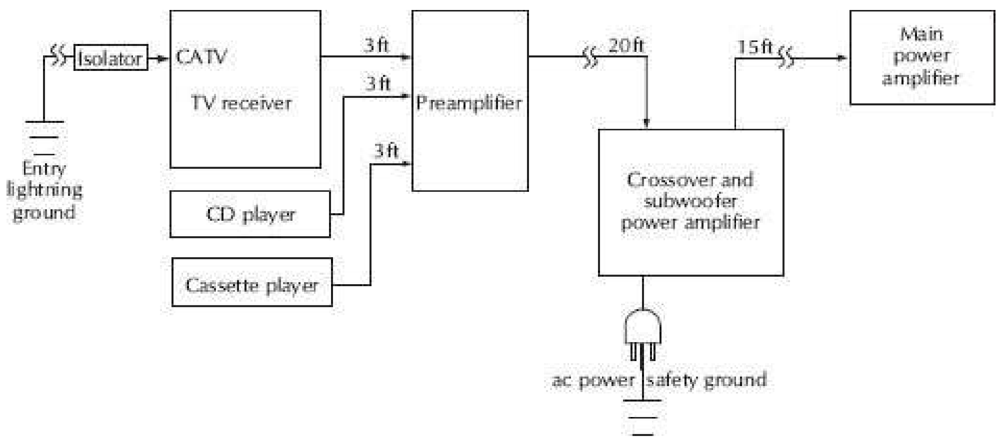


Figure 36-48. Using a CATV ground isolator to break the loop.

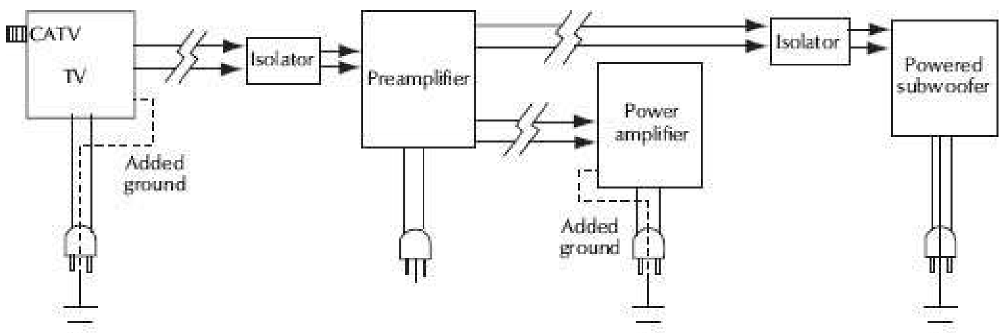


Figure 36-49. Grounding floating equipment when isolators are installed. From Jensen AN004.

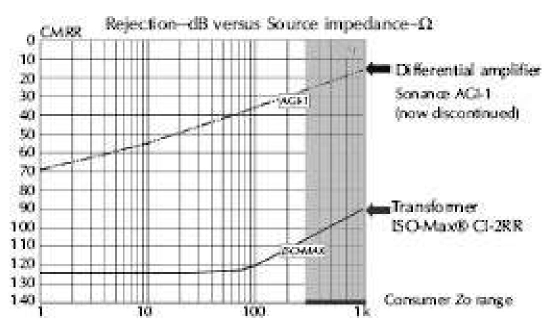


Figure 36-50. Measured hum rejection versus source impedance, active differential amplifier versus input transformer isolator.

36.6.3.3 Unbalanced to Balanced and Vice Versa

The reader is referred to Chapter 15 *Audio Transformers*, section

15.2.2, for a more detailed discussion of these applications.

Beware of RCA to XLR adapters! Fig. 36-51 shows how using this adapter to connect an unbalanced output to a balanced input reduces the interface to an unbalanced one having absolutely *no* ground noise rejection! The potential noise reduction benefit of the balanced input is completely lost.

Proper wiring for this interface, shown in Fig. 36-52, results in at least 20dB of noise rejection even if the balanced input is one of typically mediocre performance. The key difference is that, by using shielded twisted pair cable, the ground noise current flows in a separate conductor that is *not* part of the signal path.

Driving an unbalanced input from a balanced output is not quite as straightforward. Balanced equipment outputs use a wide variety of circuits. Some, such as the one in Fig. 36-53, might be damaged when one output is grounded. Others, including most popular servo-balanced output stages, can become unstable unless the output is directly grounded at the driver end, which reduces the interface to an unbalanced one with no noise rejection.[42] Unless a balanced output already utilizes a built-in transformer, using an external ground isolator such as the one shown in Fig. 36-53 is the only method that will simultaneously avoid weird or damaging behavior and minimize ground noise when used with virtually any output stage. A signal attenuation of 12dB, provided by a 4:1 stepdown turns ratio in the transformer, is also desirable to normalize “pro” and “consumer” operating levels and prevent the very real possibility of overloading the consumer input. This approach is used in the ISO-MAX® Pro model PC-2XR pro-to-consumer interface.

36.6.3.4 RF Interference

As mentioned earlier, immunity to RF interference or RFI is part of good equipment design. Testing for RFI susceptibility is now mandated in Europe. Unfortunately, much of the equipment available today may still have very poor immunity. Under unfavorable conditions, external measures may be needed to achieve adequate immunity.[43]

For RF interference over about 20MHz, ferrite clamshell cores shown in Fig. 36-54, which are easily installed over the outside of a cable, can be very effective. Some typical products are Fair-Rite #0431164281 and Steward #28A0640-0A.[44,45] In most cases, they work best when placed on the cable at or near the receive end. Often they are more effective if the cable is looped through the core several times.

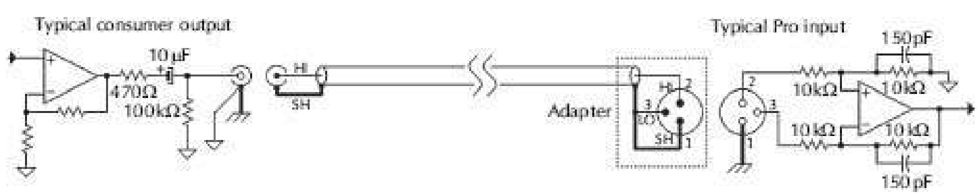


Figure 36-51. Incorrect connection of an unbalanced output to a balanced input.

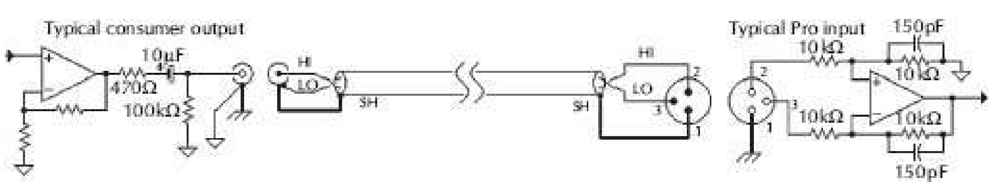


Figure 36-52. Correct connection of an unbalanced output to a balanced input.

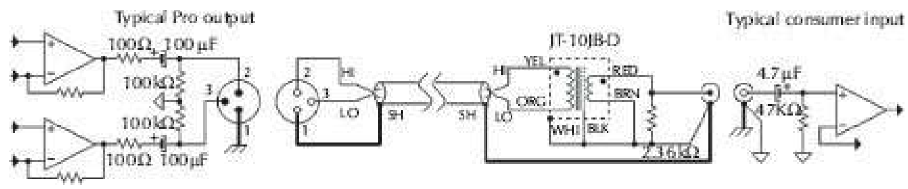


Figure 36-53. Bullet-proof connection of balanced output to unbalanced input. Use the shortest possible cable on the RCA side to avoid common-impedance coupling.

Figure 36-54. Ferrite cores.

If this is inadequate, or the frequency is lower (such as AM radio) you may have to add a low-pass, i.e., high-frequency reject, RFI filter on the signal line. Fig. 36-55 shows sample 50kHz cutoff, 12dB per octave low-pass RFI filters for unbalanced or balanced audio applications. For best performance and audio quality, use NP0 (also called C0G)-type ceramic capacitors keeping leads as short as possible, under ¼ in preferred. For stubborn AM radio interference, it may help to increase the value of C up to about 1000pF maximum. The 680µH inductors are small ferrite core types such as J.W. Miller 78F681J or Mouser 434-22-681. If the only interference is above about 50MHz, a ferrite bead may be used for *L*. For the balanced filter, inductors and capacitors should be ±5% tolerance parts or better to maintain impedance balance. The balanced filter can be used for low-level microphone lines, but miniature toroidal inductors are recommended to minimize potential hum pickup from stray magnetic fields. These filters, too, are generally most effective at the receive end of the cable.

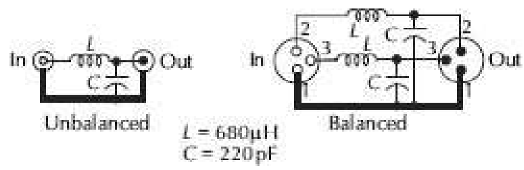


Figure 36-55. RF interference filters for audio lines.

When possible, the best way to deal with RF interference is to control it at its source. Fig. 36-56 is a schematic of a simple interference filter for solid-state 120Vac light dimmers rated up to 600W. It begins attenuating at about 50kHz and is quite effective at suppressing AM radio interference. It must be installed within a few inches of the dimmer and, unfortunately, the components are large enough that it usually requires an extra-deep or a double knock-out box to accommodate both dimmer and filter. Parts cost is under $10.

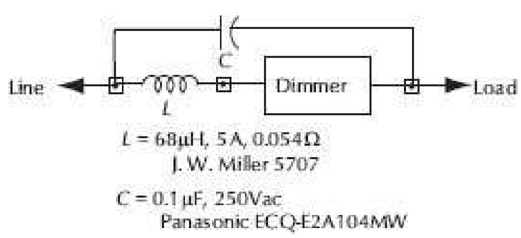


Figure 36-56. RF interference filter for solid-state light dimmer.

A speaker cable can also become an antenna. In a strong RF field, enough voltage can be delivered to the semiconductors in the power amplifier that they become a detector and interference can be heard in the loudspeaker even though the amplifier may be unpowered!

More commonly, this problem occurs when the amplifier is powered and RF enters its feedback loop. In either case, the solution depends on the frequency of the interference. Ferrite cores on the cable near the amplifier may help. In stubborn cases, a 0.1μF or a 0.22μF capacitor directly across the output terminals of the amplifier may also be required.

36.6.3.5 Signal Quality

Audio transformer design involves a set of complex tradeoffs. The vast majority of available audio transformers, even when used as directed, fall short of professional performance levels. As Cal Perkins once wrote, "With transformers, you get what you pay for. Cheap transformers create a host of interface problems, most of which are clearly audible."[46]

The frequency response of a high-quality audio transformer is typically ruler flat, ±0.1dB from 20Hz to 20kHz and –3dB at 0.5Hz and 100kHz. The extended low-frequency response is necessary to achieve low phase distortion.[47] The high-frequency response is tailored to fall gradually, following a Bessel function. This, by definition, eliminates overshoot on square waves and high-frequency response peaking. Dramatic improvements in sonic clarity due to the Bessel filter action are often reported by Jensen customers who add transformers at power amplifier inputs. On the other hand, cheap transformers often have huge ultrasonic peaks in their response that are known to excite particularly ugly intermodulation distortions in even the finest downstream power amplifiers.[48]

Accurate *time domain* performance, sometimes called *transient response*, requires low phase distortion to preserve musical timbre and maintain accurate stereo imaging. Phase distortion not only alters sonic quality, it can also have serious system head room effects. Even though it may have a flat frequency response, a device having high phase distortion can increase peak signal amplitudes up to 15dB. Phase distortion should never be confused with phase shift. Linear phase shift with frequency is simply a benign time delay: only deviations from linear phase or DLP create true phase distortion.[49] This DLP in a high-quality audio transformer is typically under 2° across the entire audio spectrum.

Harmonic and intermodulation distortion in audio transformers is unusually benign in character and cannot fairly be compared to electronic distortion. By their nature, transformers produce the most distortion when driven at high levels at very low frequencies, where the major distortion product is third harmonic. Transformer distortion mechanisms are frequency selective in a way that amplifiers, for example, are not. Electronic nonlinearities tend to produce harmonic distortions that are constant with frequency while high-quality transformer harmonic distortions drop to well under 0.001% at frequencies over a few hundred Hz. Transformers also tend to have remarkably low intermodulation distortion or IMD, to which the ear is particularly sensitive. Compared to an amplifier of comparable low-frequency harmonic distortion, a transformer typically has only a tenth the IMD. While cheap audio transformers use steel cores producing 1% low-frequency harmonic distortion at any signal level, high-quality transformers use cores of special nickel-iron-molybdenum alloys for vanishingly low distortion.

Of course, noise rejection or CMRR is often the most important property of a ground isolator. As discussed in section 36.6.3.1 and

Chapter 15 *Audio Transformers*, a transformer requires an internal Faraday shield (not a magnetic or case shield) to maximize CMRR. Most commercial isolators or hum eliminators consist of tiny imported telephone-grade transformers that do not contain such a shield. Beware of products with vague or nonexistent specs! For example, distortion described as under 0.1% is meaningless because frequency, signal level, and source impedance are not specified. The most common problems with inexpensive isolators are marginal noise reduction, loss of deep bass, bass distortion, and poor transient response. Of course, ad copy and specifications of these transformers will put on their best face, withholding the ugly truth! However, isolators using well-designed and properly applied audio transformers qualify as true high-fidelity devices. They are passive, stable, reliable, and require neither trimming, tweaking, nor excuses.

36.6.3.6 Tips for Balanced Interfaces

Be sure all balanced line pairs are twisted. Twisting is what makes a balanced line immune to interference from magnetic fields. This is especially important in low-level microphone cabling. Wiring at terminal or punch-down blocks and XLR connectors is vulnerable because the twisting is opened up, effectively creating a magnetic pickup loop. In very hostile environments, consider starquad cable because it has less susceptibility to magnetic fields. Magnetic coupling is also reduced by separation distance, cables crossing at right angles rather than running parallel, and shielding with magnetic material such as steel EMT conduit.

Pay attention to cable shield grounding. As discussed in section 36.5.3, the shield must be grounded at the driven end, it may be grounded at both ends, but never grounded only at the receive end. As a standard practice, grounding at both ends is recommended for two reasons:

1. If the device input has marginal RF suppression, grounding the shield at the input will usually reduce problems,
2. It doesn't require the use of a specially wired cable that might find its way into another system and cause unexpected problems. If special cables are made—to deal with a pin 1 problem, for example—be sure they are clearly marked.

Don't terminate to reduce noise. Nearly every practical audio system should use unterminated audio circuits. This is standard professional audio practice worldwide. While a 600Ω termination resistor at an input may reduce noise by up to 6dB or more, depending on the driver output impedance, it will also reduce the signal by the same amount, so nothing is gained. If noise is caused by RF interference, installation of a suitably small capacitor at the input may be much more appropriate.

Use ground isolators to improve noise rejection. As discussed in section 36.4.1, common balanced input circuits have generally unpredictable noise rejection in real-world systems. Actual in-system CMRR can be as little as 30dB when using balanced sources and as little as 10dB when using unbalanced sources. A quality transformer-based ground isolator can increase the CMRR of even the most mediocre balanced input to over 100dB.

Beware of the pin 1 problem. As much as 50% of commercial equipment, some from respected manufacturers, has this designed-in problem. If disconnecting the shield at an input or output reduces a hum problem, the device at one or the other end of that cable may be the culprit. See section 36.5.3 for test methods. Loose connector-mounting hardware is a major cause of pin 1 problems. Never overlook the obvious!

36.6.3.7 Tips for Unbalanced Interfaces

Keep cables as short as possible. Longer cables increase the coupling impedance. Serious noise coupling is nearly certain with 50ft or 100ft cables. Even much shorter cables can produce severe problems if there are multiple grounds. And never coil excess cable length.

Use cables with heavy gauge shields. Cables with shields of foil and thin drain wires increase the common-impedance coupling. Use cables with heavy braided copper shields, especially for long cables. See section 36.7.4 for a recommended high-performance cable. The only property of cable that has any significant effect on audio-frequency noise coupling is shield resistance, which can be measured with an ordinary ohmmeter.

Bundle signal cables. All signal cables between any two boxes should be bundled. For example, if the *L* and *R* cables of a stereo pair are separated, nearby ac magnetic fields will induce a current in the loop formed by the two shields, causing hum in both signals. Likewise, all ac power cords should be bundled. This will tend to average and cancel the magnetic and electrostatic fields they radiate. In general, keeping signal bundles and power bundles separated will reduce coupling.

Maintain good connections. Connectors left undisturbed for long periods can oxidize and develop high contact resistance. Hum or other interference that changes when the connector is wiggled indicates a poor contact. Use a good commercial contact fluid and/or gold-plated connectors to help prevent such problems.

Don’t add unnecessary grounds! Additional grounding almost always *increases* circulating noise currents rather than reducing them. As emphasized earlier, *never* disconnect or defeat the purpose of safety or lightning ground connections to solve a noise problem—the practice is both illegal and very dangerous!

Use ground isolators at problem interfaces. Transformer-based isolators magnetically couple the signal while completely breaking the noise current path through the cable and connectors. This eliminates common-impedance coupling and can improve immunity to RF interference as well.

Predict and solve problems before an installation. For systems that consist mostly of devices with two-prong power cords, some very simple multimeter measurements on each system device and cable makes it possible to actually predict hum levels and identify the problem interfaces before a system is installed.[50]

36.7 Alternative Treatments and Pseudoscience

The audio industry, especially the high-end segment, abounds with misinformation and myth. Science, evidence, and common sense are often discarded in favor of mysticism, marketing hype, and huge profits. Just remember that *the laws of physics have not changed!* See Fig. 36-57.

36.7.1 Grounding from Technical to Bizarre

In most commercial buildings, the ac outlets on any branch circuit are saddle grounded or SG-types mounted in metallic J-boxes. Since SG outlets connect their safety ground terminals to the J-box, the safety ground network may now be in electrical contact with plumbing, air ducts, or structural building steel. This allows coupling of noisy currents from other loads (which might include air conditioning, elevators, and other devices) into the ground used by the sound system. In a scheme called *technical* or *isolated* grounding, safety grounding is not provided by the J-box and conduit but by a separate insulated green wire that must be routed back to the electrical panel alongside the white and black circuit conductors to keep inductance low. The technique uses special insulated ground or IG outlet (marked with a green triangle and sometimes orange in color) which intentionally insulates the green safety ground terminal from the outlet mounting yoke or saddle. The intent of the scheme is to isolate safety ground from conduit. Noise reduction is sometimes further improved by wiring each outlet as a home run back to the electrical panel or subpanel, making each outlet essentially a separate branch circuit.[51] This technique is covered by NEC Article 250-74 and its exceptions. Combining this technique with L-N twisting, as discussed in section 36.4.4.3, can make truly remarkable improvements to system noise performance.

Figure 36-57. Officer Einstein of the Physics Police. Courtesy Coil-Craft.

Many people, believing that the earth simply absorbs noise, have a strong urge to install multiple earth ground rods to fix noise. This is desperation-mode thinking. Code allows extra ground rods, but only if they are bonded to an existing properly implemented safety ground system. Code does not allow them to be used as a substitute for safety grounding—soil resistance is simply too high and unstable to be relied upon to divert fault currents, see Fig. 36-16.[52]

Equipment grounding via the standard power cord safety ground is logical, easy to implement, and safe. It's highly recommended for all systems and is the only practical method for portable or often reconfigured systems.

36.7.2 Power-Line Isolation, Filtering, and Balancing

Most sound systems use utility ac power. If it is disconnected, of course, all hum and noise disappears. This often leads to the odd conclusion that the noise is brought in with the power and that the utility company or the building wiring is to blame.[53] Devices claiming to cleanse and purify ac power have great intuitive appeal and are often applied without hesitation or much thought. A far more effective approach is to locate, and safely eliminate, ground loops that cause coupling of noise into the signal. This solves the real problem. In reality, when system designs are correct, special power treatment is rarely necessary. *Treating the power line to rid it of noise is analogous to repaving all the highways to fix the rough ride of a car. It's much more sensible to correct the cause of the coupling by replacing the shock absorbers!*

First, when any cord-connected line filter, conditioner, or isolation transformer is used, Code requires that the device as well as its load still be connected to safety ground as shown in Fig. 36-58. Cord-connected isolation transformers cannot be treated as separately derived sources unless they are permanently wired into the power distribution system per Code requirements. Sometimes makers of isolation transformers have been known to recommend grounding the shield and output to a separate ground rod. Not only does this violate Code, but the long wire to the remote ground renders the shield ineffective at high frequencies. It is a sobering fact that, while a device may control interference with respect to its own ground reference, it may have little or no effect at the equipment ground.[54,55] Because all these cord-connected devices divert additional 60Hz and high-frequency noise currents into the safety ground system, they often aggravate the very problem they claim to solve. External, cord-connected filters, or those built into outlet strips, can serve to band-aid badly designed equipment. As shown in Fig. 36-24, some equipment is sensitive because common-mode power line disturbances, especially at high frequencies, have essentially been invited in to invade the signal circuitry!

Second, the advertised noise attenuation figures for virtually all these power line devices are obtained in a most unrealistic way. Measurements are made with all equipment (generator, detector, and device under test) mounted to a large metal *ground plane*. Although the resulting specs are impressive, they simply don't apply to performance in real-world systems where ground connections are made with mere wires or conduit. However, these devices can be very effective when installed at the power service entrance or a subpanel, where all system safety grounds are bonded to a common reference point.[56] For thorough, accurate information about separately derived power distribution and its application to equipment racks, the author highly recommends reference 60.

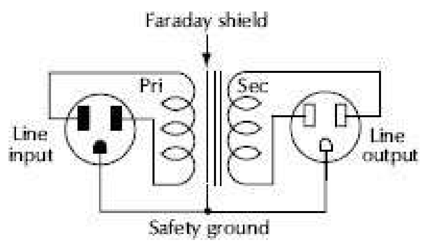


Figure 36-58. Power isolation transformer.

Balanced power, more properly *symmetrical power*, is another seductively appealing concept shown in Fig. 36-59. If we assumed that each system box had neatly matched parasitic capacitances from each leg of the power line to its chassis ground, the resulting noise current flow into the safety ground system would be zero, the interchassis voltage would be zero, and the resulting system noise due to these currents would simply disappear! For example, if C_1 and C_2 had equal capacitance and the ac voltages across them were equal magnitude but opposite polarity, the net leakage current would indeed be zero. However, for the overwhelming majority of equipment, these capacitances are not equal or even close. In many cases, one is several times as large as the other—it's just a reality of power transformer construction. Even if the equipment involved has two-prong ac power connections, actual noise reduction will likely be less than 10dB and rarely exceed 15dB. And it's unlikely that equipment manufacturers will ever pay the premium to match transformer parasitic capacitances or use precision capacitors in power line EMI filters. If the equipment involved has three-prong (grounding) ac power connections, the leakage current reduction, if any, provided by symmetrical power will pale by comparison to the magnetically induced voltage differences described in section 36.3.4. In fact, many of the benefits attributed to symmetrical power may result from simply plugging all system equipment into the same outlet strip or dedicated branch circuit—which is always a good idea.

A GFCI (ground-fault circuit interrupter) works by sensing the difference in current between the hot and neutral connections at an outlet. This difference represents current from the hot conductor that is not returning via neutral. The worst-case scenario assumes that the missing current is flowing through a person. When the difference current reaches 4–7mA, producing a very unpleasant but non-life-threatening shock, an internal circuit breaker removes power in a fraction of a second. Some power conditioners feature a ground lift switch, touted to eliminate ground loop problems, at their outputs. The National Electrical Code requires that all balanced power units have GFCI-protected outputs (among other restrictions on the use of balanced power). Although safe, ground lifting makes a GFCI-protected circuit prone to nuisance trips. For example, consider the system hook-up shown in Fig. 36-60.

For equipment having a grounding (three-conductor) power cord, UL listing requires that its leakage current be no more than 5mA. Normally, this current would flow through the safety ground path back to neutral and would not trip a GFCI that has an intact safety ground connection. However, if the safety ground is lifted and the equipment is connected to other system equipment via signal cables, the leakage current will flow in these cables to reach ground, and ultimately neutral. Because the current is *not* returning via the equipment's own power cord, the GFCI considers it hazardous and may trip, since 5mA is within its trip range. If multiple pieces of equipment are plugged into a single GFCI-protected circuit, the cumulative leakage currents can easily become high enough to trip the GFCI. This problem severely limits the ability of the GFCI/ground-lift combo to solve ground loop problems—even when balanced power partially cancels leakage currents.

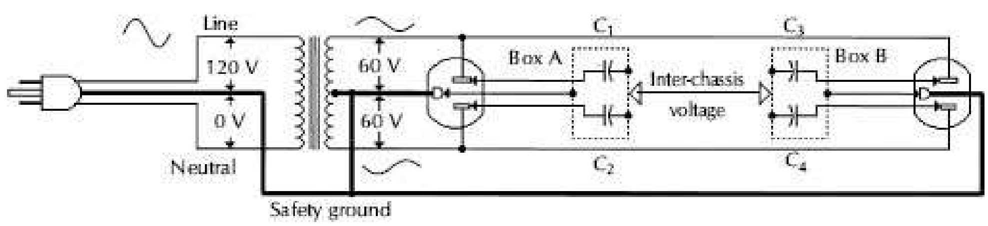


Figure 36-59. Balanced power hopes to cancel ground currents.

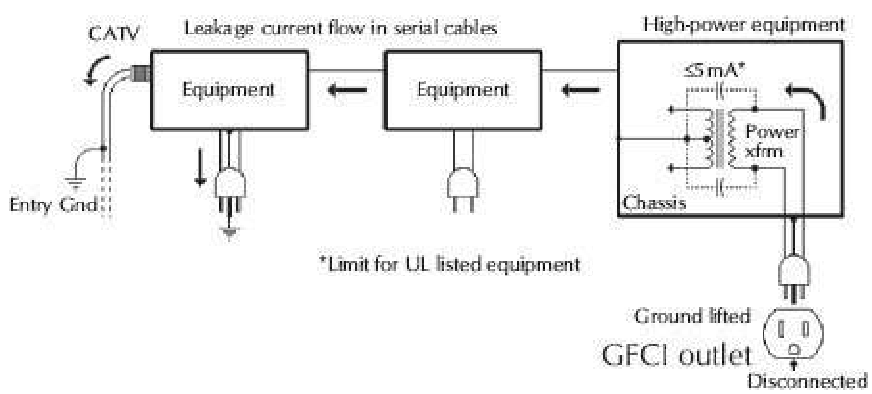


Figure 36-60. Common scenario to produce nuisance trips of GFCI in power conditioner.

36.7.3 Surge Protection

Haphazard placement of common *surge protectors* can actually result in damage to interface hardware if the devices are powered from different branch circuits.[57] As shown in Fig. 36-61, very high voltages can occur should there be an actual surge. The example shows a common protective device using three metal-oxide varistors, usually called MOVs, which limit voltage to about $600V_{peak}$ under very high-current surge conditions.

For protection against lightning-induced power line surges, this author strongly recommends that MOV protective devices, if used at all, be installed only at the main service entry. At subpanels or on branch circuits to protect individual groups of equipment, use series-mode suppressors, such as those by Surge-X, that do not dump surge energy into the safety ground system, creating noise and dangerous voltage differences.[58,59]

36.7.4 Exotic Audio Cables

In the broadest general meaning of the word, every cable is a transmission line. However, the behavior of audio cables less than a few thousand feet long can be fully and rigorously described without transmission line theory. But this theory is often used as a starting point for pseudotechnical arguments that defy all known laws of physics and culminate in outrageous performance claims for audio cables. By some estimates, these specialty cables are now about a $200 million per year business.

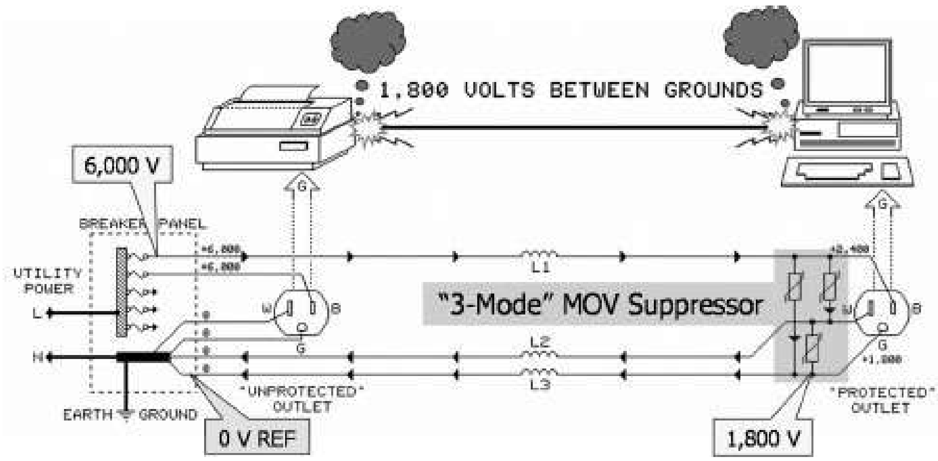


Figure 36-61. Surge protection can actually invite equipment damage.

Beware of cable mysticism! There is nothing unexplainable about audible differences among cables. For example, it is well known that the physical design of an unbalanced cable affects common-impedance coupling at ultrasonic and radio frequencies. Even very low levels of this interference can cause audible spectral contamination in downstream amplifiers.[60] Of course, the real solution is to prevent common-impedance coupling in the first place with a ground isolator, instead of agonizing over which exotic cable makes the most pleasing subtle improvement. Expensive and exotic cables, even if double or triple shielded, made of 100% pure unobtainium, and hand braided by holy virgins, will have NO significant effect on hum and buzz problems! As discussed in section 36.5.4, shielding is usually a trivial issue compared to common-impedance coupling in unbalanced interfaces. It's interesting to note that some designer cables selling for $500/meter pair have no overall shield at all—ground and signal wires are simply woven together.

Some exotic audio cables have very high capacitance and can seriously degrade high-frequency response, especially if cables are long and/or a consumer device drives it. For demanding high-performance applications, consider a low-capacitance, low-shield-resistance cable such as Belden #8241F. Its 17pF/ft capacitance allows driving a 200ft run from a typical 1 kΩ consumer output while maintaining a −3dB bandwidth of 50kHz. And its low 2.6mΩ/ft shield resistance, equivalent to #14 gauge wire, minimizes common-impedance coupling. It's also quite flexible and available in many colors.

References

1. F. W. Grover, *Inductance Calculations*, New York: Van Nostrand, 1944, p. 35, eq. (7).
2. H. Ott, *Noise Reduction Techniques in Electronic Systems*, Second Edition, John Wiley & Sons, 1988, p. 116.
3. D. and C. Davis, *Sound System Engineering*, Second Edition, Focal Press, 1987, p. 382.
4. *Audio Test Procedures*, ATP1.1 through ATP1.4 and Appendix ATP. C, British Broadcasting Corporation, circa 1978–1980.
5. R. Hess, *Voltage Transmission for Audio Systems*, Audio Engineering Society 67th Convention, 1980, Preprint #1708.
6. H. Ott, op. cit., p. 31.
7. D. Gerke, and W. Kimmel, *Kimmel Gerke Bullets*, Kimmel Gerke Associates, Summer 2000, p. 3.
8. H. Ott, op. cit., pp. 47–49.

9. R. Morrison and W. Lewis, *Grounding and Shielding in Facilities*, New York: Wiley-Interscience, 1990, pp. 144–145.

10. D. Gerke, and W. Kimmel, *The Mysteries of Grounding*, IEEE EMC Society Newsletter, Issue 178, Summer 1998, pp. 13–14.

11. R. Morrison and W. Lewis, op. cit., pp. 47–49.

12. B. Whitlock, “Surge Protection: The Enemy Within,” *Sound & Video Contractor*, July 1999, pp. 94–95.

13. B. Whitlock, “Ground Rods: Magical Noise Sinks?” *Sound & Video Contractor*, November 2000, pp. 102–103.

14. B. Whitlock, “Neutral Ground-Finding Swaps in Power Wiring,” *Sound & Video Contractor*, June 2001.

15. H. Denny, *Grounding for the Control of EMI*, Don White Consultants, Inc., Gainesville, VA, 1989, pp. 9.8–9.9.

16. S. Hiser, *1997 Electrocutions Associated with the Use of Consumer Products*, U. S. Consumer Products Safety Commission, 2000, www.cpsc.gov

17. J. Mah, et al., *1997 Residential Fire Loss Estimates*, U.S. Consumer Products Safety Commission, 2000, www.cpsc.gov.

18. H. Denny, op. cit., pp. 7.2–7.3.

19. H. Ott, op. cit., p. 81.

20. H. Ott, op. cit., p. 106.

21. R. Morrison, *Solving Interference Problems in Electronics*, First

Edition, John Wiley & Sons, 1996, p. 35.

22. R. Morrison and W. Lewis, op. cit., pp. 168–169.

23. B. Whitlock and J. Fox, *Ground Loops: The Rest of the Story*, Audio Engineering Society 129th Convention, 2010, Preprint #8234.

24. *Reference Data for Radio Engineers*, Fifth Edition, Howard W. Sams, 1972, section 11–1.

25. B. Whitlock, “Balanced Lines in Audio Systems: Fact, Fiction, and Transformers,” *J. Audio Eng. Soc.*, June 1995, pp. 454–464.

26. B. Whitlock, *A New Balanced Input Circuit for Maximum Common-Mode Rejection in Real-World Environments*, Audio Engineering Society 101st Convention, 1996, Preprint #4372.

27. N. Muncy, “Noise Susceptibility in Analog and Digital Signal Processing Systems,” *J. Audio Eng. Soc.*, June 1995, p. 436.

28. H. Ott, op. cit., p. 105.

29. R. Morrison, *Noise and Other Interfering Signals*, John Wiley & Sons, 1992, p. 55.

30. B. Whitlock, “Balanced Lines in Audio Systems: Fact, Fiction, and Transformers,” *J. Audio Eng. Soc.*, June 1995, pp. 460–462.

31. N. Muncy, op. cit., pp. 441–442.

32. C. Perkins, “Automated Test and Measurement of Common Impedance Coupling in Audio System Shield Terminations,” *J. Audio Eng. Soc.*, June 1995, pp. 488–497.

33. J. Windt, “An Easily Implemented Procedure for Identifying Potential

Electromagnetic Compatibility Problems in New Equipment and Existing Systems: The Hummer Test," *J. Audio Eng. Soc.*, June 1995, pp. 484–487.

34. B. Whitlock, *Hum & Buzz in Unbalanced Interconnect Systems*, Application Note AN004, Jensen Transformers, Inc., 1996.

35. R. Morrison, *Noise and Other Interfering Signals*, John Wiley & Sons, 1992, pp. 136–137.

36. N. Muncy, op. cit., p. 436.

37. H. Ott, op. cit., p. 134.

38. D. Jensen, and G. Sokolich, *Spectral Contamination Measurement*, Audio Engineering Society 85th Convention, 1988, Preprint #2725.

39. B. Whitlock, "Understanding and Controlling RF Interference," *Sound & Video Contractor*, February 1999, pp. 78–80.

40. L. Fielder, "Dynamic Range Issues in the Modern Digital Audio Environment," *J. Audio Eng. Soc.*, May 1995, pp. 322–339.

41. R. Pease, *Troubleshooting Analog Circuits*, Butterworth-Heinemann, 1991.

42. B. Whitlock, *Interconnection of Balanced and Unbalanced Equipment*, Application Note AN003, Jensen Transformers, Inc., 1995.

43. B. Whitlock, *Understanding and Controlling RF Interference*, Sound & Video Contractor, February 1999, pp. 78–80.

44. Fair-Rite Products, One Commercial Row, Wallkill, NY 12589, www.fair-rite.com.

45. Steward, 1200 East 36th Street, Chattanooga, TN 37401, www.steward.com.

46. C. Perkins, "To Hum or Not to Hum," *Sound & Video Contractor*, March 1986, p. 41.

47. W. Leach, "The Differential Time-Delay Distortion and Differential Phase-Shift Distortion as Measures of Phase Linearity," *Journal of the Audio Engineering Society*, September 1989, pp. 709–715.

48. D. Jensen, and G. Sokolich, *Spectral Contamination Measurement*, Audio Engineering Society 85th Convention, 1988, Preprint #2725.

49. D. Jensen, *High-Frequency Phase-Response Specifications—Useful or Misleading?* Audio Engineering Society 81st Convention, 1986, Preprint #2398.

50. B. Whitlock, *Hum & Buzz in Unbalanced Interconnect Systems*, Application Note AN004, Jensen Transformers, Inc., 1996.

51. R. Morrison and W. Lewis, op. cit., pp. 95–98.

52. B. Whitlock, "Ground Rods: Magical Noise Sinks?" *Sound & Video Contractor*, November 2000, pp. 102–103.

53. R. Morrison, *Solving Interference Problems in Electronics*, First Edition, John Wiley & Sons, 1996, p. 1.

54. R. Morrison, and W. Lewis, op. cit., pp. 106, 157–158, 177.

55. R. Morrison, *Solving Interference Problems in Electronics*, First Edition, John Wiley & Sons, 1996, p. 9.

56. R. Morrison, *Solving Interference Problems in Electronics*, First Edition, John Wiley & Sons, 1996, pp. 140–146.

57. B. Whitlock, “Surge Protection: The Enemy Within,” *Sound & Video Contractor*, July 1999, pp. 94–95.

58. J. Harford, “Will My Surge Suppressor Work?” *Sound & Video Contractor*, September 1997.

59. R. Morrison and W. Lewis, op. cit., pp. 159–160.

60. D. Jensen, and G. Sokolich, op. cit.

61. Middle-Atlantic Products, *Integrating Electronic Equipment and Power into Rack Enclosures—Optimized Power Distribution and Grounding for Audio, Video and Electronic Systems, 2008, free white paper download at* www.middleatlantic.com/power.htm.

Suggested Reading

Henry Ott, *Noise Reduction Techniques in Electronic Systems*, Second Edition, John Wiley & Sons, 1988.

Ralph Morrison and Warren Lewis, *Grounding and Shielding in Facilities*, John Wiley & Sons, 1990.

Ralph Morrison, *Grounding and Shielding Techniques in Instrumentation*, Third Edition, John Wiley & Sons, 1986.

Trademarks

ISO-MAX® is a registered trademark of Jensen Transformers, Inc.
InGenius® is a registered trademark of THAT Corporation.
Romex® is a registered trademark of General Cable Industries, Inc.

Chapter 37

System Gain Structure

by Pat Brown

37.1 Introduction

This chapter is devoted to the explanation and establishment of the proper gain structure of the sound reinforcement system. It has been the author's experience that sound systems are rarely producing the optimum performance that would be indicated by the specification sheets of the individual components. Tangible improvements in performance can often be achieved by some simple adjustments of level controls.

Most technical subjects can be best explained using ideal relationships, and this one is no exception. The real world always falls short of the ideal case, but the ideal can present a model and goal for our efforts. It is the responsibility of the sound practitioner to form an understanding of the trade-offs and apparent contradictions through experience and endless hours in the field. What follows is only an introduction that will benefit those who supplement it with lab and field work.

37.2 Interfaces

An interface exists when two components are to be interconnected for the purpose of transferring a signal. One component will be the source (sending) device and the other the load (receiving) device for the electrical signal. At least three major topologies exist for interconnecting devices, the major difference being which electrical parameter of the signal that the interface optimizes the passage of—i.e., voltage, current, or power. This is primarily a function of the ratio between the source impedance and load impedance. At this point we will make our first simplification by assuming that the impedance of these devices is purely resistive with no appreciable reactive component. This is actually a pretty accurate assumption for most electronic components in the signal processing chain.

37.2.1 The Matched Interface

A matched interface means that the source and load impedances are equal. This topology has some admirable attributes:

1. Power transfer is maximized.

2. Reflections from load-to-source are reduced or eliminated.

Impedance matching is required when the electrical wavelengths of the audio signal are shorter than the interconnecting cable. Examples include antenna circuits, digital interfaces, and long analog telephone lines. A drawback of this interface is that power transfer is optimized at the expense of voltage transfer, and therefore the source device might be called on to source appreciable current. It is also more difficult to split a signal from one output to multiple inputs, as this upsets the impedance match. A component that is operated into a matched impedance is said to be *terminated*. While the telephone company must use the matched interface due to their electrically long lines, the audio industry departed from the practice many years ago in favor of the voltage-optimized interface for analog interconnects.

Fig. 37-1 shows a matched interface. It is important to note that the selection of 600Ω as the source and load impedance is arbitrary. It is the impedance ratio that is of importance, not the actual value used.

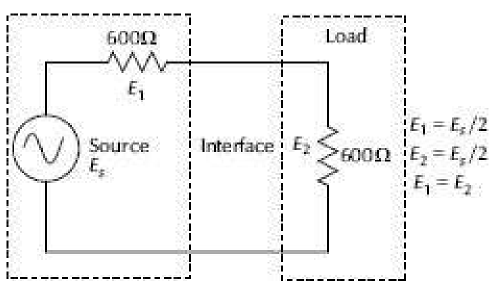


Figure 37-1. Matched interface, where *E* is the voltage.

37.2.2 The Constant Voltage Interface

Most analog sound system components are designed to operate under *constant voltage* conditions. This means that the input impedance of the driven device is at least ten times higher than the output impedance of the source device. This mode of operation assures that output voltage of the driving device is relatively independent of the presence of the driven device—hence the term constant voltage, Fig. 37-2. Constant voltage interfaces can be used in analog audio systems since the typical cable length is far shorter than the electrical wavelengths of the audio signal propagating down the cable. This makes such lines immune to the detrimental effects of reflections and standing waves. Radio, digital, and telephone engineers are not so fortunate, and impedance matching is required at component interfaces. Constant voltage (sometimes called bridged) interfaces are inherently simpler than their impedance matched cousins. Advantages include the ability for a single output to drive multiple high-impedance inputs (in parallel) without signal loss or degradation. Also, the constant voltage interface does not require that manufacturers standardize their input and output impedances. As long as the output impedance is low (typically less than 1000Ω) and the input impedance is high (typically greater than 10kΩ) then the two devices are compatible. In practice most output impedances are fairly low (<100Ω), allowing a single low-impedance output to drive several high-impedance inputs, Fig. 37-3.

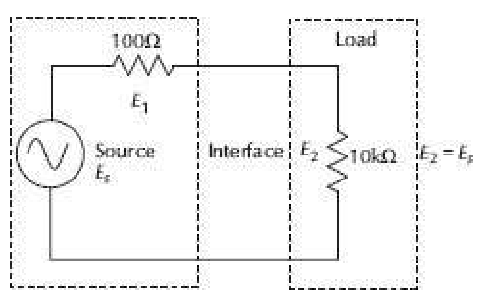


Figure 37-2. Constant voltage interface.

If the source impedance is large when compared to the load, then a constant current interface is formed. In this topology, the current flow from the source is determined by the source impedance and is independent of the load impedance. Constant current interfaces are not often used to interface electronic components, and are usually reserved for specialized applications, such as the construction of impedance meters. I will not consider this interface further in this chapter.

37.3 Audio Waveform Fundamentals

In a sound reinforcement system, *program sources* provide information that is to be reinforced and presented to a listener. This information can originate in the form of an acoustical wave (acoustic musical instruments or human voices) or an electrical wave (electronic instruments or storage media such as compact disc). In either case, the waveform must be in the electromagnetic domain prior to being presented to the sound system. Acoustic signals must be converted into electromagnetic signals with an appropriate transducer such as a microphone or accelerometer. We will refer to electromagnetic waves within the bandwidth of the human auditory system as *audio* waveforms. Typical audio waveforms are quite complex and are continuously changing in value over time. This makes it difficult to describe them numerically. Several parameters can be used to describe the characteristics of an audio waveform. These include the following:

Peak-to-Peak Voltage. The number of volts between the largest positive and largest negative peak of the waveform.

Peak Voltage. The highest peak of the waveform, regardless of whether it is positive or negative. For a waveform with amplitude symmetry, it will be one-half the peak-to-peak voltage.

Average Voltage. The average of all ± amplitude values of the waveform.

Root-Mean-Square (rms) Voltage. Sometimes called the *effective* value of the waveform, rms describes the ac voltage in terms of the equivalent dc voltage that would produce the same amount of heat into a resistive load. Rms is useful because it indicates the heating value of the waveform. The rms level of a complex audio waveform is also related to its perceived loudness if it is used to drive a loudspeaker. For a sine wave, the rms voltage is 0.707 times the peak voltage. Complex waveforms also have an rms voltage, but finding it requires integration of the waveform over time. The peak-to-rms ratio of a waveform is called its *crest* factor. Crest factors must be described in terms of a finite span of time. A span of 50ms correlates well with the integration time of the human hearing system, but other values can be used.

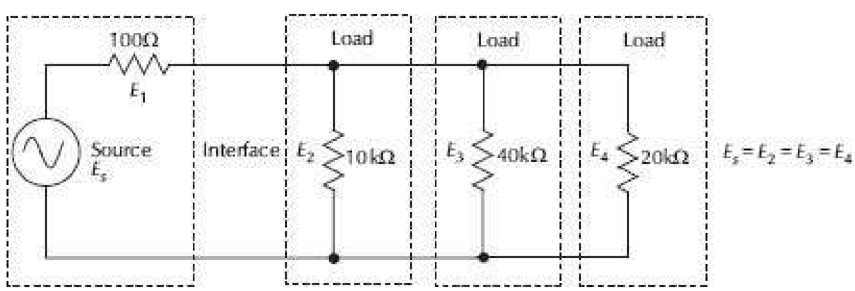


Figure 37-3. Constant voltage bridged interface driving multiple inputs. Voltage E_1 is negligibly small.

Fig. 37-4 shows a sine wave and speech waveform on a common plot of amplitude versus time. An oscilloscope provides this representation of the data, as does a wave editor application for a personal computer.

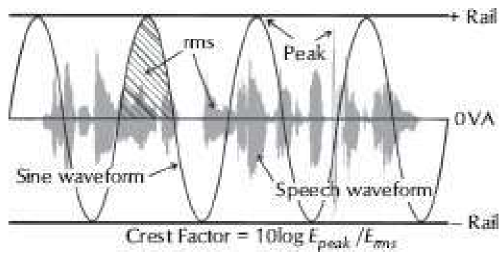


Figure 37-4. Crest factors for sine waves and speech waves.

The peak voltage of the waveform must pass through the sound system component without being clipped. It is the parameter of interest when establishing the gain structure of the system. The crest factor of the signal determines the energy content and therefore the audio power produced by the amplifier and delivered to the loudspeaker. It is of interest when considering the heat that the loudspeaker must dissipate. Additionally, the rms voltage is the signal parameter that relates most closely to the loudness of the signal as perceived by a human listener.

Since the goal of an audio system is to reproduce appropriate waveforms for a given application, these waveform principles have universal application for all parts of the sound system.

37.4 Gain Structure

The following are general terms relevant to gain structure beginning with how it applies to an individual piece of electronic equipment. It matters little whether the equipment is a mixer, equalizer, amplifier, or other active system component. By active we mean that the device has a power supply for its internal active circuitry. This can be as simple as an internal battery or two, or as complex as an internal or external ac line-powered supply. The power supply voltages establish the maximum amplitude that a waveform can take on as it passes through the component, Fig. 37-4. In audio equipment, most power supplies form a bipolar set of rails—a fixed positive and negative dc voltage that the waveform is developed between. The value of the rail voltage determines the peak amplitude that the waveform can take on. Exceeding this peak value will cause deformation of the wave, commonly known as *clipping*. We will proceed under the assumption that the rails are fixed, and indeed they are for most signal processing devices. Some power amplifier topologies use multivalued or fluctuating rails. The principles are the same, but the rail voltages are dynamic rather than static.

Under a no input signal condition, all audio components will still emit a residual output signal. Thermal noise is generated at the molecular level and is present at the output of all system components whether active or passive. The level of the thermal noise determines the noise floor of the component. In practice, other factors can also make a contribution to the residual noise of an electronic device. An undersized power transformer or poor shielding can elevate some frequencies above the broadband thermal noise floor, Fig. 37-5. Equipment designers try to minimize thermal noise by component selection and careful design, but it can never be eliminated. We must accept the fact that it exists. Part of the reason for establishing the proper gain structure of a system is to render the effects of thermal noise insignificant. The thermal noise floor is affected by the settings of the component's level controls. While a low noise floor can be achieved with all controls set at minimum, this is not realistic, as we cannot operate it that way. The controls should be set at a point appropriate for the operation of the device. A good starting point is a setting that produces the same voltage at the device output that is present at the device input, often called *unity gain* by audio practitioners. Level controls placed at their 0dB setting generally produce this condition, and represent a good starting point for setting up a system.

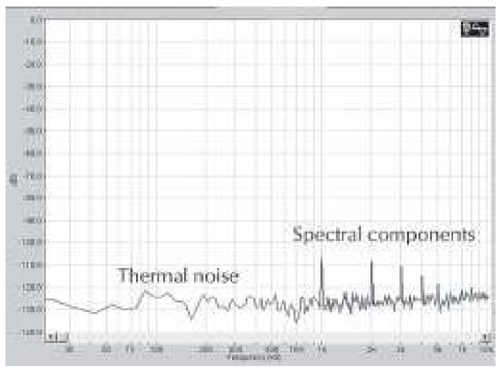


Figure 37-5. Thermal and spectral noise.

The supply rails and noise floor establish the dynamic range for the device—the difference in level between the highest possible undistorted peak and the lowest level that the signal can take on without being buried in the noise. The dynamic range is what *can* happen when a signal is passed through a device. It is a range of possible values that the waveform can take on. The possibilities are infinite (within a device's dynamic range) for an analog component and finite for digital components, since digital signals are made up of discrete samples that must be quantized to fixed steps. Fig. 37-6 shows a 1kHz sine wave driving a component to just below clipping. The level difference between clipping and the noise floor describes the component's dynamic range.

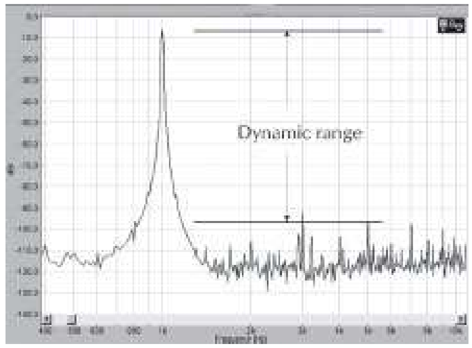


Figure 37-6. Dynamic range.

An example is in order. Let us consider a line level audio signal processor. We can pick any rail voltage that we like, since no universally recognized standard exists. A rail voltage of +17.5Vdc (and −17.5V to for the negative rail) will allow a peak voltage of 17.5V to be realized at the device output. It is customary to express this voltage in terms of the rms value of the largest sine wave that the device (peak × 0.707, or peak level minus 3dB) can produce with an acceptable amount of distortion. An oscilloscope allows observation of the wave and any deformation due to clipping. This rms voltage becomes the maximum output voltage, and when expressed in decibels becomes the component's maximum output level. This level is expressed in dBm (dB ref. 0.001W) for impedance matched interfaces (note that knowledge of the circuit impedance is required) or dBV (dB ref. 1V) or dBu (dB ref. 0.775V) for constant voltage interfaces (assuming the bridged impedance condition is maintained).

$$L_{out} = 20\log(V_{peak}) - 3 \text{ dB in dBv} \qquad (37\text{-}1)$$

$$L_{out} = 20\log\left(\frac{V_{peak}}{0.775}\right) - 3 \text{ in dBu} \qquad (37\text{-}2)$$

$$
\begin{aligned}
L_{out} &= 20\log(17.5) - 3 \\
&= 21.8\ \text{dB dBV} \\
&= 20\log\left(\frac{17.5}{0.775}\right) - 3 \\
&= 24\ \text{dBu}
\end{aligned}
$$

Assume that the thermal noise measured at the device output is about 200μVrms, as measured using an rms broadband voltmeter. Expressed as a level in dBV, the thermal noise floor becomes:

$$L_{noise} = 20\log(\text{output noise}) - 3\text{dB} \qquad (37\text{-}3)$$

$$L_{noise} = 20\log\left(\frac{\text{output noise}}{0.775}\right) - 3\text{dB} \qquad (37\text{-}4)$$

$$
\begin{aligned}
L_{noise} &= 20\log(0.0002) \\
&= -74\ \text{dBV} \\
&= 20\log\left(\frac{0.0002}{0.775}\right) \\
&= -71.8\ \text{dBu}
\end{aligned}
$$

This audio component ($L_{out} - L_{noise}$) thus has a dynamic range on the order of 100dB, a very good figure, and one that is typical for a well-designed piece of audio equipment, whether analog or digital.

With the dynamic range established, it is still necessary to use it effectively. If a weak signal is fed to the component, it may fall far short of the clipping point established by the power supply voltages, placing it unnecessarily close to the component's noise floor. This will produce a poor signal-to-noise ratio, *SNR*, even in a component that has a wide dynamic range, Fig. 37-7.

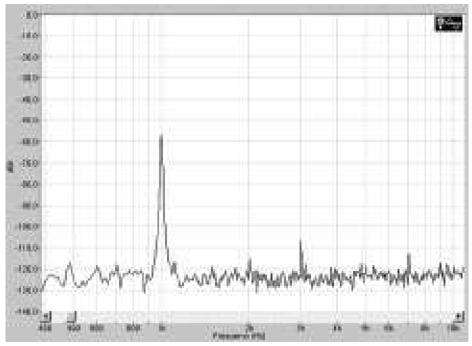


Figure 37-7. Poor *SNR*.

If the input level control is increased beyond unity, the thermal noise will likely increase with the signal voltage, and no increase in *SNR* is realized. Increasing the drive (source) voltage will improve the *SNR*, assuming that the sending device has a noise floor lower than the driven device. In some cases, an additional gain stage may be required, as with microphones and phonograph cartridges.

If too strong a signal is fed to the component, the highest amplitude parts of the waveform may not fit within the constraints of the power supply voltages and may drive the component into a nonlinear mode of operation (clipping). This may yield an excellent signal-to-noise performance, but a distorted output signal rich in harmonic distortion, Fig. 37-8. Gain structure, from a component perspective, is passing the signal through at optimum amplitude—not too strong and not too weak. As such, a system component can be overdriven, underdriven, or optimally driven by a signal source. It is important to note that the *SNR* of the program source is often the determining factor for the SNR of the entire system, since it can only be improved with very specialized signal processing that is not found in typical reinforcement systems. The old adage "garbage in, garbage out" certainly applies. The *SNR* will be degraded as it passes through other system components, which is why care must be taken to properly calibrate each stage of the system.

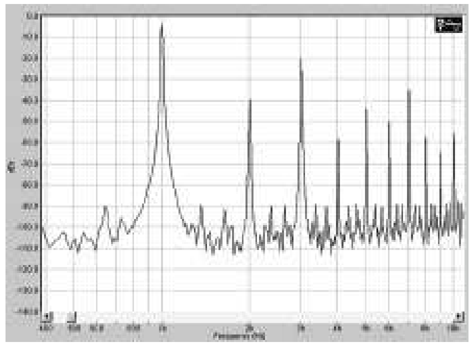


Figure 37-8. Overdriven signal characteristics.

37.5 System Gain Structure

Audio components have been evolving since early in the last century. In the process the dynamic ranges of system components have become similar, and in many cases approach the theoretical limits dictated by nature. While overall dynamic range may be similar, the clipping levels and noise floors are not identical between manufacturers or even within product lines. While we won't go into all of the reasons for this, it is unfortunate that at least clipping levels for line level components and amplifier input stages aren't standardized within the sound reinforcement industry. As such, it is possible for a system component to be operating optimally within its own dynamic range, and yet be overdriving or underdriving the next component. This reality forces us to consider gain structure from system perspective.

Before discussing the gain structure of a sound system, it is necessary to consider a method for determining the internal gain structure of a system component. This can be done by introducing a stimulus to the component and observing its output signal. It is common practice for technicians to use a stable and repeatable waveform for calibrating the signal-processing chain. A sinusoidal waveform, commonly called a sine wave, is such a waveform. The sine wave is a single frequency tone that is easily generated, fed to an input, and observed at the output of each component in the chain. The previous graphs have shown sine waves displayed on a magnitude versus frequency plot. They resemble a vertical spike due to the narrow frequency bandwidth. An alternate and equally valid display is amplitude versus time. The oscilloscope displays the amplitude of the waveform as a function of time. More advanced models will even provide some statistics, such as peak voltage, rms voltage, frequency, crest factor, level, etc. Let us proceed. A 1000Hz sine wave is developed across the input terminals of one channel of the mixer. An amplitude is chosen that does not overdrive the input, and that is of sufficient level to drive the mixer to its full undistorted output voltage with all level controls in the signal path set at unity. For microphone inputs, about 0.1Vrms (−20dB ref. 1V) is generally sufficient. As much as 1Vrms might be required for a line level input. The level controls of the mixer are set as follows:

- Master at unity or 0dB.
- Channel at unity or 0dB.
- Trim at unity or 0dB.

Under these conditions, the voltage amplitude of the output signal should be the same as the voltage amplitude of the input signal—an amplification factor of one or unity, and a gain of 0dB.

The input voltage has been increased until the main meter of the mixer reads zero. We will speak in more detail about zero later, but for now we will assume that it indicates a voltage in the optimum operating range of the mixer's overall dynamic range (typically −20dB rel. clipping). Since program audio waveforms are constantly changing, this operating level allows some room for peaks in the audio waveform to pass undistorted. It is instructive at this point to measure the output voltage of the mixer at meter zero. Using the oscilloscope, either the peak or rms value of the waveform can be measured. It is traditional to measure the rms value, since it is readily measured with much less sophisticated voltmeters than the oscilloscope and correlates well with the loudness and heat production of the signal.

For historical reasons, a common voltage measured at a mixer output with the meter indicating zero is 1.23Vrms, corresponding to an rms open circuit level of +4dB ref. 0.775V (+4dBu). This level might be termed the *operating level* of the mixer. A volume indicating meter describes the waveform in a way that correlates with its loudness as perceived by human listeners. The indication of such a meter is in VU, or volume units. When sound systems used the impedance matched interface, this voltage was developed across the input impedance (usually 600Ω) of the next input stage. Since voltage and impedance were known, the power equation could be used to calculate the output power of the mixer, which became

$$W = \frac{E^2}{R} \tag{37-5}$$

$$\begin{aligned} L_{out} &= 10\log\frac{V_{rms}^2}{600} \\ &= 10\log\frac{1.23^2}{600} \\ &= 4\,\text{dBm} \end{aligned} \tag{37-6}$$

The power transfer was relevant since the matched interface was used, optimizing the circuit for power transfer. One milliwatt (mW) is a useful reference, as it falls in the middle of the range of power levels found in the sound system. A voltmeter calibrated to read zero at 0.775V could directly indicate the circuit power level in dBm (assuming a 600Ω matched impedance interface). This calibration would make the voltmeter a dBm meter when placed across a 600Ω circuit. When dBm meters are used at other impedances, a correction factor is required. As the sound reinforcement industry migrated to the constant voltage interface, the 0.775V reference lived on due to the proliferation of voltmeters so calibrated, and signals were then described in dB ref. 0.775V or dBu. In modern systems, the term *level* is used to describe the field quantity of interest at a component interface, which is the signal voltage for the constant-voltage interface. This a good place to note that many modern mixers do not use the +4dBu reference level for meter zero, so the reader is advised to consult the literature or perform a measurement. A more common meter zero level today is 0dBV, or 1Vrms.

Let us now advance the trim control (or the drive voltage) until the waveform becomes distorted when viewed on the scope. Some mixers have a clipping indicator to warn of this condition. When the waveform flattens on top, reduce the trim control until the waveform appears undistorted. Since mixers are made up of several stages, it is usually informative to move each the main fader, channel fader, and trim control until clipping is observed to assure that each stage is clipping simultaneously. This produces the maximum output voltage of the mixer at the mixer's output terminal. Using the scope or a voltmeter, measure the voltage of the waveform. Note that the clipping occurs on the peak of the waveform, yet it is standard practice to measure the rms value of the waveform and include it on the specification sheet. Ideally, this maximum output voltage is at least ten times the voltage measured at the meter zero indication, providing 20dB of peak room above meter zero. Once the maximum output level has been measured, the drive level (or trim control setting) should now be reduced to produce the meter zero operating level of the mixer.

We now have knowledge of the operating and clipping level of the mixer (e.g., +4dBu and +24dBu respectively). These values should be recorded in the system documentation. The noise floor of the mixer can be measured by muting the input signal and measuring the mixer's no signal output voltage, but this is of little interest in practice.

37.6 The Unity Method

The mixer is now at an optimum operating level with good *SNR* and 20dB of peak room. The signal from the mixer is fed to the next component in the chain. If the component has input and/or output level controls, they are adjusted to produce the same level of the mixer at that component's output terminal. In like manner the signal is fed through subsequent signal processors (if present), and the mixer's level eventually becomes the input level of the power amplifier. The amplifier's input sensitivity control is set for the desired output voltage to the loudspeaker. This establishes the playback level of the system). As the amplifier's voltage is impressed across the loudspeaker load, the amplifier supplies current flow as determined by the impedance of the loudspeaker. Power will flow, but the signal level is a linear function of the applied voltage over the useful operating range of the amplifier. So, the output voltage is the parameter of interest in a properly configured amplifier-to-loudspeaker interface under normal operating conditions. Fig. 37-9 shows such a processing chain for a mixer with a 0dBV meter zero. The unity gain method has a number of advantages, which include:

1. Ease of calibration.
2. Fast implementation.
3. Easy substitution of components.

Unfortunately, there are some drawbacks to this approach, mostly due to the non-standardization of clipping levels between product lines and manufacturers. A mixer operating at 0dBV that clips at +20dBV will have 20dB of operating peak room for transient peaks. If the component after the mixer clips at +18dBV, that component will only have 18dB of operating peak room. In this case an undistorted full-scale waveform from the mixer would cause clipping in the next component. The mixer could be turned down a bit if the overdrive is not severe. If the level mismatch is more than a few dB then a different solution may be required. It should be pointed out that this condition is not as prevalent as it once was, as many post mixer product manufacturers have modified their products to handle the higher output voltages produced by modern mixing consoles.

37.7 An Optimized Method

The drawbacks of the unity gain method can be overcome with an optimized method of establishing the gain structure of the system. While the unity method establishes a consistent operating voltage from component to component, the optimized method sets each device to clip simultaneously, regardless of the actual signal level. A mixer outputting +24dBV and an equalizer outputting +20dBV are both set to reach their clipping point simultaneously. This method often requires the insertion of a resistive attenuator between the mixer and equalizer, allowing the mixer to output its maximum voltage and still not overdrive the equalizer.

To optimize the system gain structure, feed a sine wave to the mixer in the same manner previously described, but this time advance the trim control until clipping occurs at the mixer output. All power amplifiers should be off or fully attenuated. The clipping can be detected with the aid of an oscilloscope or a spectrum analyzer capable of handling at least +30dB ref. 1V (+30dBV). With the mixer set just short of clipping, connect the output of the mixer to the input of the next component (i.e., an equalizer). Set all controls on the equalizer at their unity setting. Move the clipping indicator (scope) to the output of the equalizer and note whether the waveform is clipped. If it is not, the equalizer is capable of passing the full output voltage of the mixer. If the equalizer is clipping, first try reducing the setting of its input level control. This often doesn't work since the stage being overdriven likely precedes the level control stage. Some manufacturers design their equipment to handle higher input levels than they can output, and the input level control may indeed eliminate the overdrive condition. If not, an attenuator is placed between the mixer and equalizer and adjusted to produce an undistorted waveform from the equalizer. The same procedure is repeated for each subsequent piece of equipment in the processing chain. Compressor/limiters should be set at their highest threshold setting and lowest compression ratio. Crossover networks require either:

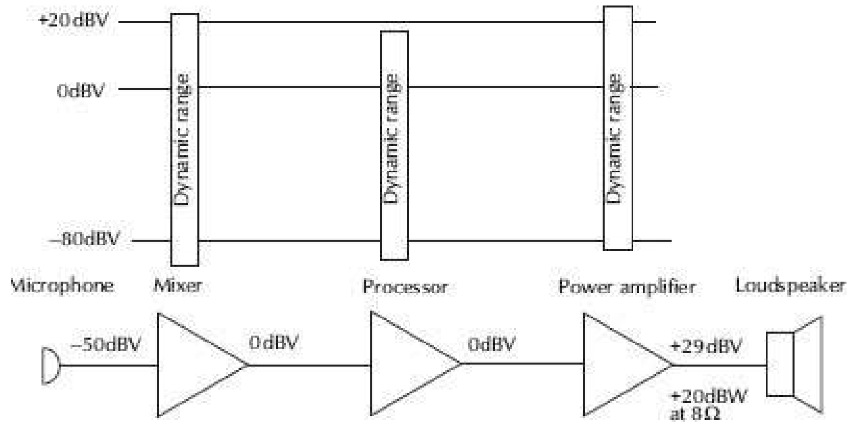


Figure 37-9. The processing chain.

1. Selection of a sine wave within the pass band of the output being tested, or
2. Readjust the crossover frequency control to allow the output being tested to pass the test signal. Remember to restore the setting before turning on the amplifier!

Resistive pads for the “too hot” source are available commercially. If a component is being *underdriven* by the previous one (i.e., the full output voltage of the mixer is insufficient to clip the equalizer), it may be advantageous to increase the underdriven component’s input level until just short of clipping. This will provide a stronger drive voltage to the *next* component (and possibly an improved signal to noise).

The advantages of the optimized method include:

1. The ability to mix at meter zero without the potential to clip a component farther down the processing chain.
2. Optimized *SNR* in all components in the chain.
3. The mixer’s meter can now serve as an accurate indicator for all subsequent components, since all clip simultaneously.

As with all things audio, the optimized method is not without drawbacks. These include:

1. It requires more time and expertise to set up.
2. It requires a method to determine device clipping (scope or spectrum analyzer).
3. It requires the purchase of or construction of pads.
4. It makes component substitution more difficult, since the replacement component may have a different clipping level than the defective one.

A pad of 5–15dB may be necessary for a professional mixer driving a consumer recorder.

Fig. 37-10 shows a system whose gain structure has been optimized in this manner. The benefit of either method is that a healthy drive voltage with a good *SNR* is delivered to the power amplifier input. Both methods also assure that any digital components in the processing chain are being driven with a voltage high enough to produce optimized A/D conversion.

37.8 Setting the Amplifier Sensitivity

Ideally the amplifier’s input stage should handle the full output level of the preceding device without clipping. It is possible for the amplifier input circuit to clip prior to its output stage. This can be tested by setting the amplifier attenuator at a very low level and observing the output waveform of the amplifier when driven with the full undistorted signal level of the preceding device. If clipping is apparent at the amplifier output at a low attenuator setting, the input stage is being overdriven. Insertion of a pad or a reduction in drive voltage at the output of the previous stage will be required.

If an active crossover is in the signal chain, its proper settings should be established prior to switching on the amplifier and setting the its input sensitivity. These settings are best obtained from the loudspeaker manufacturer.

The amplifier could be calibrated in the same manner as the other components—simply adjust its input sensitivity (volume) control to produce an output signal just short of clipping. Since this may be too loud, it is better to use a broadband program source (pink noise or music) and adjust the amplifier for the desired L_P at the listener position. The procedure is as follows.

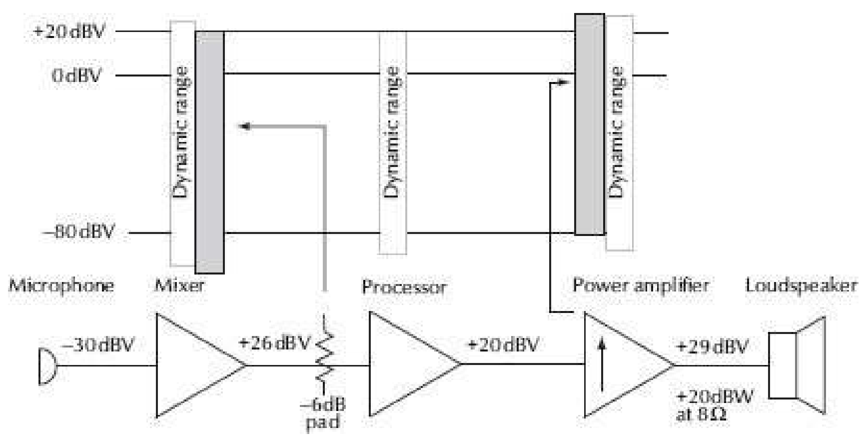


Figure 37-10. Optimized system gain structure.

With the program material being input to the mixer, the mixer is set to produce a zero meter indication as previously described. Note that this assumes a loudness (VU) meter and not a peak program meter. The input attenuator of the power amplifier is then advanced until:

1. The desired acoustic level is reached in the audience, or
2. The amplifier begins to indicate clipping.

In either case, don't turn it up any higher. The gain structure is complete and the system is producing its maximum undistorted L_P. The technician can now proceed to fine tuning of the signal processor to finish the system calibration.

37.9 Power Ratings

It is important to note that the amplifier's wattage rating must be *appropriate* for the loudspeaker. The loudspeaker must not be driven beyond its ability to dissipate heat buildup or beyond its limits of mechanical excursion. In most systems, heat will be a bigger problem than excursion, so the rms value of the amplifier's output voltage waveform must be managed. Since the continuous power (based on the rms voltage) delivered to the loudspeaker is largely a function of the type (crest factor) of the program material, a signal similar to the intended program material in spectrum and crest factor should be used. Pink noise is a common choice.

37.9.1 Amplifier Power Ratings

The power ratings of power amplifiers and loudspeakers have little in common. The amplifier is usually rated in accordance with the maximum continuous power that it can deliver reliably with sinusoidal input for a specified span of time into a specified load impedance. This yields a large number for the amplifier power rating (due to the high rms voltage of the sine wave), and most likely a wattage that the amplifier will never be called on to deliver, since the signals that we present to audiences usually bear little resemblance to sine waves. Even so, this rating can be useful for amplifier comparisons and selection. Just remember, you won't get that much rms voltage across the loudspeaker with real-world program material.

37.9.2 Loudspeaker Power Ratings

The loudspeaker's continuous power rating describes the loudspeaker's ability to dissipate heat on a continuous basis. A meaningful rating must state at a minimum:

1. The type and crest factor of the signal used.
2. The bandwidth of the signal.
3. The time duration of the test.
4. The rms voltage of the signal.
5. The impedance of the loudspeaker under test.

If the signal used has a crest factor of 6dB, and the loudspeaker is rated at 50W continuous, the amplifier size required to run the test would be 200W (peak) or 100W (sine wave rating.)

$$\begin{aligned} 17\ \text{dBW} + 6\ \text{dB} &= 23\ \text{dBW} \\ 23\ \text{dBW} &= 200\ \text{W}_{\text{peak}} \end{aligned}$$

Power specifications are of little use to system technicians. They must be converted to an equivalent rms voltage so that the system tech can measure the signal with a voltmeter. A simple conversion for a loudspeaker is to multiply the power rating by eight and take the square root to get the voltage. Bear in mind that if the power rating has been exaggerated, the voltage resulting from this conversion will be too.

Due to the high crest factors of audio program material, power amplifiers normally deliver far below their theoretical maximum sine wave power. This makes it possible (and necessary) to use an amplifier whose continuous power rating exceeds the continuous power rating of the loudspeaker if the intent is to produce the maximum L_P possible. Care is required on the part of the user to insure that the crest factor of the program material is not reduced excessively by dynamic range control devices (compressors and limiters) and then used to drive the amplifier to the point of clipping. Figs. 37-11 and 37-12 show the same waveform. The peak output voltage of each waveform is the same. A peak limiter was used to reduce the dynamic range of the second waveform, resulting in a 6dB increase in applied rms voltage (and continuous power) to the loudspeaker (2× the voltage and 4× the power). This example shows how that amplifier power (and loudspeaker power dissipation) are highly dependent on the nature of the waveform, not just the amplifier rating. The amplifier selection and setting should ideally depend on the target sound level in the audience. There are no ramifications to operating a loudspeaker below its power rating, and in fact it is good design practice.

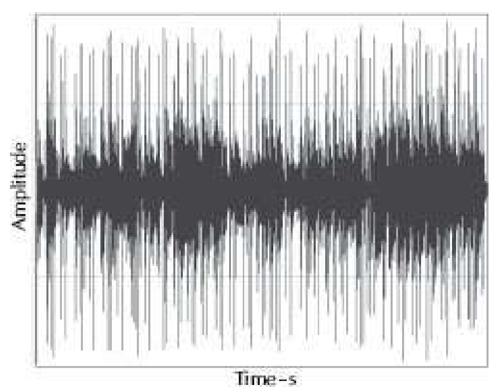


Figure 37-11. Output voltage of a complex waveform with large peaks.

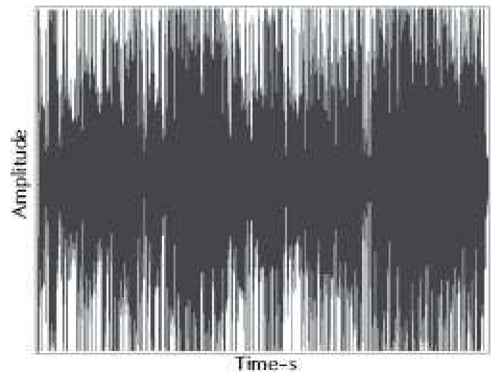


Figure 37-12. Output voltage of the waveform in Fig. 37-11 utilizing a 6dB limiter and normalized to full scale.

One is reluctant to formalize a formula for determining the required amplifier size for several reasons, including:

1. The continuous output power of an amplifier is a function of the crest factor of the program material and can vary by 20dB or more (a 100:1 ratio!).
2. Amplifier power ratings are based on signals that can bear little resemblance to audio program material.
3. Monitoring actual power delivered to the loudspeaker requires sophisticated equipment and a knowledgeable operator.
4. Standard loudspeaker power handling tests require that the loudspeaker be driven to the point where no permanent damage occurs. This is a bit ambiguous. The author utilizes a power handling test that drives the loudspeaker with increasing rms voltage until its response changes by 3dB from the small signal (typically 3Vrms) response. This rms voltage is used to determine the continuous rating of the loudspeaker, either in volts rms or power into a rated impedance.

Even so, a conservative approach is as follows:

1. Determine the loudspeaker's continuous power rating in watts (from the specification sheet). Determine the maximum rms voltage by taking the square root of eight times the power rating. This assumes an 8Ω loudspeaker. This voltage is necessary for level setting and verification.
2. Quadruple this rating for the required amplifier size. This will allow program peaks to exceed the continuous rating by 6dB.
3. Be careful to not clip the amplifier, Fig. 37-13.

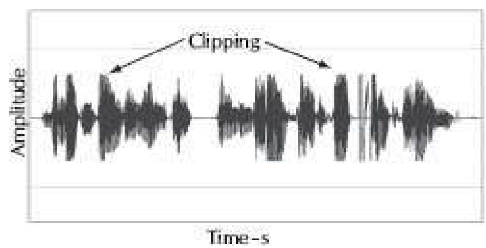


Figure 37-13. Waveform clipped due to insufficient head room.

If the crest factor of the program material exceeds 6 dB, and the amplifier is operated without clipping, the loudspeaker will simply be operating further below its continuous rating, increasing its reliability and longevity. A careful operator could use a significantly larger amplifier, provided that a high crest factor is maintained and clipping is avoided, Fig. 37-14.

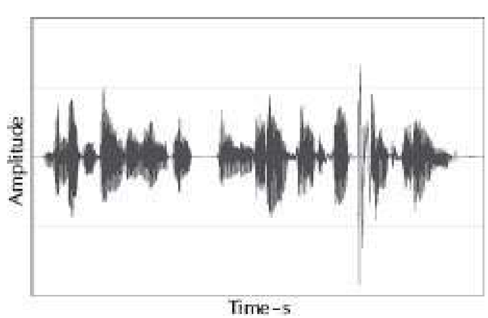


Figure 37-14. Increasing the amplifier size produces additional peak room.

In essence, buy a big amplifier but use it carefully. Don't overdrive the loudspeaker or the audience! A sound level meter should always be used to check the L_P produced by the system, and this value should be within OSHA exposure guidelines.

37.10 Conclusion

A properly calibrated sound system allows the operator to mix at or near meter zero on the mixer without danger of clipping any system component or damaging the loudspeaker. Meter zero should also correlate with the maximum desired L_P in the audience. In effect, all components in the system are now functioning as one component, the only difference being that they are housed in separate chassis and interconnected with cables.

Bibliography

Milton Kaufman and Arthor Seidman, *Handbook for Electronics Engineering Technicians*, 1976 McGraw-Hill, Inc. Don Davis and Eugene Patronis, *Sound System Engineering*, 2006 Focal Press, Burlington. MA.

Chapter 38

Sound System Design

by Pat Brown

38.1 Introduction

The design of a sound reinforcement system must be governed by some objectives. These depend on the intended use of the system and can be different for speech systems and music systems. Either type of system should deliver the information it carries to the listener with clarity.

The design process must be applicable to existing rooms as well as future ones. If the room exists, go there and listen. Listening is a vital part of the sound system design process. What follows is a logical and meaningful way to evaluate the performance of a loudspeaker in the room. The first assumption I will make is that the loudspeaker is functioning properly, and that the sound quality is as intended. This means that the electrical signal has sufficient bandwidth, output level and dynamic range, and that the loudspeaker is working properly and has not been stressed or damaged.

There are two major inhibitors to sound clarity in auditoriums. These are *noise* and *reflected* sound. Reflected sound can be subdivided into specular reflections and reverberation. I'll assume that the reflected sound is reverberation, and that it is quantified by the room's decay time, and by the level it builds to when the room is continually excited with sound. These assumptions grossly simplify the consideration of the room's acoustic behavior, but simplify the analysis by being fairly consistent throughout the space. It may or may not be possible for the sound system designer to manipulate the ambient noise and/or reverberation time of the space. The design process put forth here will assume that these quantities are fixed and must therefore influence in the design process.

38.2 The “Back-Away” Test

The first exercise is to evaluate the spoken word in the space. I like to use a “talker loudspeaker”, Fig. 38-1, for this purpose, as this allows me to be the listener, and allows me to conduct the test alone in a quiet room.

Queue up a speech track for playback. Beginning at a 1m distance, listen to the loudspeaker. Most likely, it is easy to understand because at one meter the speech level is likely well above the room’s ambient noise, and well above the level of any reverberation that develops in the space. If echoes are present, they are likely low in level relative to the talker’s voice. Staying on axis, double the axial distance by backing away from the source. You are now at 2m. The direct field from the source is reduced by 6dB, while the other sound fields remain about the same. This increase in distance has likely degraded the transfer of information. Double the distance again and the same thing happens. Repeat until the sound becomes unintelligible. Note the furthest distance at which the sound clarity is acceptable. This is the maximum working distance from the source.

Figure 38-1. A "talker" loudspeaker provides a starting point. Courtesy NTI Audio.

What has happened? Moving away from the talker has reduced the direct-to-reverberant ratio (*DRR*) and the signal-to-noise ratio (*SNR*). The clarity of the sound has been reduced by one or both of these factors.

Next, turn "off" any noise sources driving the room. These can include fans, lights, pumps, etc. Return to the farthest listener position and re-assess the clarity. If it has improved, then the room noise was reducing the clarity of the source, and you should note that *SNR* will be a factor in this space. Noise must be addressed at its source (e.g. a noisy HVAC system or fountain).

Next, substitute a loudspeaker of medium directivity. This should be a physically larger device that utilizes "horn loading" to make it more directional. Fig. 38-2 shows an example.

With the source producing the same axial level as the talker loudspeaker, repeat the listening experiment and note any increase in the working distance from the source. If it increased, then reverberation was inhibiting the clarity. Reducing reverberation requires room treatment, increased source directivity, or both.

This demonstration has yielded a wealth of information about the room, and about the required characteristics of the loudspeaker that can tame it. The two major mechanisms for increasing the clarity of the sound are *level* and *directivity*. Increasing the level improves the *SNR*. Increasing the directivity improves the *DRR*.

Figure 38-2. A well-controlled medium directivity loudspeaker. Courtesy Mitchell Acoustic Research, (www.frazierspeakers.com).

This simple example is the heart of the sound system design process. A sound system's job is to extend the experience of the listener located at the maximum listening distance to all of the listening positions in the space. We have amplifiers and loudspeakers that can provide the increased level. We have directional loudspeakers that can keep the sound energy on the audience and off of the room.

Our investigation only considered a listener on-axis with the source. Of course in a real auditorium there are listeners at many angles from the loudspeaker, and at many distances. Loudspeaker specifications include polar data that give the sound pressure level

(L_P) at all angles around the source.

The flowchart in Fig. 38-3 provides a logical roadmap for designing a sound system. It establishes the thought process for assuring that the design will provide adequate clarity for all listeners. It will guide the system designer to make an appropriate loudspeaker choice given the acoustic properties of the space and the placement of the loudspeaker(s).

38.3 In Situ Design

Given enough time and money, the entire system design process could be carried out "in situ" using the procedure just described. A loudspeaker could be placed at a central position, aimed, and its coverage evaluated by walking the audience area and listening. This would take all of the variables at play into account. If unsatisfied with the sound, a different loudspeaker could be substituted for the first one and the process repeated. The loudspeaker could then be elevated to a practical trim height, re-aimed and the process repeated. If one device cannot cover the entire audience, additional loudspeakers could be added and the process repeated. While maybe impractical, this would actually be a very good way to design a sound system. What you hear is what you get.

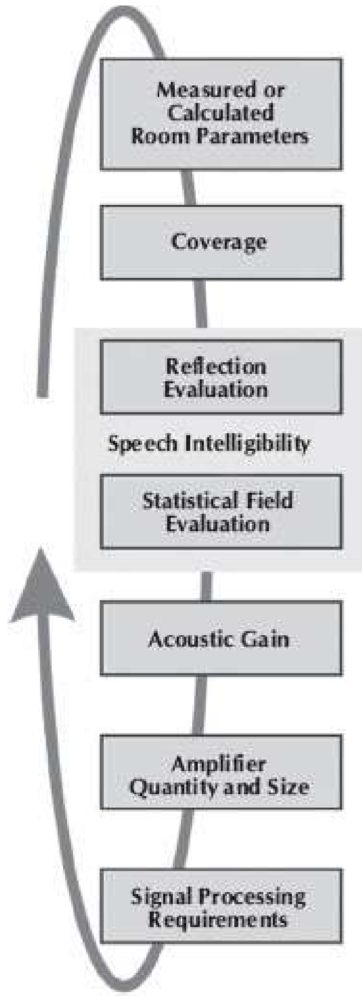


Figure 38-3. A sound system design flowchart.

38.4 A Walk Through the Flow Chart

Note that the flow chart begins with “Measured or Calculated Room Parameters.” The performance of a sound system is not independent of the room that it is placed in. At a minimum the reverberation time (*RT*) and noise floor (L_N) are needed. If the room exists, the *RT* should be measured. If it doesn’t, it should be estimated by calculation. Measured data is always preferred to calculated data, as errors in acoustic predictions can be high due to the many variables involved and methods used.

So, the first consideration is that the system sounds good somewhere, and this somewhere is on-axis to the loudspeaker, and it must extend to the furthest listener position. This is what was evaluated with the back-away test. The second consideration is how to achieve coverage of the audience area, since in most rooms a single loudspeaker will not likely cover all of the seats. The assumption is that each listener needs to clearly hear one loudspeaker, so we begin with one, evaluate its coverage by moving away from the axial position, and then make changes and re-evaluate. For a simple system it can be done by listening. The only practical way to do this for large rooms is with computer modeling, which I will cover later in this chapter.

When a loudspeaker is placed in a room, acoustic reflections are produced. This is unavoidable. Room reflections may be good or bad, depending on their arrival time, level and direction. Reflection evaluation is based on geometric principles, where the angle-of-incidence equals the angle-of-reflection. We will assume that the sound behaves like light, which is only an approximation. As with coverage, the only practical way to consider all of the possible source-surface-receiver paths is by using a room modeling program.

In some rooms, the sound introduced from the loudspeaker produces a reverberant field that may mask the information contained in the direct field. This requires consideration of the reverberant sound *level*, and its ratio to the direct sound level. This is one aspect of the design that can be grossly estimated with some classical reverberation equations, and for many years this was the heart of the design process. Computer room modeling is a much more powerful approach, especially since few rooms meet the criteria that make the classical equations valid.

The acoustic gain block of the flow chart begs the question “Will the system get loud enough before going into regenerative feedback?” This of course assumes that there will be live microphones driving the system, and that these are in the same room as the loudspeakers. It is good design practice to minimize the L_P directed to the areas where microphones will be placed (e.g. the stage). It is also good practice to keep source-to-microphone distances as short as possible. I will assume that the combination of these two measures allows for sufficient acoustic gain.

Now that the loudspeakers and their placements are known, the amplifiers can be selected. Enough audio power is needed to produce the required direct field L_P at the listener with sufficient room for signal peaks. We must also assure that the drive signal will not damage the loudspeaker, so information regarding its power handling is needed. In some cases it may be necessary to start the design from scratch, should it prove to be impossible to generate the needed L_P with the current approach.

The required signal processing can only be determined after the selection and placement of loudspeakers. Digital signal processors (DSP) make extraordinary amounts of processing available to the system designer. Many auditorium designs include one DSP channel and one amplifier per loudspeaker. This can allow multiple settings based on system use, as well as experimentation with various configurations. The establishment of the signal processing chain within the DSP is the last step of the design process.

System design is a trial-and-error process, and one can return to the beginning at any time.

38.5 Performance Objectives

There are at least three aspects of the system's performance that the designer must quantify prior to beginning the process.

1. How loud must it be?
2. What is the furthest distance at which it must be this loud?
3. What is the required bandwidth?

The answers to these questions are necessary for initial selections of loudspeakers and amplifiers. The target L_P must be carefully considered, with input from the client and consideration of the intended use of the system. Coming up short on L_P is an expensive mistake.

38.6 Loudspeaker Specifications

Loudspeakers have numerous attributes and specifications that describe them. A very complete set of metrics is provided by the Common Loudspeaker Format (CLF). The CLF is a file format that serves two purposes.

1. It provides important loudspeaker specifications of interest to system designers.
2. It is a data file that can be imported into major room modeling computer programs for drawing board performance investigations.

Fig. 38-4 shows a screen shot of the freeware CLF Viewer (Microsoft Windows).

While all of the information contained in the CLF is informative and useful, there are two specifications that are vital early in the design process. These include the sensitivity and the maximum input voltage (*MIV*). Here is an overview of each.

Sensitivity is a measure of the axial (on-axis) L_P at a reference distance that results from the application of a known voltage to the loudspeaker. The applied voltage is usually 2.83Vrms and the reference distance is usually 1m. Large loudspeakers are usually measured at a greater distance and the result normalized to one meter using the inverse-square law rate of level change. The applied signal should be broadband, and usually takes the form of pink noise or swept tone (chirp). Sensitivity is measured and specified by the loudspeaker's manufacturer. It should be given as a plot of level vs. frequency, and the average over the useful bandwidth may be used as a "one number" rating for design work.

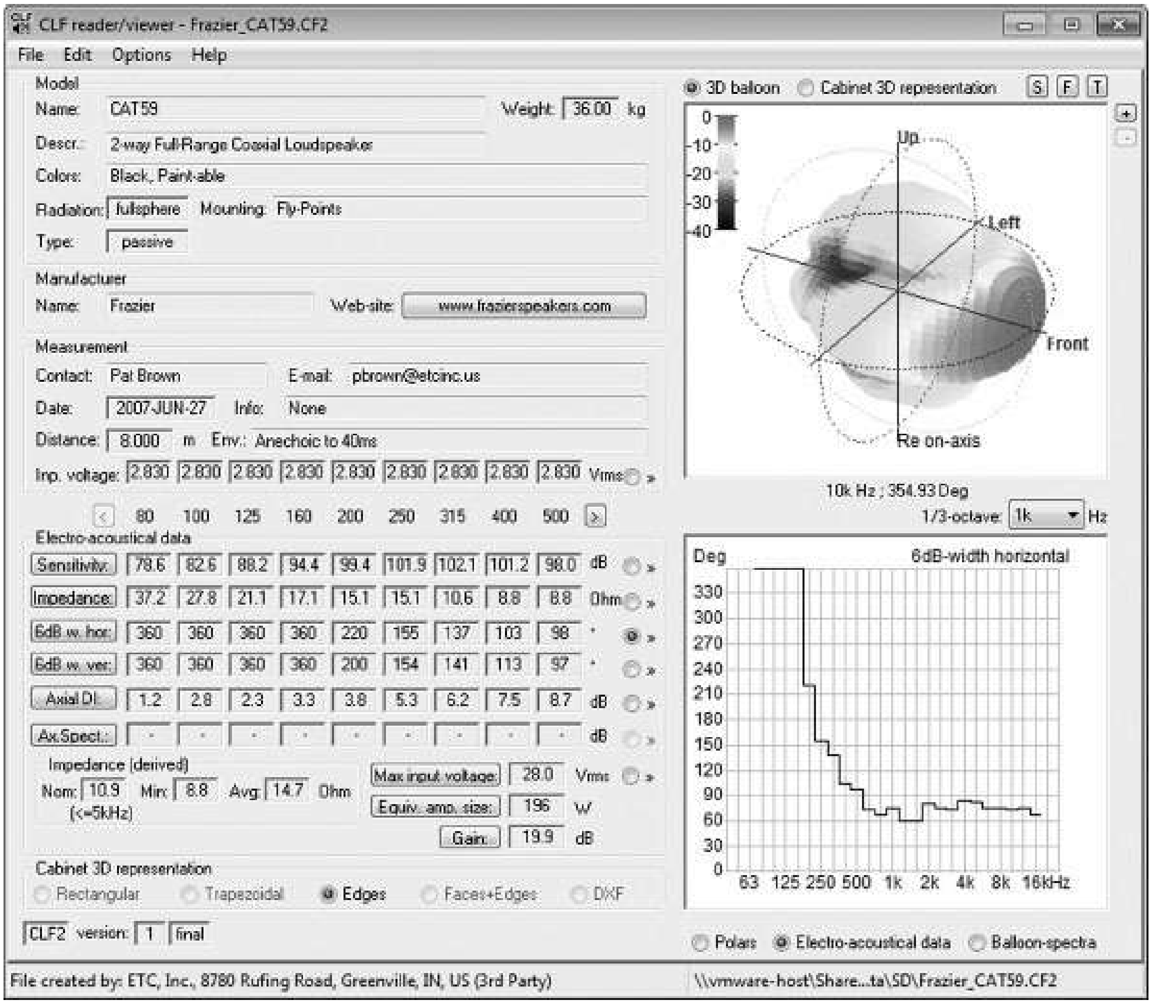


Figure 38-4. The freeware CLF data viewer (www.clfgroup.org).

The *MIV* is the maximum sustainable drive voltage that can be applied to the loudspeaker without damaging it. The axial sensitivity can be scaled by the *MIV* to determine the maximum L_P that the loudspeaker can produce at one meter, and this level can be extended to any distance using the inverse-square law. This will reveal whether the loudspeaker can produce the needed level at the furthest listener distance. If this isn't possible, there is little reason to investigate the off-axis performance of the loudspeaker.

What are appropriate targets for the level and bandwidth of the system? These depend on the system type.

38.7 Speech Systems

The most effective communication system known to humans is face-to-face communication. A sound reinforcement system for speech should emulate it. It consists of one sound source (the talker) and a two-channel receiver (the listener). The sound pressure level (L_P) is about 60dB A-wtd for an average male talker at 1m. This should produce an adequate signal-to-noise ratio (*SNR*) in quiet environments for the information carried in the acoustic signal to be received and decoded by the listener. It should produce an adequate direct-to-rever-berant ratio (*DRR*) in all but the most reverberant spaces, mainly because the distance between the talker and listener is short (1m). In other words, hopefully we are designing a system for a room in which face-to-face communication is possible.

As this distance is increased, the *DRR* quickly degrades and the speech may become unintelligible. The sound system designer must understand the principles required to extend the *DRR* of the face-to-face scenario further into the room. The system can be designed to produce nearly any L_P that humans can tolerate, but a comfortable listening level is preferred.

38.8 Music Systems

A music system is a speech system with extended bandwidth. It will provide intelligible speech if spoken over. This requirement is often overlooked by system designers, but has become increasingly important for voice evacuation reasons.

38.9 System Bandwidth and Frequency

Resolution

Frequency resolution is a variable that audio and acoustic practitioners deal with in many areas. The 20Hz–20kHz bandwidth of the human auditory system must be subdivided for processing, measurement and reproduction. One means of subdividing the spectrum is the use of musical intervals, such as decades, octaves and octave fractions. Fig. 38-5 shows the resolutions important to sound system designers.

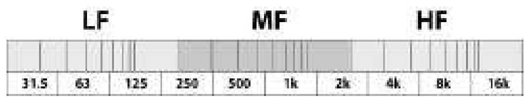


Figure 38-5. Dividing up the audible spectrum.

One-decade resolution is of significance to loudspeaker system designers, since a given fixed-size radiator can be expected to have about one decade of bandwidth when used at levels consistent with sound reinforcement. The audible spectrum can be divided into three decades – low frequency (LF), mid frequency (MF) and high frequency (HF). It can be further subdivided into octaves. There are ten octaves in the audible spectrum, and their center frequencies are given in Fig. 38-5. A high quality speech system should cover the 125Hz to 8kHz octave bands. This is sufficient bandwidth do a respectable job of reproducing musical program material for many applications, such as background music (BGM) and foreground music (FGM). The bandwidth may be extended further (e.g. 10 octaves) for high fidelity systems such as used for cinema and contemporary concerts.

This chapter will present a design process based on 1/1-octave resolution. It requires the level, coverage and clarity to be considered for at least 7 octave bands. While higher resolution (e.g.

1/3-octave) is possible, it is usually not practical to map a room 30 times to evaluate the system's coverage for each 1/3-octave band. 1/1-octave resolution is also the accepted norm for acoustics work, which is a major part of the design process. Higher resolution may be useful for special investigations and for the direct sound field.

38.10 The Listening Environment

The performance of a sound reinforcement system is dramatically influenced by the space that encloses it. In fact the room has nearly as much to do with the sound as the sound system itself. A good system can sound bad in a bad room. A bad system can sound acceptable in a good room. A well-designed sound system in a pleasant acoustic environment may deliver what might be referred to as "good sound."

The design process must begin with the analysis of the room. This requires the collection of the room impulse response (RIR), described later in this chapter, and in Chapter 49, *Test and Measurement*. Using the RIR, the space can be categorized as dead, live or reverberant. An acoustically dead space offers little competition to the direct field produced by the loudspeaker. Outdoor environments and indoor environments with heavy upholstery qualify. If a positive *SNR* can be achieved, it is reasonable to expect that communication can occur over the system. But how can I know for certain?

Fortunately the back-away test can be conducted mathematically with a few assumptions.

1. The L_P falls off at the inverse-square law rate of level change. Since the wave front is spherical in shape, each time you double

the listener distance from the 1m initial position, the sound level drops 6dB.

2. The reverberant field that builds in the room is uniform throughout the room. This is a fundamental characteristic of true reverberation. It should be noted that many rooms are semi-reverberant, so this condition may only be partially true.

Before the use of computers, the sound system designer would sit down with measured or predicted room reverberation times, loudspeaker polar measurements, and a calculator (or slide rule) and perform direct-to-reverberant sound energy calculations at a sampling of listener seats. This is still how it is done, but by using computer room modeling we can consider many more seats, semi-reverberant spaces, and handle some special cases that don't fit our idealized assumptions, all in less time than a manual approach.

38.11 Computer-Aided Sound System Design

Given the number of variables in play, and the complex ways that they interact, a modern sound system design process must be aided by the personal computer. Computer room modeling programs allow the designer to quickly evaluate the feasibility of their ideas.

I'll begin this section with an important statement—"Computers can't design sound systems." While this may not be true in the future, it is true now. Many potential sound system designers have been dismayed after purchasing an expensive design program, to find out that it is basically a calculator that executes algorithms that are based on simplified assumptions of sound wave behavior. That is not to say that these programs aren't sophisticated. Some are incredibly complex, a mixture of deterministic calculations and proprietary algorithms for simulating acoustic behavior based on geometric principles. Do sound waves behave like light rays? “Yes,” in some ways, but “No” in others.

For an excellent overview of the history of room acoustics modeling, see *“The Early History of Ray Tracing in Room Acoustics*” by Peter Svensson.[1] This work clearly shows that acoustics modeling is a game of approximations, assumptions and compromise. It’s not simply a matter of number crunching.

The remainder of this chapter is devoted to enabling the sound system designer to utilize the room modeling program of their choice to its fullest.

Major principles of electroacoustic behavior have been presented elsewhere in this text. It is requisite for the sound system designer to understand how sound radiates from loudspeakers, and how it interacts with the acoustic environment. Computer modeling attempts to model this interaction, and provides a valuable tool that allows the sound system designer to investigate various loudspeaker selection and placement scenarios. In this chapter I will examine some universal principles for the effective use of room modeling software. All simulations were performed with CATT-Acoustic (CATT-A)(www.catt.se).

38.12 Spherical Loudspeaker Data

Axial and polar measurements have long been used to characterize the performance of loudspeakers. Room modeling programs require spherical data for modeling the behavior of a point source, or group of point sources in a room. A spherical data set characterizes the loudspeaker’s directivity, because it includes data that was measured over a full sphere of measurement positions around the loudspeaker.

The author has built a spherical loudspeaker measurement system from scratch, and currently produces loudspeaker data files for many loudspeaker manufacturers for use in room modeling programs. The many years of invested time and money have provided some insights into what matters, and what doesn't with regard to loudspeaker data. "More" is not necessarily "better."

A modeling program assumes that the emerging wavefront is a sphere, and expands spherically as the sound propagates away from the source. This is only true in the far field of the source. The loudspeaker data must be measured in the loudspeaker's far field for the calculations that it is used for to be accurate.

38.12.1 Near Field vs. Far Field

The term "point source" has both theoretical and common-usage definitions in audio engineering. A literal point source is infinitely small. Directivity is achieved by interference. Since interference requires mass, a literal point source is omnidirectional and would emit the same sound pressure wave in all directions. The spherical waves fall off at the inverse-square law rate of level change as they propagate outward from the source. This means that when the radius of the sphere doubles, the area that the sound passes through quadruples. Since the same amount of sound energy is passing through a progressively larger area, the sound intensity level L_J and resultant sound pressure level L_P are reduced as the distance from the source increases. In room modeling, this behavior is simulated by rays that radiate outward from the point source.

A physically realizable loudspeaker has size and mass. The sound may not radiate evenly from all of its surfaces, so the wavefront may not be spherical near the source. This definitely true of multi-way devices as well as line arrays. Even though the wavefronts from these devices are not spherical when they are formed, the waves travel at the same speed in all directions. This means that they become increasingly spherical as the distance from the source increases. The distance from the source at which the waves can be considered to be spherical is the beginning of the far field. The inverse-square law applies from this distance outward. All loudspeakers obey the inverse-square law at remote distances. Note that "spherical wave" does not mean that the loudspeaker is omnidirectional. While the balloon shape is spherical in the far field, the sound is likely more intense in the axial direction due to the use of horns or wave guides. An omnidirectional source would produce the same sound intensity over the entire surface of the sphere.

A loudspeaker has a near field where the emerging waves are not spherical. It has a far field where they are. There is a frequency-dependent transition between the near and far fields. The shape of the loudspeaker's axial transfer function is distance-dependent in the near-field. It is independent of distance in the far-field, except for the frequency-dependent effects of air absorption.

There are both low frequency and high frequency criteria for the beginning of the far field. The point of observation must be:

1. At least one wavelength from the source at the lowest frequency of interest. This satisfies the low frequency criteria.
2. At least 10× the longest dimension of the source. This satisfies the high frequency criteria. This assumes that the high frequency sound energy emanates from the entire surface of the device. Often it does not, and the 10× criteria can be relaxed.

So, a 1 meter tall loudspeaker would have a far field that begins at approximately 10m, for frequencies above 30Hz. Ten meters is the acoustic wavelength of 30Hz. In practical cases this can be relaxed, and can be determined by measurement with a pull-away test that compares the axial response of measurements made at increasing distances from the source.

A practical distance limit for measuring spherical data is approximately 8m. This allows measurements above 43Hz for a device up to 0.8m (longest dimension normal to the axis). In practice, longer devices may be measured at 8m. The reasons include:

1. Most loudspeakers do not emit significant high frequencies from their entire frontal area.
2. The 10× distance criteria can be relaxed with an acceptable loss of high frequency accuracy, since for room modeling purposes, data is only needed through the 8kHz octave band.

38.12.2 Spherical Loudspeaker Data

The radiation properties of a real world loudspeaker must be determined by measurements made on the surface of the far field sphere previously described. The loudspeaker is placed in a free field—an environment that is free of acoustic reflections. A measurement microphone is placed on-axis and in the far field of the loudspeaker. The axial impulse response (time domain) or transfer function (frequency domain) is measured and recorded, and the loudspeaker is then rotated horizontally by the desired angular resolution, typically 5°, and the measurement repeated. This continues until the microphone is 180° off-axis. The series of

39 measurements is referred to as an “arc.” The loudspeaker is returned to the axial position, rotated 5° about its aiming axis, and another arc is collected. The process continues until a sufficient number of arcs have been measured to fully characterize the radiation from the loudspeaker. The exact number of arcs depends on the required number of quadrants, which in turn depends on the acoustical symmetry of the loudspeaker. Fig. 38-6 shows the measurement arcs around a loudspeaker.

The end result, using 5° resolution, is a set of about 2600 impulse responses. The IRs can be transformed to the frequency domain using the Fourier Transform to yield the transfer function, or frequency response magnitude and phase, for each measurement position. This data set is then processed into a set of loudspeaker balloon plots. There is one balloon plot for each 1/*n*-octave band. One octave resolution is generally used for room acoustics work. One-third octave may be used for loudspeaker coverage mapping and special investigations. Fig. 38-7 shows the coverage balloon for a directional loudspeaker.

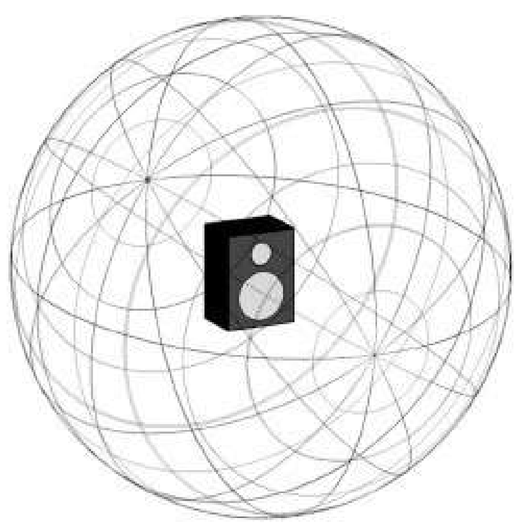


Figure 38-6. Grid for performing balloon measurements.

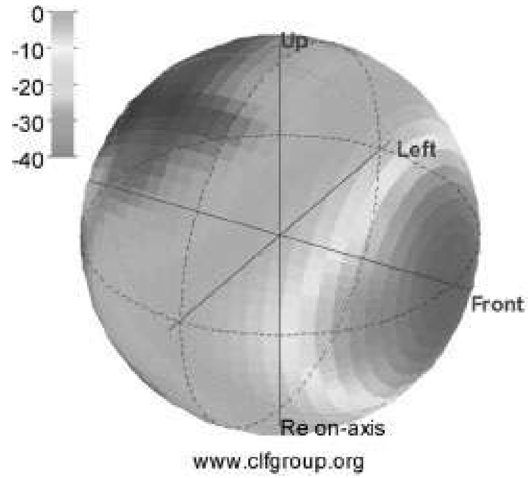


Figure 38-7. Loudspeaker radiation "balloon" (2kHz).

The balloon data is assumed to represent the directional behavior of the device at any distance, even though it was measured at 8m. This makes it inaccurate for the loudspeakers near field. In other words, while the data balloon shape measured at 8m may not be accurate for shorter distances, it is accurate beyond the measurement distance. If the data were actually measured at one meter, and this distance is in the near field, then it would become increasingly inaccurate as it is extrapolated outward. This measurement process is necessary to allow the balloon data to be accurately extrapolated to remote listener positions, which are typically more abundant in large rooms than seats near the loudspeaker. The 1 meter sensitivity (also measured in the far field and corrected to 1 meter using the inverse-square law) of the loudspeaker is used to scale the relative balloon data to an absolute level. The room modeling program extrapolates the balloon until it intersects with an audience plane, and weights the L_P using the balloon data. The resultant L_P is presented as a coverage map of the audience area, Fig. 38-8. The loudspeaker data file also includes the maximum rms voltage that can be applied to the loudspeaker. The level difference between this voltage and the voltage used to measure the sensitivity is used to calculate the maximum L_P possible for the device.

A loudspeaker measured in this manner is commonly referred to as a point source.

Figure 38-8. Direct field map of audience area in room model (CATT-A).

38.12.3 Loudspeaker Arrays

Small loudspeaker arrays, such as line arrays of 2m length or less, may be measured and modeled as point sources, since that is how they behave in their far field. Their near field behavior cannot be known if measured in this manner, but typically the bulk of the audience is in the far field, at least the ones that will experience poor *DRR* and *SNR*. Longer arrays may be assembled from multiple point sources. This requires that one of the sources be measured, and then multiple sources be stacked for a vertical array, Fig. 38-9. The relative arrival time of each can be computed for any listener position within the model. This allows the complex (magnitude and phase) interaction to be modeled and the pattern control to be visualized. Some modeling programs offer dedicated modules to facilitate array construction.

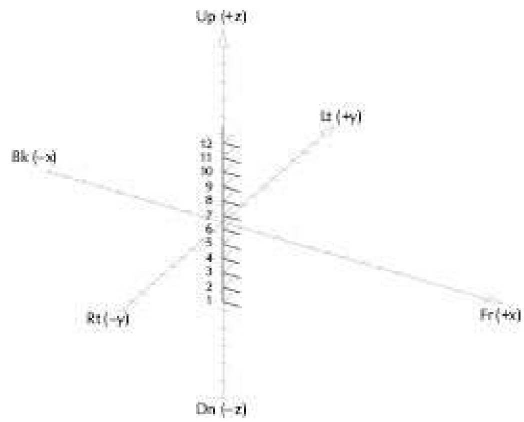


Figure 38-9. Array modeled from point sources (CATT-A).

A loudspeaker data balloon may consist of magnitude-only or magnitude plus phase data. Phase data can improve the prediction results for some array types, namely those that rely on complex interactions between the individual elements to form the radiation pattern, and are too large to measure as a single unit.

38.12.4 Direct Field Modeling

Direct field L_P and coverage predictions can be done without consideration of the room's acoustics. There is no need to build a complete, enclosed room model if all you need to know is the required mounting height and aiming angle for the loudspeaker to optimally cover an audience area. Modeling programs are extremely useful and accurate at showing how the spherical data balloon intersects the flat audience plane. This assures that an appropriate loudspeaker(s) has been positioned as to allow even direct field L_P over the audience area.

The direct field coverage map considers the L_P radiated from each point around the loudspeaker and the distance to the audience area. The resultant L_P is frequency-dependent, so independent maps are produced at the desired 1/n-octave resolution. The 1/n-octave bands can be summed and weighted to produce broadband L_P maps.

Coverage is not intuitive, and failure to model can lead to gross errors. Direct field coverage mapping should be the first step of any sound system design process, whether or not room acoustics modeling is performed.

38.12.5 Room Model Detail

Once the designer is satisfied that the direct field coverage is acceptable, the reflected energy from the room must be considered. A wire-frame model of the room geometry is constructed for this purpose.

The amount of detail required in the room model is the subject of ongoing debate. It would seem intuitive that an accurate visual model is an accurate acoustical model. This would also allow the convenience of using an existing CAD drawing as provided by an architect. Unfortunately this is generally not the case. A good visual model is not necessarily a good acoustic model.

The interaction of sound with room objects is quite complex, being a combination of reflection, resonance and diffraction. Ray-tracing or image/source methods can only approximate the behavior of sound, so they can never be described as "accurate" regardless of the detail used in the model. Too much detail can dramatically increase the calculation time, without increasing the accuracy of the predictions. The room model should be viewed as an acoustic "sketchpad," and is the acoustic equivalent to the architect's scale model made from foam board and paper. It is a highly programmable reverberation processor that can be tailored to the room's geometry while considering the directivity of the source.

The accuracy of acoustic predictions tends to follow the nature of the sound from the source. The direct field emitted by the loudspeaker can be measured with relatively high accuracy, and its behavior in the far free field can be estimated using the inverse-square law. This means that direct field measures, such as L_P and coverage are predictable. There should be excellent agreement between the various room modeling programs on the direct field, since its calculation is deterministic. Once the sound reflects, the term “accuracy” no longer applies. We are now dealing with approximations, as each ray that encounters a room boundary produces a new acoustic source that is as complex as the original source. The behavior of the sound becomes increasingly complex with each reflection order, eventually becoming diffuse and defying deterministic prediction completely. Fortunately, the needs of the sound system designer tend to track this accuracy progression. I can have high confidence in the direct field predictions and possibly a few reflection orders. But, the errors compound and the higher order reflections and the diffuse field behaviors are only estimates.

Modeling programs provide modules for creating the wireframe model, Fig. 38-10. Alternately they are done in a third party CAD program and imported. Each room boundary is given an absorption coefficient that determines the how much the L_P of the reflection is reduced when the ray encounters the boundary. It can also be given a scattering coefficient, which randomizes some or all of the reflected energy. Scattering coefficients are invaluable for estimating the behavior of complex room surfaces. Both absorption and scattering coefficients are estimates, and actual numerical values can be grossly affected by how they were measured.

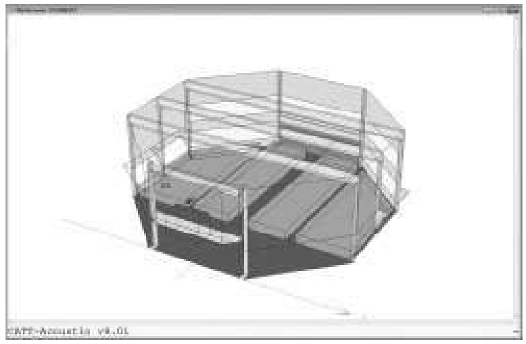


Figure 38-10. The wire frame room model provides a virtual design environment (CATT-A).

38.13 Predicting Room Reflections

Room modeling programs primarily use two methods to predict room reflections. An image-source method is deterministic, and can be visualized by considering the room surfaces to be mirrors. If you were seated at the listener position, the image of the loudspeaker would be visible in each boundary that produces a specular reflection for that seat. Similarly, in the real room if the source were replaced with a laser, and the images replaced with mirrors, the laser dot would end up on you. So, by using optics and geometry, the surfaces producing specular reflections can be identified.

The computational intensity of this method increases as the reflection order is increased. The accuracy decreases as the reflection order increases, since the errors compound. This suggests that a different approach is needed to simulate the late decay of the room. Raytracing and its variants (i.e. cone tracing) emit thousands of virtual rays from the source, traces them geometrically to a user-specified reflection order, and then counts the ones that arrive at the listener position. The listener in the model is actually a "counting balloon"—a target sphere of fixed or variable radius. The exact method of predicting the reflected field differs between modeling programs, as do the results. The method used may be proprietary to a modeling program, and is influenced by the knowledge, skills, intuition and prejudices of the programmer. The importance of the need to consider the modeling program's pedigree cannot be overstated. Don't buy into the marketing hype that acoustics modeling is a highly accurate process.

So, let's forget about "accurately" predicting the room's impulse response. From a sound system design perspective, what I need to identify are the room surfaces that will produce acoustic problems for listeners. This can be determined by constructing a simple room model, and then applying a scattering coefficient to surfaces that will not likely produce specular reflections, such as an audience area. The scattering coefficient randomizes the reflection from the boundary to simulate how a room surface with significant relief at the frequency (wavelength) of interest would behave. Individual modeling of each physical seat is both inaccurate and unnecessary. It is better to use a flat plane and assign an absorption and scattering coefficient. This facilitates the simulation of various audience sizes and distributions.

Computers make computations using the algorithms with which they are programmed. These algorithms, in turn, are formulated by humans based on some assumptions that simplify the modeling of acoustic behavior. For any given source/listener combination in an enclosed space, there are a myriad of variables that determine the room impulse response (RIR). The *RIR*, in turn, is the best summary that we have regarding what that seat sounds like.

The objective of the design process is to synthesize an approximation of the RIR. There are numerous similarities and parallels between the synthesized *RIR*, which I will refer to as an echogram, and the actual *RIR*. For one to be a good modeler, one must first be a good measurer. It is only through an appreciation of the sensitivities of measured data that one can grasp the difficulty of predicting the RIR. Only then can we avoid wasting time on the fine details and subtleties that many assume are accounted for by the mathematics used by the modeling software.

The measured *RIR* provides a reference for creating a virtual environment whose behavior emulates the actual room to the degree necessary to select and place loudspeakers. The following sections apply to both measured and modeled acoustical data.

38.14 Room Acoustics

The speed of sound vibrations through air is very slow compared to the speed of light. This results in a human-detectable time offset between the direct sound from the loudspeaker arriving at the listener and the reflections produced from room surfaces within the space. The art and science of room acoustics deals with these reflections.

The room is passive and produces no sound of its own. When sound is produced in the room from a source, a number of sound fields develop. Rooms are initially analyzed in the time domain since the arrival of reflections is function of time. For any listener position, there exists a ratio between the direct sound and the "room sound." This ratio determines how well the information from the sound system is conveyed to the audience. Let's look at how the room response is evaluated.

While the "frequency response" is a much more popular way of describing the sound from a system, the frequency response is determined by the time response—the complex interaction of the direct field and multiple reflections that are unique for each listener position. Most signal processing tools (i.e. equalizers) are "time blind" and therefore fail to address the real causes of poor sound quality. Since an equalizer affects all of the sound fields heard at a listener position, it cannot alter the ratio between them, which is the root of most sound clarity and speech intelligibility problems.

38.14.1 The Room Impulse Response

A hand clap in the space will produce a series of reflections at a listener position. Each reflection is a modified facsimile of the original event. This series can be broken down into several distinct sound fields. The hand clap is a crude *RIR*. In formal investigations the hand clap is replaced by methods that are calibrated and consistent. It is important to understand that regardless of the method used to collect it, the *RIR* is the most fundamental acoustic test. It is the primary means of analyzing the acoustic behavior of a room, and its synthesis is the ultimate goal of the design process. The room will have the same effect on any sound coming from the loudspeaker that it has on the impulse. A wireframe scale model of the room provides a virtual environment for sculpting the synthesized *RIR*.

The sound fields that the impulse produces include:

- The Direct Sound Field and its resultant level L_D.
- The Early-Reflected Sound Field and its resultant level L_{ER}.
- The Late-Reflected Sound Field and its resultant level L_{LR}.
- The Reverberant Sound Field and its resultant level L_R.

For pure acoustics work, the impulse may be a balloon pop or starter's pistol. For sound system work the stimulus may be pink noise or a sine wave sweep that is played through a loudspeaker, recorded and mathematically processed to yield the *RIR* by an analyzer. This technique allows the *RIR* to be collected without using an actual impulse and its attendant drawbacks. These include the requirement for a very quiet room as well as possible damage to the loudspeaker if the playback level of the impulse is too high.

In the same way that a sound source and receiver position are placed in a physical room to measure the *RIR*, a virtual source an receiver are placed in the computer model to predict the *RIR*. The room model serves as a virtual measurement environment.

A measured *RIR* is shown in Fig. 38-11. It is a time domain plot where sound pressure is the dependent variable and time is the independent variable. The vertical axis is linear.

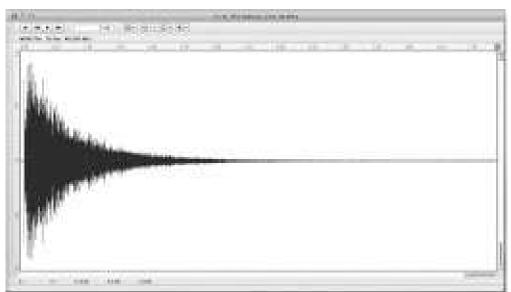


Figure 38-11. The Room Impulse Response displayed in a wave editor.

38.14.2 Post-Processing the RIR

As with other time domain amplitude data considered in the previous lessons, the absolute values of the amplitudes are displayed on a vertical axis of relative dB for observation and investigation. This is the log-squared *RIR*, Fig. 38-12.

Please pause and carefully consider the graph shown. It is the time domain representation of the relative levels of sound arrivals produced by an impulsive source placed at a unique position in the room, and collected at a different unique position. Since an infinite number of such positions exist, the investigator must select each based on what they are investigating. The answer to “Why?” determines “Where.” Strategically meaningful source and receiver positions are selected for both making room measurements and for evaluating the system performance in a computer model. Both are heavily influenced by answers to two questions:

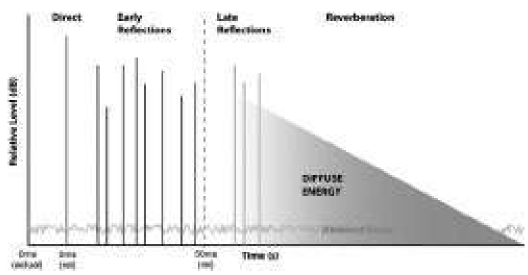


Figure 38-12. An idealized log-squared *RIR* showing the sound fields of interest to a system designer.

1. “Where can loudspeakers be placed?”
2. “Where will listeners be located?

A “measure” or “metric” is a score used to quantify some aspect of the *RIR*. One popular measure is the reverberation time (*RT*) in seconds.

38.14.3 Absorption

A sound absorber terminates the sound wave at a room surface by converting the sound energy into heat. A rating that describes the absorptiveness of a material is the absorption coefficient or alpha (α). The α is a number between zero and one, where zero is a perfect reflector and one is equivalent to an open window from which no sound returns. The α is frequency-dependent and usually specified at one-octave resolution.

In both the actual room and the room model, the α is multiplied by the surface area of the boundary to produce the *sabins* of absorption contributed by that boundary to the space. Each boundary can have only one α, and large boundaries can be subdivided if they have multiple surface coverings. One English sabin is one square foot of open window. One metric sabin is one square meter of open window. The more sabins, the less reflected sound. Sabins can be added or removed to modify the sound of the room.

If all room surfaces have α =1 for all octave bands, the room would be anechoic, or without echoes. An anechoic chamber utilizes heavy α to simulate this condition. In practice complete absorption is not possible and the criteria for an anechoic environment is that all reflections must be −20dB relative to the direct sound at the measurement microphone over the bandwidth of interest.

Absorptive materials have an acoustic impedance similar to air. Effective absorbers include soft, fuzzy materials such as fiberglass and mineral wool as well as some types of foams. Adding surface relief increases the surface area and yields more sabins for a given area. The ideal minimum thickness of an absorber is 1/4-wavelength at the lowest frequency of interest. This makes low frequency absorption difficult to accomplish for practical reasons due to the required material thickness.

Absorption coefficients are usually measured in a reverberation chamber. This accounts for the energy loss (or more correctly, the conversion to heat) of sound striking the material at random angles. The material's effect on a specular reflection with a specific angle of incidence may be quite different. Remember, these are estimates. Be conservative if you have to guess an absorption coefficient.

The absorption coefficient also plays a central role in the computer model. Here, the coefficient is a percentage, Fig. 38-13. Each surface in the room model is given an α, usually taken from tables of values created from actual measurements of various surface coverings, often with obscure origins where the details of how they were measured are rarely given. Practical values range from 1% to 99%. A practical resolution is 1/1-octave. While there is a push toward higher (i.e. 1/3-octave) a data, “more” is not necessarily “better” for several reasons. These include:

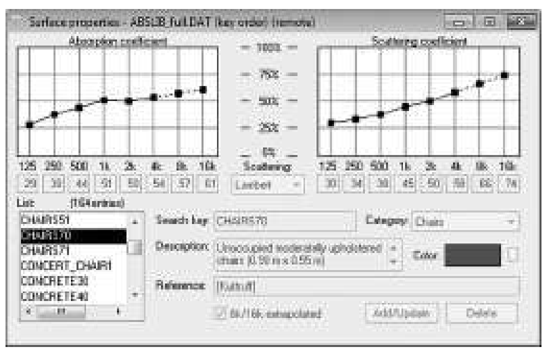


Figure 38-13. 1/1-octave absorption and scattering data as used by a room modeling program (CATT-A).

1. No matter how complete your materials library, there is always some educated guesswork involved in assigning an α to a room surface. 1/3-octave resolution complicates the guessing process, as I have to guess at 3 times as many values.
2. Sharp spikes in the absorptive characteristics of a material tend to support the case for higher frequency resolution, but such spikes (or dips) tend to be very sensitive to the variables that cause them. Even high resolution ABS data may not properly quantify the behavior of such a surface.

3. The predicted performance of both loudspeakers and rooms must be done at a practical resolution. It is a time-intensive task to characterize the direct field sound coverage and various performance measures at 1/1-octave resolution, let alone 1/3-octave.
4. Well-behaved loudspeakers do not have sharp directivity changes as function of frequency. One can visually interpolate the loudspeaker's performance between the octave bands. Ironically, the poorest loudspeaker designs can require the highest measurement resolution to characterize, as higher angular resolution may be required to resolve their erratic response. Is it better to increase the resolution to quantify such a device, or select a device that can be characterized by a lower angular resolution? The debate goes on.

38.15 Sound Behavior

When the wavelengths are short relative to the internal volume of the space, sound is modeled as rays of light that behave geometrically. The angle of incidence will equal the angle of reflection, as with a mirror. As frequency decreases, the light model becomes increasingly inaccurate and the sound must be modeled as a wave. The transition of a room from "ray behavior" to "wave behavior" is a gradient with no clear single transition frequency. The "ray assumption" is why we can use straight lines with arrows to indicate the direction of sound travel. One must always be mindful of the limitations of their assumptions. It's a big jump to consider the sound as a wave rather than a ray, which is why most acoustic prediction methods don't work at low frequencies which require wave methods due to their long wavelengths.

Room modeling programs assume geometric behavior of sound, treating it as a ray or particle that is emitted from a source. The Schroeder frequency estimates the transition region from “wave” behavior to “ray” behavior. In practical cases, the ray model for sound has validity down to possibly 100Hz, to be charitable. This means that the octave band centered at 125Hz is the lowest band that can be considered in the room model. At the other end of the spectrum, the 8kHz octave band extends to beyond 10kHz. The very short wavelengths beyond the 8kHz octave band do not predict well, due to their complex interaction with the room and the influence of air absorption and temperature gradients. They also have little if any influence on speech intelligibility. The critical frequency F_C provides an estimate for the transition of sound from wave behavior to ray behavior, Fig. 38-14.

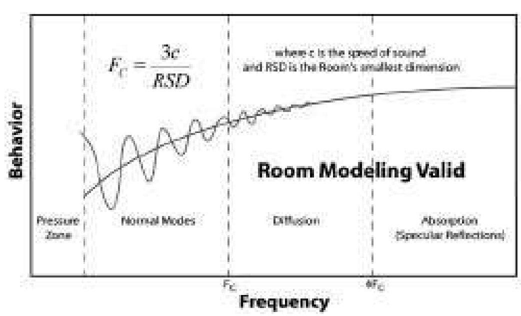


Figure 38-14. The audible spectrum subdivided based on critical frequency F_C.

For these reasons, sound system performance predictions are best limited to the 125Hz through 8kHz octave bands.

The following discussion assumes that the room size is very large relative to the wavelength of the sound, and therefore the sound behaves geometrically. This is usually the case for much of the audible spectrum for room sufficiently large to require a sound reinforcement system. The computer modeling programs used by sound system designers assume geometric sound behavior, and should therefore restrict acoustic calculations to the 125Hz octave band and above if acoustics is being considered.

38.15.1 The Direct Sound Field

The direct field is the sound that travels straight from the source to the listener. It arrives first since it has the shortest distance to travel. The direct field is the "engineered" sound field. Loudspeaker and amplifier selection are often based on it alone, as though the system were to be used outdoors. The spectrum of the direct field may be equalized by modifying the drive voltage to the loudspeaker with electronic filters—a process often referred to as "system tuning" and improperly referred to as "room tuning." The direct field is independent of the room, and direct field equalization can be performed off-site when certain conditions are met. Consideration of the direct field is Step 1 of the modeling process in the virtual room.

38.15.2 The Early-Reflected Sound Field

The early-reflected field includes the reflections that arrive close enough in time to be integrated, or mixed with the direct field by the ear-brain system. This integration time is frequency-dependent, ranging from a few milliseconds (ms) at high frequencies to tens of ms at low frequencies. 35–50ms is often used as a "one number" integration time.

Early-reflections increase the perceived level of the sound, and may be the primary means of amplification in a lecture or recital hall. They are often called "supporting" reflections. Acousticians often use “clouds” and “shells” to provide supporting reflections to listeners or musicians.

Early-reflections also produce tonal coloration as the reflected sound acoustically superposes with the direct sound at a listener. This can be either good or bad depending on the application. It is good sound system design practice to maintain some distance between loudspeakers and room surfaces to minimize coloration of the loudspeaker’s sound. Surface treatment can be substituted for distance, as can horn loading. The result is an initial time gap *ITG* between the direct sound and first reflection that can be observed on the log-squared impulse response (measured data) or echogram (predicted data). The presence of an *ITG* of 10 or more milliseconds can dramatically improve the fidelity of a loudspeaker, and is considered essential in critical listening spaces such as recording studio control rooms. It is equally important in an auditorium system, where the objective is to achieve a similar response from seat-to-seat. Strong very early reflections can make this impossible.

In the room model, the early-reflected field is determined by an image/source prediction algorithm. Rather than rely on the probability of sound striking a room surface, as does ray-tracing, the image/source method uses a deterministic approach. This becomes too computationally intense for high order reflections, so most modeling programs transition from image/source methods to ray-tracing in order to synthesize the complete room decay, Fig. 38-15.

38.15.3 Late Reflections or Echoes

Late reflections arrive beyond the ear’s integration time. They are perceived as a blurring of the sound or in extreme cases as echoes.

Hall designers use room geometry and acoustic treatment to control late reflections. Sound system designers utilize loudspeaker pattern control, placement and aiming to the same end.

In most cases, strong, late reflections are low in order. Common offenders include rear walls, balcony faces, windows, doors, etc. The predicted echogram can often identify the surfaces that are likely to be offending for a given listener position. As in physical rooms, the most likely problem spots are the first rows of the audience and the stage.

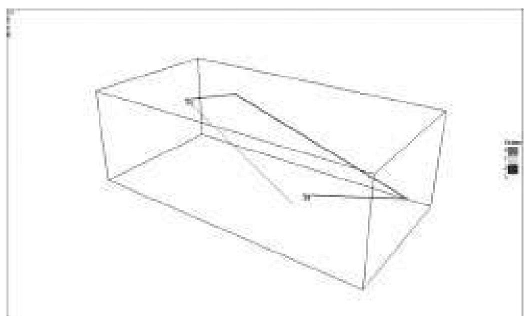


Figure 38-15. The path of a specular reflection between source and listener (CATT-A).

38.15.4 Reverberation

The reverberation time of a room is a measure of "sound persistence." By definition, the *RT* is the time required for the sound from an interrupted source to decay by 60dB. 60dB was chosen for largely practical reasons, as it represents the amount of decay that can be heard in a quiet space. The *RT* is more formally termed the RT_{60}. This has been shortened in later standards to T_{60} to be consistent with other time domain measures. A common measure is the T_{30}, which is the time required for 60dB of decay, extrapolated from the time required for 30dB of decay. It is described in ISO 3382 – “Acoustics—measurement of the reverberation time of rooms with reference to other acoustical parameters.” The *RT* is the easiest acoustic parameter to measure, and casual investigations require only pulsing the room and counting.

Reverberation time is frequency-dependent, with one-octave being a practical resolution.

Reverberant energy results from the sound persisting in a space long enough to produce so many reflections that the sound becomes diffuse and random. Contrast this with a reflection which had a definitive direction of travel and that can be attributed to a specific room surface. A diffuse sound field is noise-like, except that it is repeatable. The two characteristics of reverberation that are of interest are the reverberation time and the reverberation level. I have previously defined reverberation time. The reverberant level is the level that the reverberant sound builds to as the room is continuously excited by a source. A room may have a long reverb time, but if the reverberant level is made low by careful loudspeaker selection and placement, communication may not be impaired by the reverberant field. As one consultant puts it “The room won’t laugh if you don’t tickle it!”

The reverberation time can be estimated mathematically with the Sabine equation and its variants for rooms that meet certain criteria. These include low average absorption, uniform absorption distribution and a mixing geometric shape. Few rooms meet these criteria, so the reverberation equations only provide estimates. Due to its random and mixing nature, the reverberant field tends to be consistent throughout spaces where a significant reverberant field develops.

Reverberation is often used as a “catch-all” term for reflected sound, but it is important for sound system designers to consider the type of reflected sound as outlined here. Most rooms are semi-reverberant and all of the sound field types may exist at a given listener position. The investigator must determine which sound field is relevant to the problem being investigated.

Since the computer room model synthesizes the total echogram for a receive position, the T_{30} can be determined from this data. Of all of the acoustic measures, the T_{30} is the least “seat-specific,” meaning that it tends to vary less than the other measures as one moves around the space. It can serve as a general measure of sound persistence for a space, and it is meaningful to speak of working in a “3-second room.” The room modeling program shown has a special ray-tracing sub-program for quickly estimating the T_{30}. This is useful for achieving a general match between a measured T_{30} and the T_{30} of the model, Fig. 38-16.

38.15.5 Room Measures

The sound arriving at a listener position from a loudspeaker can be given ratings based on the previous sound field descriptions. This task is handled by an acoustic analyzer that collects the *RIR* and processes the data.

These include T_{30}, Early-Decay Time (*EDT*), Clarity (C50) and the Speech Transmission Index (*STI*).

Room acoustics is a field of study in and of itself. So is the measurement and characterization of the intelligibility of a sound system. Both are presented here as an introduction to familiarize you with their existence, and their importance in the sound system design process. Sound system practitioners must recognize the influence of room acoustics on the sound from their systems. It is possible, and even commonplace, to achieve adequate L_P and direct field coverage, only to have it swamped out by a high reverberant field level, or degraded by and echo. One objective of computer room modeling is to identify these problems at the drawing board.

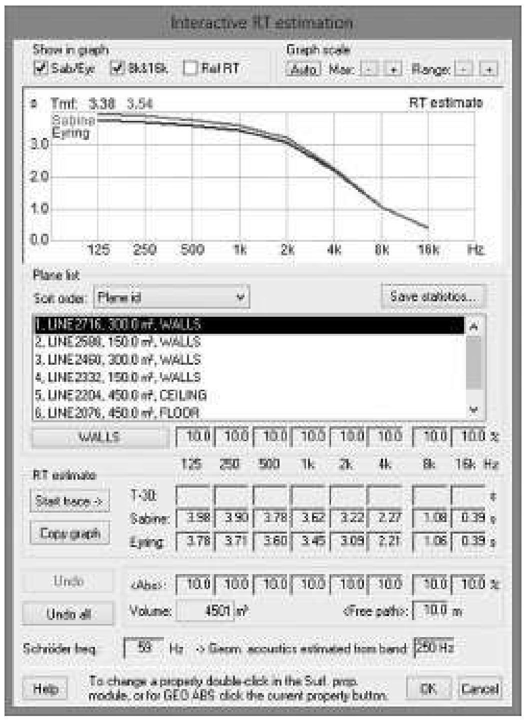


Figure 38-16. A utility for estimating the *RT* of a room model (CATT-A).

38.15.6 Qualifying the Room Model

Computer room modeling programs create a virtual acoustic environment to aid in loudspeaker selection and placement. It is important to build a bridge between the physical environment and the modeled one using measured data. Skipping this step can lead to wildly inaccurate predictions that over or under estimate the clarity of the system. Following is the method that I use to qualify the room model for the purpose of designing a sound system.

Step 1 – Field measurements

1. A well-controlled loudspeaker is placed on stage, well away from room boundaries (other than the floor). A properly measured data file (e.g. CLF) for this loudspeaker is mandatory and should be provided by the manufacturer. It is not so important what you use as that what you use is well-defined. The talker loudspeaker described at the beginning of this chapter is internally processed and powered, has an extremely smooth axial transfer function, and a well-defined directivity characteristic similar to a human talker. It is an excellent candidate for this process. Place it where a live talker might stand in the space, away from hard surfaces (ear height on a stand).
2. The furthest measurement position is selected, on-axis with the loudspeaker and usually near the back row of the auditorium but away from room boundaries (other than the floor). The *RIR* is collected at this position.
3. The distance is halved, and the measurement repeated.
4. The distance is halved again, and the measurement repeated.

This provides three axial positions for which I can easily determine the direct field level, *RT*, clarity, etc. They are related by the inverse square law, as the direct field level should diminish by approximately 6dB relative to the closest measurement for each subsequent measurement. This allows the sound behavior at in-between positions to be estimated if needed. The data can be analyzed in the measurement program, or listened to using convolution. Fig. 38-17 shows a program that analyzes the measured *RIR* and calculates the various acoustic metrics.

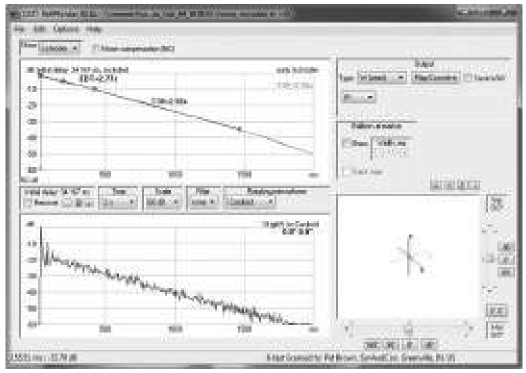


Figure 38-17. ReflPhinder™ software program (www.catt.se).

Step 2 – Wire-Frame Generation

1. A wire-frame model of the room's geometry is produced in the modeling program or other CAD environment. Each room boundary is assigned a color and a descriptor, i.e. wall, ceiling, floor, audience plane. These will be useful later for assigning absorption and scattering coefficients.
2. In the room modeling program, I initially set the a for all room surfaces to 0.1, or 10%. The *RT* is then calculated in the model and compared to the measured *RTs* from the room. 1/1-octave is sufficient resolution. Take note whether the predicted *RT* is higher or lower than the measured *RT*, as this indicates whether the average alpha must be increased or decreased.
3. Refine the model by estimating the absorption coefficients for the walls, ceiling, floor etc. These values may be taken from tables of measured data, guessed at, or both. Continue this process until the 1/1-octave predicted values are similar to the measured values.
4. Visually identify the room surfaces with significant relief and assign scattering coefficients based on depth of the relief to the wavelength ratio for the octave band of interest. If the relief

equals half the wavelength, assign a scattering coefficient of 0.5. Increase it by 10% for each successively higher octave and halve it by 10% for each successively lower octave, see Fig. 38-13. The purpose of this step is to "de-specularize" room surfaces that will not likely produce a specular reflections due to their complex shape. Audiences, pews and organ pipes are examples of room surfaces that will exhibit high scattering.

Step 3 – Correlation

We now have sufficient information to evaluate the correlation between the measured data and the room model.

1. Place a virtual source in the room model at the same position as the source in the physical room. Specify the data file that is correct for your test loudspeaker.
2. Place listener positions in the room model at the same coordinates as those used for the measured data.
3. Generate an echogram for each listener position.
4. Compare the Clarity-C_{50} for the measured and modeled data. Tweak the room model until the C_{50} scores are within a dB or two for measured vs. modeled, for each 1/1-octave band. This can take some time, and illustrates why 1/1-octave resolution is preferred to 1/3-octave for acoustics modeling.

Step 4 – Design the System

We now have a qualified room model for trying our design ideas. Substitute different loudspeaker makes and models at the positions you choose. Place additional listener seats as needed. Use coverage maps (overall) and echograms (seat-specific) to assure that the

Clarity is acceptable at all listeners.

A "Family" of Loudspeakers. There are hundreds of loudspeakers in the marketplace. How do I know which ones to try? I like to approach each design using a family of loudspeakers that are similar in design, but of increasing size and directivity, Fig. 38-18.

This gives me a logical progression for determining the required directivity for acceptable intelligibility. Start with the smallest one, and move up until you have acceptable performance at the furthest listener distance.

The point of this exercise is to identify the type of loudspeaker that can work in the environment. Once you know the sensitivity and directivity of an appropriate device, you can replace it in the model with a similar device from the manufacturer of your choice. For example, I may determine that the loudspeaker should be a large-format device, with a directivity index of 20dB and an average sensitivity of 100dB SPL. Most all major loudspeaker manufacturers offer such a device. Knowing what you are looking for dramatically narrows the search.

A Logical Design Process. Now that you have a qualified virtual environment to work in, you can set out to achieve your design objectives. Here is the process I used to arrive at the proper loudspeaker selections and placements.

Since the source is initially on the stage, the front-to-back coverage of the room will not be even.

Raise the loudspeaker to an appropriate trim height, re-aim it to the back row and regenerate the coverage map. The coverage should now be more even since the distance differential from the front to back rows is less.

Replace the source with one more appropriate for a sound reinforcement system. This typically means higher directivity and higher sensitivity. Start with the smallest member of your “family” of loudspeakers. Remap the coverage and change the aiming until the audience is covered as evenly as possible. A good target criteria is no more than 6dB differential over the audience area. Map the clarity and evaluate it.

Continue to increase the directivity until the clarity and intelligibility is acceptable initially at the axial position, but ultimately over the audience area. If you can achieve this with a single source, great. If not, then you may need to break the audience up into zones and start over, assigning one loudspeaker per zone.

You might also consider other approaches, such as the use of line arrays or pew back systems. Your modeling environment will allow you to try these and others, comparing them with the initial “single source” approach. Ultimately something will win out due to performance, aesthetics, cost or some combination of the three. Let the customer decide. Your job is to present some designs that can work. Many approaches are possible, but these steps will get you started. Never lose site of the objective—to extend the face-to-face listening experience to all listeners in the space.

Once the loudspeaker types and placements have been determined by the modeling process, the necessary amplifiers and signal processing can be determined. A valuable resource for amplifier selection is the Common Amplifier Format (CAF). A freeware amplifier-sizing calculator is available at www.cafgroup.org.

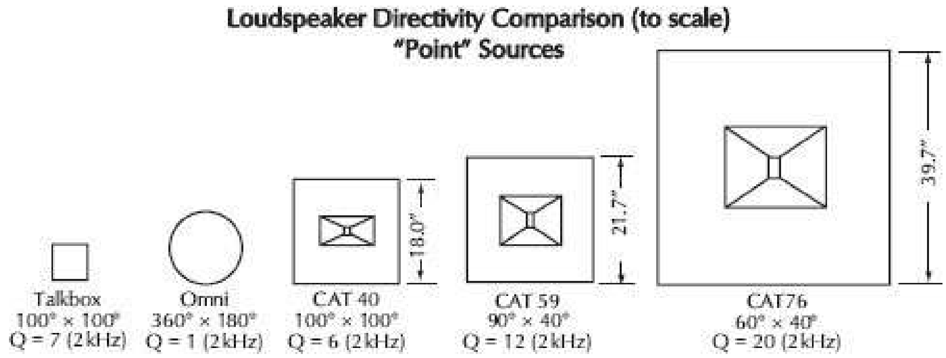


Figure 38-18. A family of point source loudspeakers with increasing directivity. Courtesy Mitchell Acoustic Research, (www.frazierspeakers.com).

The Final Step. An often overlooked step in the design process is comparing the measured performance of the finished system to the performance of the modeled system. This requires that you collect some *RIR* measurements of the finished system at the listener seats that you used in the room model. If all has gone as planned, and you have been conservative in your estimation of absorption coefficients, the measured performance should have higher clarity and intelligibility than the modeled performance. So, we are not trying to accurately predict the response of the system. We are trying to make sure that if the worst case materializes, the system still works. It is better to estimate the *RT* at 3s and it turn out to be 2s, than vice-versa.

By correlating the measured response with the predicted response, you will gain valuable insight into the nuances of modeling, and will be better equipped for cases where the room has not yet been built. Experience is a great teacher.

References

1. Peter Svensson, *The Early History of Ray Tracing in Room Acoustics*.

2. Davis, Patronis, Brown, *Sound System Engineering*, Focal Press, 2013.

3. Bengt-Inge Dalenback, *CATT Acoustic Manual*, www.catt.se.

Chapter 39

Computer-Aided Sound System Design

by Dr. Wolfgang Ahnert, by Stefan Feistel, and Hans-Peter Tennhardt

39.1 Introduction

For more than 2000 years acoustic phenomena have been perceived and manipulated subjectively. Reference can be made in this context to Marcus Vitruvius, the ancient Roman architect who described at this early time the application of acoustic laws in theatrical spaces. But only since the end of the medieval times and particularly during the last century acoustics has developed into an independent science.

Highlights on the way to a scientifically calculated design:

- Roman/Greek Times/Medieval Times: knowledge based on experience and first trial and error reports—e.g., by the Roman Architect Vitruvius 15 BC since the end of 18th century: theoretical investigations, e.g., Chladni 1810 or in 1875 Lord Rayleigh, Prof. Helmholtz.
- Since 1900: room acoustical basics, Prof. Sabine/USA 1923; radiation of sound, H. Stenzel/Germany 1930, and H. F. Olson/USA1947.
- By 1935: measurement in models and "Auralization" in physical models, Prof. Spandöck München, Prof. Reichardt, Dresden/Germany.
- Since 1965: computer-model investigations, Prof. Krokstad, in Trondheim/Norway, afterwards many similar works have been done.
- Since 1995: Auralization by means of computer models has been introduced.

In Fig. 39-1 a measured sound-field structure in three-dimensional, so-called waterfall form (decay of sound energy as a function of time and frequency) is shown. The sound level is marked on the ordinate, the frequency (range of frequency 63Hz–8kHz) on the abscissa and the time on the third axis (0ms–direct sound to 3.5s reverberation).

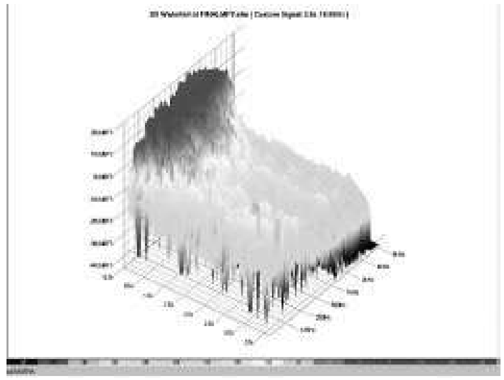


Figure 39-1. Typical waterfall presentation.

These sound field structures are depending on listener locations. In the old days a wanted sound decay for concerts or for speech transmission was generated by changing the primary or secondary structure of a room (see Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, section 9.3.2). Now with sound systems it is possible to generate any sound fields subjectively desired.

Today we can derive the basic items in sound design. Listening comfort and intelligibility are influenced by:

- Reverberation time and volume.
- Early and late reflections.
- Ambient noise level.
- Directivity of the loudspeakers, new array constructions or beam forming.
- New loudspeaker types.
- Interference effects.

Some basic measures of the performance of a sound system are:

- Intelligibility STI, Alcons etc.
- Loudness in dB(SPL_{tot}).

- Direct sound in dB(SPL_{dir}).
- Frequency response in ±dB from flat.
- Coverage in ±dB from even.

The goal of modern sound system design is to calculate in advance the complete sound field structure in a hall or in open spaces by means of Computer-Aided Acoustic Design (CAAD) programs enabling you to prevent any "surprises" that become evident after a sound system has been installed. You just describe in advance the expected properties of your sound system and the overall acoustic properties of your room using the new sound system.

In the following considerations there are included personal experiences of the authors, especially with the CAAD program EASE,[1] but features of other programs are also explained.

39.2 Sound Design Basics for Acoustic Simulation

Application of physical or computer models for acoustic and sound system design:

Prior to 1965. Physical models since the 1930s and after WWII mainly in selected cases in research centers by means of huge computers.

Since 1970. The programmable pocket computer offered the first algorithms for acoustics.

1981. The PC and PC-XT have been available.
1984: Reverberation time and intelligibility calculations for simple rooms:

CADP / JBL.

TEF-10 Analyzer: First coverage plots became reality.

1985: PHD Program. By Prohs/Harris in spring: first version for TEF Analyzer:

1. Room-acoustic calculations like different reverberation times.
2. Loudspeaker cluster design.
3. Power calculation for horn radiators and corresponding drivers.
4. Alcons calculations by Peutz.

1986: BOSE-Modeler. First full-graphic CAD Macintosh-based program, Version 1, 1986 by K. Jacob, T. Birkle/Bose/USA.

1987: Acousta-CADD. First full-graphic CAD MS-DOS-based program, Version 1 by A. Muchimaru, Altec Lansing/USA.

1990: EASE. Full-graphic CAD MS-DOS based program with pop-up menus Version 1, 1990 by ADA, Germany.

1991: CADP2. Full-graphic CAD Windows 3.1 based program, by JBL/USA.

1996: ULYSSES. By IFB/Germany (P. Hallstein).

1997: CADP2. Further development stopped.

1999/2001: EASE for Windows. By Ahnert Feistel Media Group, currently version 5 in preparation.

Room Acoustics Programs including Sound System Design Features

1988: ***CATT-Acoustic***. Dalenbäck/Sweden, version 1, 2013 version 9.0c.

1991: ***ODEON***. Naylor & Rindel/Denmark, version 1, 2013 version 12.0.

1994: ***RAMSETE***. Farina/Italy, version 1, 2013 version 2.7b.

2002: ***EASE***. Ahnert & Feistel/Germany, EASE version 4.1, 2013 version 4.3.
The programs printed in italics are subject to constant advancement.

39.2.1 Measurement and Planning Methodology with Physical Models of Large Auditoriums

by Hans-Peter Tennhardt

39.2.1.1 Fundamentals

The room impulse response is obtained in a reduced model of the auditorium interior by applying the corresponding scale-model laws based on the constant ratio between the geometrical dimensions *L* of the room and of the sound wavelength λ in the model scale (index *M*) and in natural scale (index *N*):

$$\frac{L}{\lambda} = Const = \frac{L_N \cdot f_N}{c_N} = \frac{L_M \cdot f_M}{c_M} \qquad (39\text{-}1)$$

where,
c is the speed of sound,
f is the frequency.

If the scale-model test is carried out in the same sound propagation medium, there results $c_N = c_M$ and thus Eq. 39-1 becomes

$$p = \frac{L_N}{L_M} = \frac{f_M}{f_N} \tag{39-2}$$

i. e., the measurements are carried out in a frequency range that exceeds the original frequency range by the factor p (reduction scale 1:p).

A favorable compromise regarding model size and accuracy of reproduction is given with a reduction scale of 1:20, but scales between 1:8 and 1:50 are feasible depending on model size or frequency range to be studied. The sound impulse is irradiated from the location of the sound source (e.g., stage, orchestra pit, loudspeaker). The acoustical response of the room to the emitted signal is simultaneously registered at receiving positions (audience area, platform, stage) by special electroacoustic transducers (microphone, dummy head with ear simulator). The transfer function between transmitting and receiving locations is calculated from the obtained room impulse response. Very often a spark discharge generator is used as a sound transmitter in air (nowadays electronic MLS scale radiators too). With an impulse width in the model of 80 μs, it is possible to resolve path differences equaling 60cm in the original room. The reproducibility of the maximum sound pressure is within ±0.2dB.

Special model sound sources enable simulation of a talker or a singer, of an orchestra as nondirectional sound source in the center of the same, of orchestral instrumental groups (see section 39.1.2.2) and of loudspeaker lines with variable directivity characteristics.

Preferably the registration of the room impulse response is dual-channel by the microphones of the dummy head at listener seats which are representative for determined seating groups so that the binaural head-related listening parameters of the human auditory organ are optimally reproduced. In a model of scale 1:20 the diameter of this miniature dummy head must be about 11mm, Fig. 39-2.

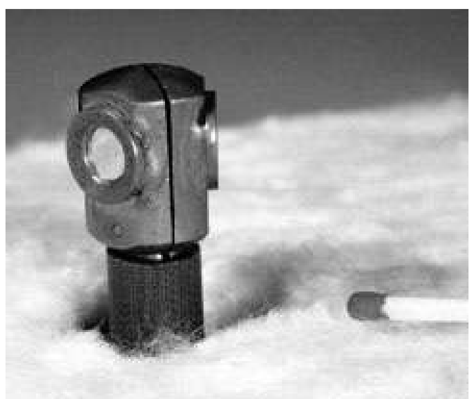


Figure 39-2. Model-size dummy head for measurements at a scale of 1:20.

The investigated frequency range lies between 5kHz and 200kHz in the scaled model which corresponds to 250Hz to 10kHz in the original. Structures whose linear dimensions fall below approximately 8cm in the original room are not reproduced in the model. Also sound absorbers and wall impedances below the studied frequency range are not considered in the model tests.

For better access to the models during the measurements, these should be carried out in air under normal pressure. Owing to the excessive atmospheric absorption occurring at the model frequencies there occurs a faulty momentary value of the sound pressure which is mathematically corrected by a real-time compensation. Without these mathematical corrections the measurements must be carried out in a nitrogen environment instead of air, in which case the drawback consists, of course, in the bad access to the model.

All physical acoustic phenomena, e.g., diffraction and scattering, are represented in a frequency-true fashion.

The obtainable accuracy is at the time being still superior to that of a computer simulation. The method is capable of providing answers to questions concerning balance investigations in rooms for music performances (see section 39.1.2.2), the influence of electroacoustical components on room-acoustical parameters, and the directional effect of wall and ceiling structures prepared as scaled models.

By using original sound source simulations (talker, singer, nondirectional sound source, orchestra instrumental groups, loudspeakers) and a dummy head as receiver unit, the described measuring procedure is applicable also for original rooms.

39.2.1.2 Balance Investigations of Music Performances

The scaled-model simulation of an orchestra can in first approximation be realized by a nondirectional sound irradiation from the center of the same. A more detailed simulation is necessary, however, if one wants to have information about the influence of a room on the balance of the different instruments at the listener's seat. A useful approximation can be obtained already by a simulation of orchestra instrumental groups in which their sound spectra are based on the frequency response chiefly reproduced in music presentations and their directional characteristics are derived from the usual playing posture see reference 2.

The simulated orchestra is subdivided into four instrumental groups, in which the percussion instruments, in view of their considerable loudness and adaptable style of playing, may be left out of consideration:

- String instruments St.
- Woodwind instruments Wo.
- Brass instruments Bl.
- Bass instruments Ba.

To this may be added the electroacoustical model transducer of a singer/talker (S).

The scaled-model simulation comprises the impulse excitation by a spark-gap generator provided with a shading reflector of defined sound attenuation so as to align it with the directional characteristic of the instrumental group in question, see references 3 and 4. The electroacoustical transducers are positioned in the center of the group in question according to the arrangement topography of the orchestra, Fig. 39-3.

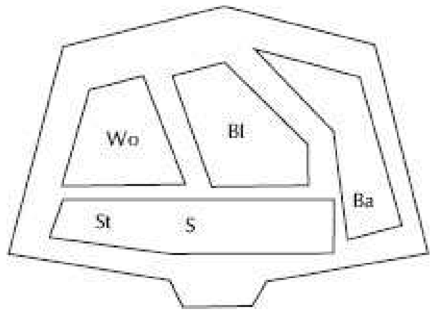


Figure 39-3. Typical arrangement of simulated orchestra-groups on a concert hall platform.

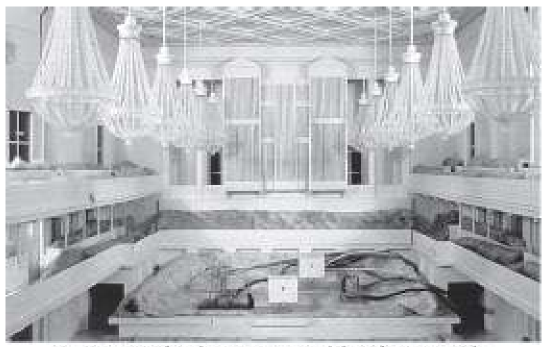
By evaluating the measured binaural room impulse responses on the basis of the register balance measure BR, see Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, section 9.2.2.15, it is possible to infer room-acoustical measures for a required architectural modification of the horizontal and vertical boundary surfaces of the performance zone and to clarify questions concerning the vertical staggering of the orchestra. The sound intensity-time behavior allows conclusions regarding the sound attack of the individual instrumental groups and regarding masking effects in the frequency and time domains from which measures for the acoustical formation of the secondary structure of the room can be inferred.

Fig. 39-4A shows an example for the use of the scaled-model measuring technique for a concert hall, as compared to the original room, Fig. 39-4B.

39.2.2 Building a Computer Model, Entering Room Data

To enter room data into a simulation program must be simple and straight forward. A combination of graphical and numerical entry of the data, planes and vertex points has to be supported. This way of entering the room into the program must be efficient in order to make the program work cost-effective and intuitive. If the room entry takes too long, the program becomes much less valuable as a real design tool. There are different ways to enter room data:

- By x, y, z Coordinates.
- By text files.
- By import from professional drawing programs like AutoCAD or SketchUp.



A. Acoustical indoor room model scale 1:20 with simulated orchestra groups.

B. Finished original room.

Figure 39-4. Konzerthaus, Berlin.

- By use of advanced drawing tools like prototypes or predefined room shapes.

In Fig. 39-5A to D we show one example of a model with different view options.

Simpler models which normally should have between 500 and a maximum of 1500 faces may be created based on simple room shapes or prototypes. A manipulation routine should allow to stretch or shrink dimensions to adapt the prototype to the requirements. This way a simple room model for basic investigations can be created within minutes.

Better would be the possibility of importing DWG or DXF or other similar architecture files directly. But the disadvantage here is that architects in the early design phase do not create 3D-models and offer only 2D-drawings. These drawings are of less use and so the acoustician has to enter the model vertex-by-vertex, line-by-line and area-by-area. Sometimes 3D-models can be built by extruding a 2D-plan and by manipulating the result.

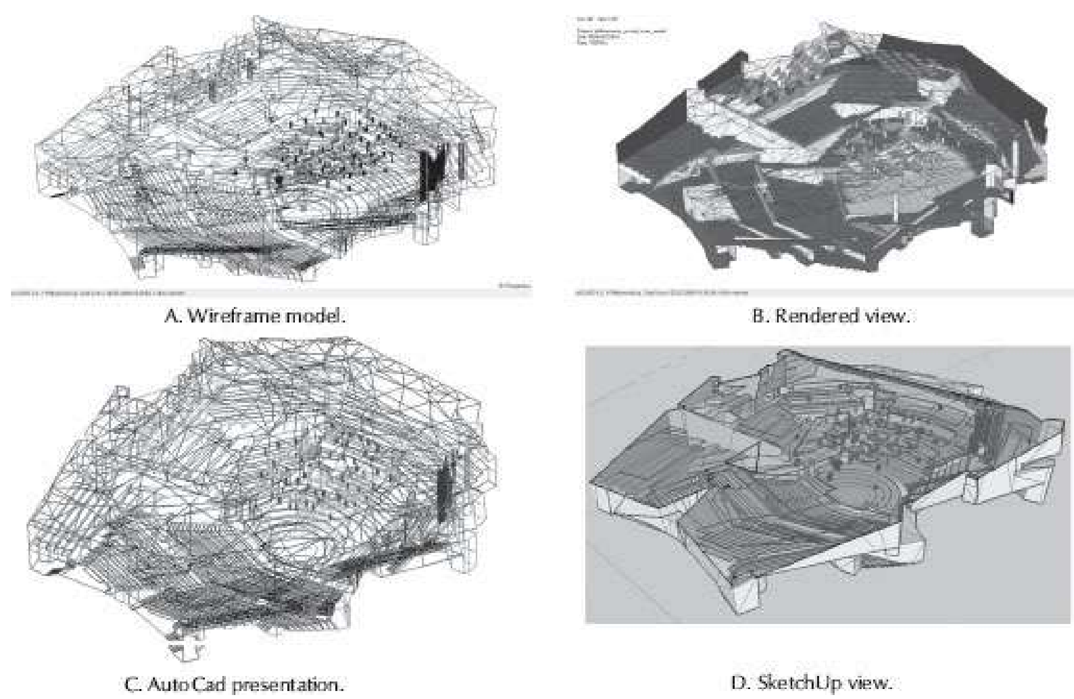


Figure 39-5. Same room model in four different views.

39.2.3 Acoustic Sources and Loudspeaker Systems

39.2.3.1 Natural Sources

Sound Reinforcement quite often has to deal with reinforcing natural sources like human voices or natural music instruments. Therefore to know the amount of reinforcing we have to know the quality of the reference sources and under following aspects:

- Loudness.
- Frequency response.
- Directivity.

Fig. 39-6 shows the level and frequency range of natural sources and instruments perceived by human beings. In this range natural sources develop their sound power and produce level components in the mentioned frequency range. Everything outside this range is on one side masked by noise (lower than 30 dB in the midrange) or dangerous for our health (pain threshold ≈120dB). Frequency components lower than 25Hz are becoming inaudible as well as frequency parts above 15–20kHz—depending on age and health.

Natural sound sources like human voices or musical instruments do not radiate sound in an omnidirectional way. It comes close to it in the case of a human voice, but the higher the frequencies the higher the head becomes an obstacle to the sound radiation backwards. Fig. 39-7A shows the directivity balloon curves of a female voice in the vertical domain. We recognize until 1000Hz an omnidirectional radiation pattern, but for higher frequencies the radiation dominates more and more in front of the head. That is also the reason that in concert halls where the audience is also behind the orchestra sometimes complaints are made about the singer's clarity.

The radiation behavior of musical instruments is much more complex. Here a lot of investigations have been done, especially by Meyer.[6] Fig. 39-7B shows the three-dimensional presentation of the directivity balloon of a horn instrument including the player. According to Meyer also the shadow effect of the player himself is considered. We observe with increasing frequencies a reduced radiation into the front domain of the player.

To model all these different natural sources right, their radiation behavior has to be known and this not only for single instruments but also for groups of them. Here a lack of corresponding data is still evident.[7]

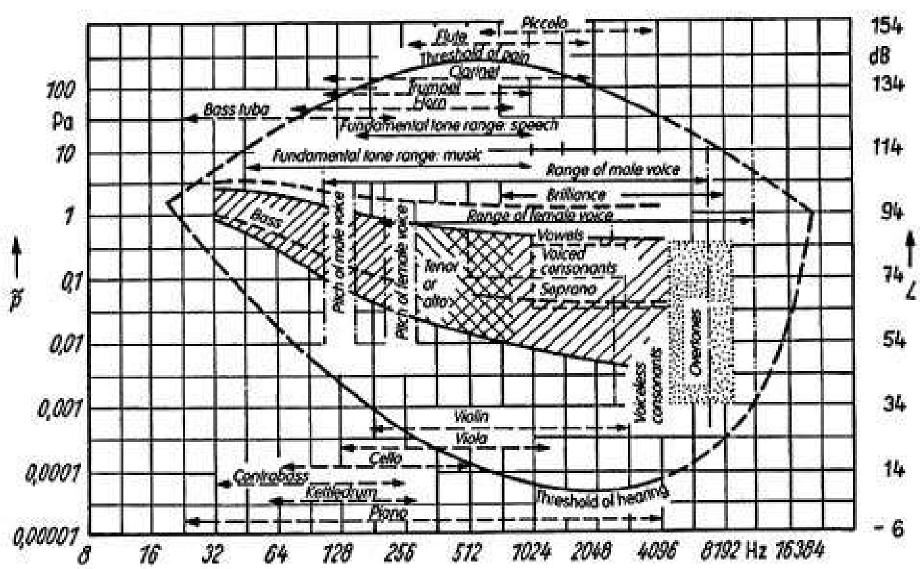


Figure 39-6. Level and frequency range of natural sources and instruments. Courtesy V. O. Knudsen 1932.[5]

39.2.3.2 Loudspeaker Types

We distinguish the following loudspeaker types:

- Point sources.
- Loudspeaker columns.

- Cluster.
- Line arrays.
- Digitally controlled sound columns.

For the use of these different sound radiators, the performance parameters of the same must be known. One will soon note, however, that the performance parameters specified by the manufacturers vary in accuracy and scope. For nearly two decades the Standards Committee of the AES updated rules and standards for a uniform approach in this respect.[8] When studying the data sheets of diverse manufacturers one will nevertheless note considerable discrepancies allowing the expert to draw conclusions as to the quality of the data given. For this reason we are going to mention the most important data to be specified in loudspeaker design. Let us start with the so-called point sources.

39.2.3.2.1 Point Sources

Point sources do not show automatically omnidirectional radiation behavior. Their directivity behavior is measured on a turntable and all directivity balloon data are referred to the point of rotation, therefore the name point sources.

Transfer Behavior. The nominal *load capacity* P_n of this loudspeaker type is the rms electrical power specified by the manufacturer according to the design characteristics.

The ratio between the sound pressure $\tilde{p}$ and the voltage $\tilde{u}$ required to attain this capacity at the radiator is called *sensitivity* T_s.

$$T_s = \frac{\tilde{p}}{\tilde{u}} \tag{39-3}$$

One distinguishes between a *free-field sensitivity Td* and a *diffuse-field sensitivity T_r*. The free-field sensitivity is normally indicated for a reference point on the reference axis at a distance of 1m from the loudspeaker. This can be expressed by

$$T_d = \frac{\tilde{p}_d\, r}{\tilde{u} r_0} \tag{39-4}$$

The diffuse-field sensitivity has to be ascertained in a diffuse field, for instance in a reverberant chamber. In order to eliminate the room property characterized by the equivalent absorption area of the room, a correction factor has to be used:

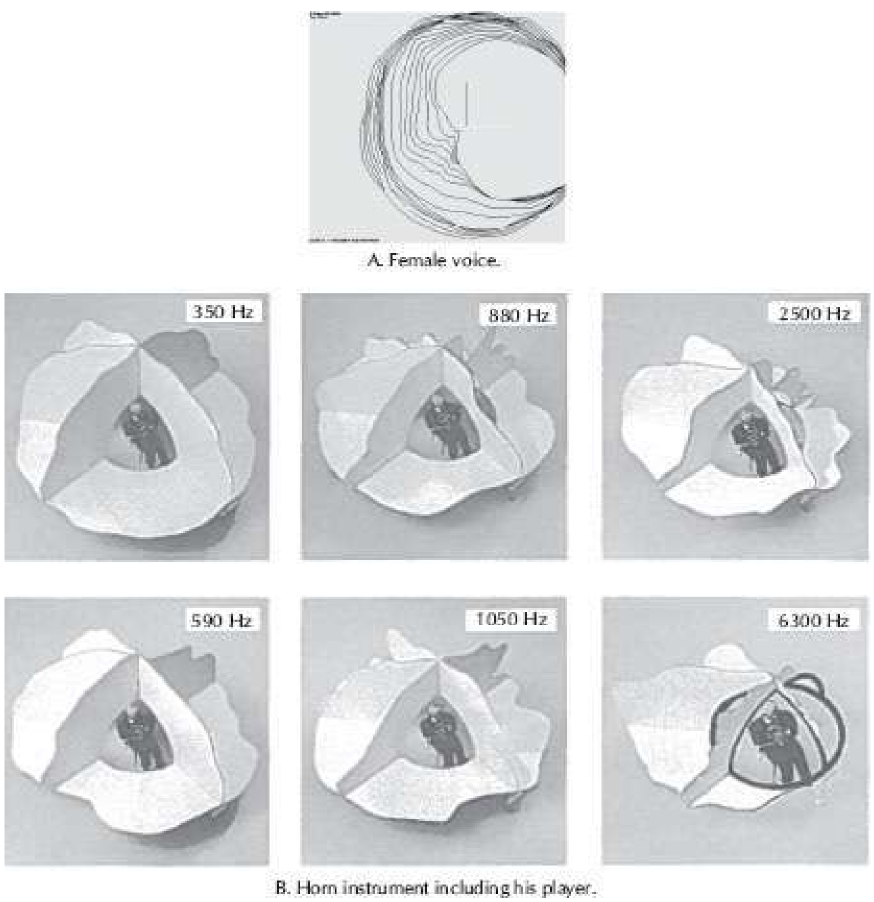


Figure 39-7. Directivity pattern of a natural source and an instrument.

$$T_s = \frac{\tilde{p}}{\tilde{u}} \tag{39-5}$$

where,
T_s and T_d or T_r are given in Pa/V.

The *sensitivity level* G_S, is defined as logarithmic quantity of sensitivity:

$$G_S = 20\log\frac{T_S}{T_0}\,\text{dB}$$

The reference sensitivity T_0 is preferably 1Pa/V. If another value is chosen, it has to be indicated.

The graphic representation of the sensitivity level as a function of frequency is called *frequency response*.

One of the quantities most frequently used in sound reinforcement engineering is the rated or characteristic sensitivity. In combination with the nominal load capacity it serves, among other things, to ascertain the maximum achievable sound pressure in the main reference axis of a loudspeaker or a loudspeaker system. According to the definition given in the Standard DIN 45570,[9] AES2-1984 (r2003),[8] and IEC 60268-5[10] it is the ratio between the sound pressure p_d averaged over a determined frequency range (mostly 250 to 4000Hz) and measured on the reference axis at a distance of 1m from the reference point of the radiator (usually the point of rotation during measurements), and the square root of the power supplied. By the standard, this power is referred to the *nominal impedance* Z_n of the radiator ($P_s = \tilde{u}^2 / Z_n$). Thus the *rated sensitivity* is

$$E_K = \frac{\tilde{p}_d}{\tilde{u}}\sqrt{Z_n \frac{r}{r_0}} = \frac{\tilde{p}_d\, r}{\sqrt{P_S}\, r_0} \quad (39\text{-}6)$$

where,
$\tilde{p}_d$ is direct sound pressure,
r is distance to the loudspeaker on the main axis,
r_0 is 1m distance.

Because of its reference to power, this expression is also designated as rated power sensitivity. According to IEC 60268-5, the logarithmic quantity of this expression is called *characteristic sound level* L_K, but also *sensitivity*/dB. It is defined by

$$L_K = 20\log\frac{E_K}{E_{K0}}\text{ dB} = 20\log E_K \text{ dB} + 94\text{ dB} \quad (39\text{-}7)$$

An important parameter for approximating the sound-field conditions in rooms is the *front to random factor γ* It characterizes the relationship between the acoustic power that would be radiated into the room by an omnidirectional loudspeaker having the same free-field sensitivity as the real loudspeaker to be assessed, and the acoustic power of the real loudspeaker:

$$\gamma = \frac{\int_S \tilde{p}_0^{\,2}\, dS}{\int_S \tilde{p}^2(\vartheta)\, dS} = \frac{S}{\int_S \Gamma^2\, dS} \quad (39\text{-}8)$$

where,
$\tilde{p}$ is sound pressure ($\tilde{p}_0$ is measured in the main front direction),
S is the globe surface around the loudspeaker,
ϑ is the room angle,
for Γ see Eq. 39-17.

A measuring procedure for ascertaining the *front to random factor* was established in the IEC publication 60268-5:[10]

$$\gamma = \left(\frac{\tilde{p}_d\, r}{\tilde{p}_r r_H}\right)^2 \qquad (39\text{-}9)$$

where,
$\tilde{p}_d$ is direct sound pressure,
$\tilde{p}_r$ is reverberant sound pressure,
r is distance to the loudspeaker,
r_H is the critical distance in the diffuse sound field, see Eq. 9-10.[11]

The *directivity factor Q(ϑ)* is often used for this term, but it is a function of the angle ϑ, see Eq. 39-20.

The logarithmic quantity of the *front to random factor is the front to random index*

$$C = 10\log\gamma\,\mathrm{dB} \qquad (39\text{-}10)$$

It corresponds to the difference between the free-field and the diffuse-field sensitivity levels:

$$C = G_d - G_r \qquad (39\text{-}11)$$

where,
G_d is the sensitivity level in the direct-field,

G_r is the sensitivity level in the diffuse-field,

or expressed by the sound levels measured at 1 m distance in the direct field of the loudspeaker (L_d) and in the diffuse field (L_r) of a room having the reverberation time RT and the volume V,

$$C = L_d - L_r + 10\log\frac{RT}{V}\,\mathrm{dB} + 25\,\mathrm{dB} \qquad (39\text{-}12)$$

where,
L_d is the direct sound level,
L_r is the diffuse sound level,
RT is the reverberation time in s,
V is the volume of the room.

An equal input power P_{el} is taken for granted.

Because of the dimensions of the radiators, the wavelength of the radiated sound in the lower frequency range is long compared to the radiating surface. Because of this difference there results only an insignificant directivity. With rising frequency the relationship changes and directivity increases.

For sound reinforcement purposes it has been proven in practice that slight increases of approximately 3dB/octave of the front to random index of the loudspeaker system are appropriate, because most of the natural sound sources show a similar increase giving rise to a corresponding timbre change.

For approximate calculations or also measurements of the front to random index one has to cover at least the range between 500 and 1000Hz. Many manufacturers are indicating, in the data sheets of their products, the frequency dependence of directivity.

By means of the front to random index C and the nominal power rating P_n it is also possible to describe the characteristic sound level of a loudspeaker system:

$$L_K = L_W + C - 10\log P_n - 11\,\text{dB} \qquad (39\text{-}13)$$

where,
L_W is the sound power level.

The *efficiency* η of a loudspeaker system is determined by the ratio between radiated acoustic power and supplied electric power:

$$\eta = \frac{P_{ak}}{P_{el}} = \left(\frac{E_K^2}{\rho_0 c} \times \frac{4\pi r_0^2}{\gamma_L}\right) \times 100\% \qquad (39\text{-}14)$$

where,
P_{ak} is the acoustic sound power,
P_{el} is the electric power applied,
E_K is the sensitivity of the loudspeaker,
r_0 is 1m distance,
γ is the front to random factor of the loudspeaker,
ρ_0 c is the characteristic acoustic impedance of air = 408Pa s/m^3 at 20°C.

By combining all constants one obtains the following approximation:

$$\eta = 3\frac{E_K^2}{\gamma_L}\% \qquad (39\text{-}15)$$

This correlation can be seen in Fig. 39-8. The efficiency of loudspeaker systems lies in reality between 0. 1 and 10%. As is the case with the rated sensitivity, the efficiency is often referred to the nominal impedance Z_n of the loudspeaker and designated as *nominal efficiency* η_η.

$$\eta_n = \frac{\tilde{p}_d^{\,2} Z_n}{\gamma_L \tilde{u}^2} \times \frac{4\pi r^2}{\rho_0 c} \qquad (39\text{-}16)$$

where,

$\tilde{p}_d d$ is the direct sound pressure,

Z_n is the nominal impedance,

γ_L is the front to random factor of the loudspeaker.

Eqs. 39-14, 39-15, and 39-16 suggest that because of the frequency dependence of the front to random index and the insignificant frequency dependence of the free-field sensitivity, the loudspeaker system efficiency may as well depend heavily on the frequency.

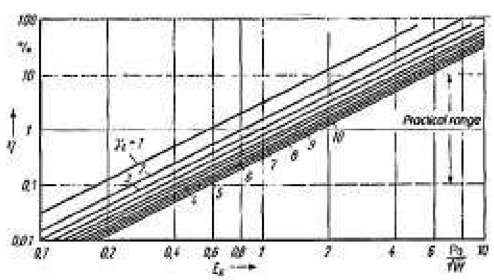


Figure 39-8. Efficiency of a loudspeaker as a function of rated sensitivity and front to random factor.

Directional Properties. All loudspeakers used in real life show a more or less pronounced directional dependence of radiation, which is frequency-dependent—just like beaming behaviors. This angular dependence of sound radiation is characterized by three quantities that are going to be considered in detail.

The *angular directivity ratio* Γ for a frequency or a frequency band is the ratio between the sound pressure p radiated at an angle ϑ from the reference axis, and the sound pressure p_0 generated on the reference axis at equal distance from the selected acoustic reference point (this reference point is selected by the loudspeaker manufacturer and must be published in data sheets; generally it is the center of gravity of the loudspeaker box).[11]

$$\Gamma(\vartheta) = \frac{\tilde{p}(\vartheta)}{\tilde{p}_0}. \qquad (39\text{-}17)$$

In general $\Gamma(\vartheta) \leq 1$. If the maximum of directional characteristics does not occur at $\vartheta = 0°$, then $\Gamma(\vartheta) > 1$.

The logarithmic quantity of the angular directivity ratio is the angular directivity gain

$$D(\vartheta) = 20\log\Gamma(\vartheta)\,\mathrm{dB} \qquad (39\text{-}18)$$

Fig. 39-9 shows the directional characteristic of the horn loudspeaker in a polar plot of the directivity gain. One sees the main maximum at 0° and several secondary maxima at higher frequencies.

An important parameter for direct-sound coverage is the angle of radiation Φ (beam width angle). It stands for the solid-angle margin within which the directivity gain drops by a maximum of 3dB or 6dB (or another value to be specified) as against the reference value. The curves of equal directivity gain are marked Φ_{-3}, Φ_{-6}, or generally Φ_{-n}; the higher the directivity the smaller the angle of radiation, see Fig. 39-10, compare IEC 60268-5, item 23/4.[10]

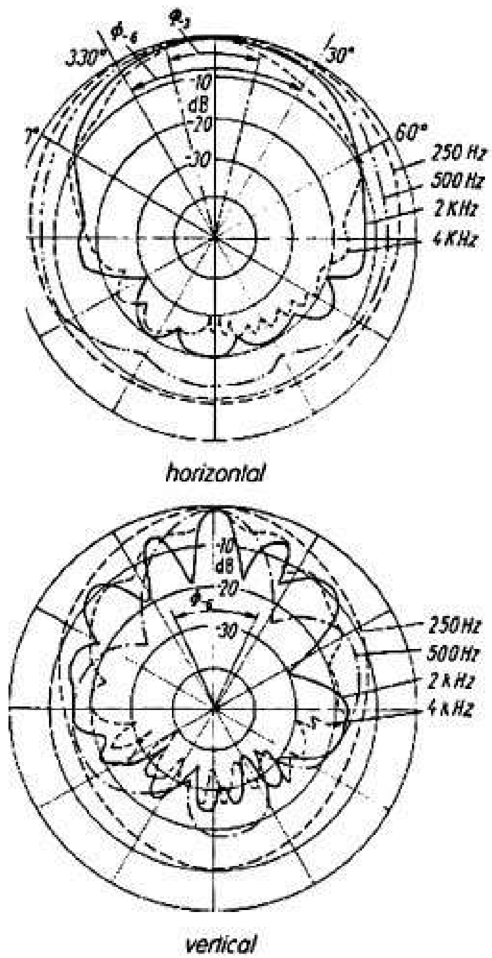
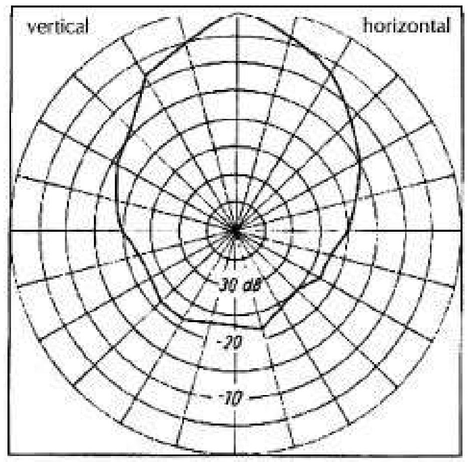


Figure 39-9. Polar plot of the angular directivity gain of a sound column with indication of the radiation angles.

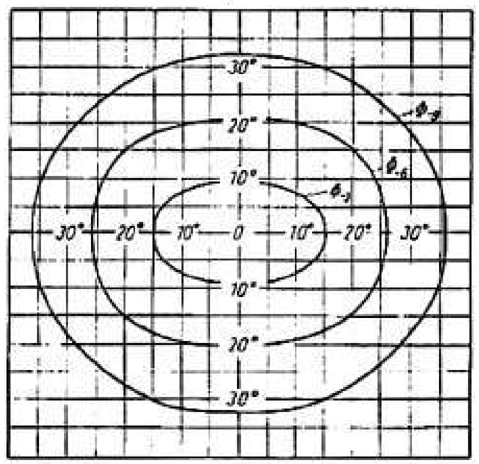
Because of the curves of equal directivity gain and the sound distribution loss, the impact of direct sound of a loudspeaker on a surface may produce elliptic curves that represent a calculated SPL isobar area of the direct sound coverage. These isobar areas are important in the planning of sound reinforcement systems as coverage areas.

For combining the influence of the directional effect as well as that of the distribution between directional and omnidirectional energy, one uses in acoustics the *directivity deviation ratio*:[11]

$$\Gamma^{*}(\vartheta) = \sqrt{\gamma}\Gamma(\vartheta) \qquad (39\text{-}19)$$



A. Directivity characteristic.



B. Coverage area.

Figure 39-10. Plot of the directional effect of the radiator DML-1122 (Electro-Voice) frequency 2 kHz; front to random index 15dB; maximum sound level at 1m distance: 125dB.

This quantity is also of high importance to sound reinforcement engineering. It characterizes the reverberation component caused by the loudspeaker in the excited room. The square of this quantity is the well-known directivity factor Q:

$$\begin{aligned} Q &= g(\vartheta) \\ &= \gamma\Gamma^{2}(\vartheta) \end{aligned} \qquad (39\text{-}20)$$

It should be noted that especially in the USA, it is common to use just Q for different angles ϑ. Nevertheless this entity is angle-dependent and therefore it should always be referenced along with the corresponding angle. The logarithmic expression of the directivity factor Q(ϑ) is the so-called directivity index *DI* (also angle-dependent)

$$\begin{aligned} DI &= \mathrm{H}(\vartheta) \\ &= 10\log g(\vartheta)\,\mathrm{dB} \\ &= 10\log Q(\vartheta)\,\mathrm{dB} \end{aligned} \qquad (39\text{-}21)$$

where,
H(ϑ) is the reverberation directional index.

In the German literature one uses for the directivity factor *Q*, the reverberation directional value g(ϑ).[11] By the same token the directivity index, *DI*, is called *reverberation directional index* H(ϑ).

The reader should be aware of the partially contradicting conventions, of which some are using *Q* and DI only for values of ϑ = 0° and others employ *Q* and DI in an angle-dependent way, sometimes without clearly stating so.

Transmission Range. According to several standards, the transmission range of a loudspeaker is the frequency range usable or preferably used for sound transmission. That region of the transmission curve in which the level measured on the reference axis in the free field does not drop below a reference level generally characterizes the transmission range. The reference value is the average over the bandwidth of one octave in the region of highest sensitivity (or in a wider region as specified by the manufacturer). In the ascertainment of the upper and lower limits of the transmission range there are not considered any peaks and dips whose interval are shorter than 1/8 octave.

This definition implies that loudspeakers have necessarily to be checked as to their transmission range before being used in sound reinforcement systems. With radiators intended for indoor use it is also necessary to consider the front to random factor, i.e., the influence of the diffuse-field component on the formation of the resulting sound pressure.

For special loudspeaker systems, e.g., studio monitoring equipment, narrower tolerance fields of the free-field sound pressure are indicated for the transmission range. Thus the OIRT Recommendation 55/1[12] permits for the range from 100Hz–8kHz a maximum deviation of ±4dB from the average value, whereas below, down to 50Hz, and above, up to 16kHz, the tolerance field widens to −8dB and +4dB.[11]

Fig. 39-11 shows exemplarily the behavior of free-field sensitivity, diffuse-field sensitivity and front to random index of a radiator.

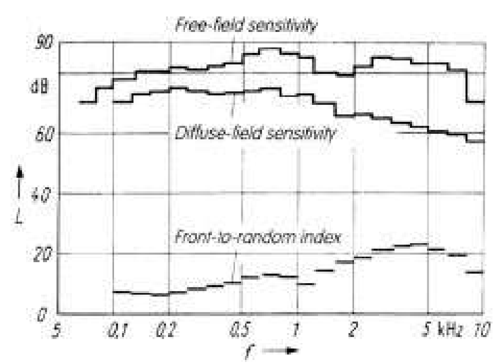


Figure 39-11. Frequency dependence of the front to random index as compared with the free-field and the diffuse-field sensitivities.

Moreover the transmission range is influenced, especially in the lower frequency range, by the installation conditions or the arrangement of the radiator. Fig. 39-12 shows that the arrangement of the loudspeaker system has a considerable influence on the transmission curve. This is due to the fact that arranging the radiator in front of, below or above a reflecting surface causes interferences of the direct sound by the strong reflections that give rise to comb-filter-like cancellations which can be proven by a narrow-band analysis of the resulting signal. These cancellations are particularly pronounced, if in case of an arrangement in front of a wall the radiator is provided in its rear part with compensating openings or if these reflections come from a distance of about 1.5m out of a room corner, e.g., between ceiling and wall.

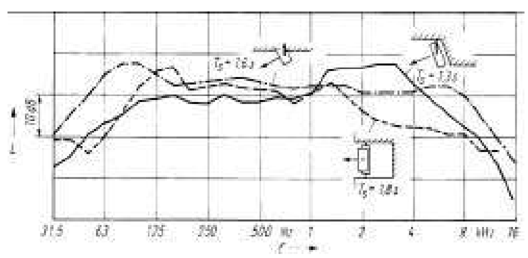


Figure 39-12. Frequency response curve of a loudspeaker at different mounting conditions. T_s is the reverberation time of the hall.

As a rule one can say that the ear does normally not perceive dips and peaks that are not measurable in a third-octave band filter analysis (unless they show pronounced periodic structures).

A good bass radiation is produced if the radiating plane is embedded in a reflecting surface, for instance a wall or a ceiling. In this case there may also exist a certain angle between the radiating plane and the surrounding surface.

Types of Loudspeakers. The different tasks of sound reinforcement engineering require different radiator types. These differ as to size and shape of their enclosures, the form of sound conduction, the types of driving systems used as well as arrangement and combination of the same. In this way one obtains different directional characteristics of sound radiation, sound concentrations, sensitivities, transmission ranges, and dimensions which facilitate solutions for diverse applications or even enable them at all.

Among the simplest radiators are single loudspeakers of smaller dimensions and ratings that are used in decentralized information systems, for instance for covering large flat rooms or for producing room effects in multipurpose halls. The integration into a wall or an enclosure of these loudspeakers avoids the “acoustic short circuit” usually seen with “no baffle” situations (suppressing the pressure compensation between the front and rear sides of the diaphragm). To this effect a baffle panel or an open or closed box may be used, Fig. 39-13.

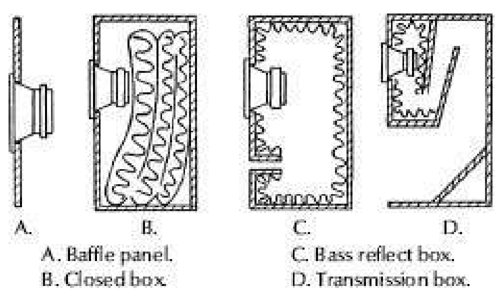


Figure 39-13. Different measures for suppressing the “acoustic short circuit.”

With a closed box one has to consider that the oscillating part of the loudspeaker functions in one direction against the relatively stiff air cushion of the box. Loudspeakers for such *compact boxes* are for this reason provided with an especially soft diaphragm suspension so that they cannot be easily used for other purposes.

Acoustically more favorable are the conditions with vented enclosures, the *bass reflex boxes* or *phase reversal boxes*. Such box loudspeakers are nowadays used less as decentralized broadband radiators, but increasingly for high-power large-size loudspeaker arrays.

Another possibility for achieving a determined directional characteristic consists in the arrangement of sound-conducting surfaces in front of the driving loudspeaker system. Given that such arrangements are mostly of horn-like design, they are named *horn loudspeakers*. Because of the high characteristic sensitivity and the high directional characteristic, this radiator design is very well suited for sound reinforcement in big auditoriums where the desirable frequency range and different target areas (coverage areas) require the use of different types of radiation pattern.

For technical reasons it is not sensible to construct a broadband horn for the overall transmission range. A better solution is several horn loudspeakers complementing each other.

Bass Horns. Owing to the great dimensions involved, the design of bass horns requires extensive compromises. Practical models of bass horns receive a horn shape, as a rule, only in one dimension, whereas at a right angle to it, sound control is achieved by means of parallel surfaces. The power-handling capacity of such bass horns, which are mainly used in music or concert systems, is about 100 to 500VA.

Medium-Frequency Horns. The greatest variety of driver and horn designs is available for horn loudspeakers for the medium frequency range of about 300Hz to 3kHz.

The drivers used are mostly dynamic pressure-chamber systems connected to the horn proper by means of a throat, the so-called *throat-adapter*.

Treble Horns. For the upper frequency range, two main types of horn loudspeakers are produced. These are the horn radiators showing similar design characteristics as the medium-frequency horns that function in the frequency range from 1 to 10kHz, and the special treble loudspeakers (*calotte horns*) used for the frequency range from 3 to 16Hz.

39.2.3.2.2 Loudspeaker Arrays with In-line Arrangement of Radiators (Loudspeaker Line, Sound Column, Line Arrays)

Classical Columns. For many tasks of sound reinforcement engineering one requires radiators capable of producing a high sound level at a large distance from their point of installation, while minimally affecting microphones located at close range to them. To have this effect they must show a determined directional characteristic and directivity. A radiator type suitable for this purpose is the loudspeaker array consisting, in the variant required, in a stacked arrangement of in phase-identical loudspeakers. In the plane perpendicular to this arrangement there occurs a pressure addition, whereas in the areas above and below this plane there is a cancellation by interference because of the early-to-late difference between the components stemming from the different loudspeakers, Fig. 39-14. Each of the individual loudspeakers radiates the sound spherically and the sound waves get favorably superposed in the far field, whereas the effect of the individual loudspeaker prevails in the near field. For the far field the following equation was given by Stenzel[13,14] and Olson[15] for the angular directivity ratio Γ, the so-called polars.

$$\Gamma = \frac{\sin\left[\frac{n\pi d}{\lambda}\sin\gamma\right]}{n\sin\left[\frac{\pi d}{\lambda}\sin\gamma\right]} \qquad (39\text{-}22)$$

where,

n is the number of individual loudspeakers,

d is the spacing of the individual loudspeakers,

γ is the radiation angle,

λ is the wavelength of sound,

l is $(n - 1)\,d$, the length of the loudspeaker line.

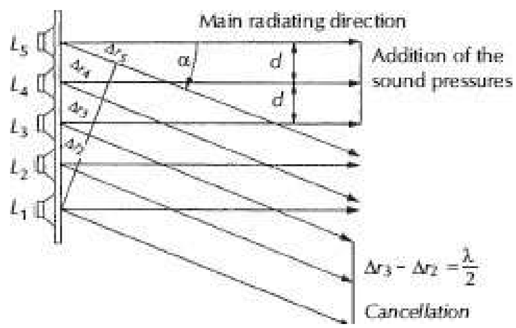


Figure 39-14. Operating principle of a classical sound column.

This directional effect of a loudspeaker line according to Fig. 39-15[16] is shown in Fig. 39-16A a balloon at 1kHz, and in Fig. 39-16B a balloon at 2kHz. The line consists of nondirectional loudspeakers arranged with a spacing of 25cm. Secondary maxima occur at frequencies above a critical frequency (wavelength = spacing of the loudspeakers), that is above 1400Hz in the example. Thus a desirable discshaped radiation without secondary maxima can be observed at 1000 Hz, whereas at 2000Hz lobes (secondary maxima) are already utterly evident.

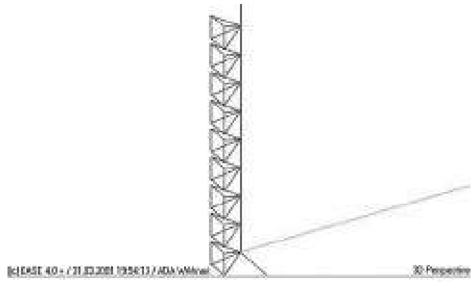
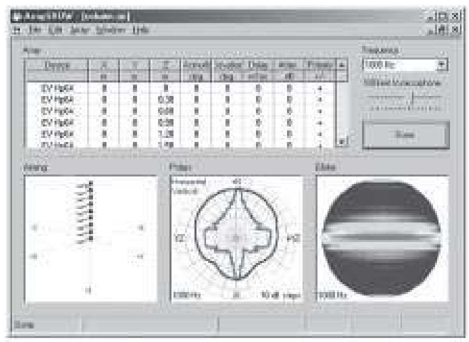
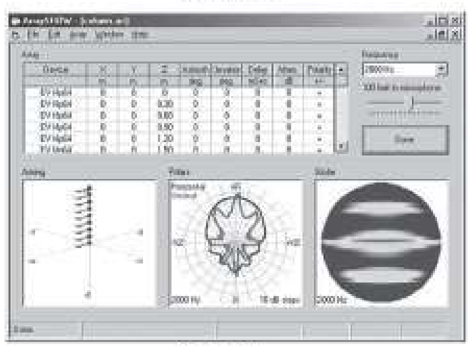


Figure 39-15. A line presentation with nine horns HP64 in the simulation program EASE.



A. 1000 Hz.



B. 2000 Hz.

Figure 39-16. Balloon presentation of the line array according Fig. 39-15 in a simulation.

The drawback of an in-line loudspeaker arrangement consists of the fact that

- The desired directional effect is given only in a range below the critical frequency, whereas above that frequency there occur additional secondary maxima.

- The directivity is frequency-dependent (*front-to-random factor* γ of the main maximum ≈ 5.8*lf* (l is the length of the column in m and *f* the frequency in kHz).[17]
- The directivity increase does not only occur in the "directivity domain," but also, owing to the distances of the individual loudspeakers, in the scattering domain, so the column is losing directivity at high frequencies.

All these frequency-dependent properties of the loudspeaker lines involve the possibility for timbre changes to occur over the width and depth of the covered auditory. In order to eliminate or limit this drawback, the lines are often subdivided in the upper frequency range. This is mostly accomplished by curving the line "banana-like" or like a so-called J-Array. Alternatively individual elements can be rotated slightly off-axis in the horizontal domain, such as in alternating angles of +10° and –10° relative to the aiming axis of the system.

Line Arrays. Modern line arrays do not consist of a line-up of individual cone loudspeakers, but instead of a linear arrangement of wave-guides of the length *l*, which produce a so-called coherent wave front. In contrast to the traditional sound columns, these arrays radiate in their near range so-called cylindrical waves. This near range is frequency-dependent and only valid up to the following distances *r*:

$$r = \frac{l^2}{2\lambda} \qquad (39\text{-}23)$$

where,
both array length and wavelength are in m.

In 1992 Christian Heil was the first to present this new design at the AES in Vienna.[18] With the product V-DOSC by L-Acoustics a new technology was introduced which can now be found with modifications in the product range of more than 40 manufacturers, Fig. 39-17.

The characteristic feature of these systems is that the sound levels decrease in the near field by only 3dB with distance doubling, and begin to decrease like those of spherical radiators only beyond the near range. This way it is possible to cover large distances with high sound levels and without having to use delay towers.

Digitally Controlled Line Arrays. A way of reducing the frequency dependence of the directional characteristics and directivity of sound lines consists of supplying the sound signal with different phases and levels to the individual loudspeakers in an array.

Figure 39-17. A Geo T series column by NEXO SA.

Duran Audio was one of the first manufacturers who reduced the length of their INTELLIVOX loudspeaker lines with increasing frequency by electronic means (so-called DDC solution). This solution resulted in loudspeaker lines with pronounced directivity in the vertical domain and constant sound power concentration in the horizontal domain.[19] Fig. 39-18 illustrates such a directional effect in three-dimensional representation.

Nowadays this type of digitally aimed loudspeaker is offered by many manufacturers, e.g., Renkus-Heinz with the ICONYX loudspeaker,[20] see Fig. 39.19, or ATEIS (Messenger),[21] EAW (DSA series),[22] and Meyer Sound Inc. (CAL series).[23]

By changing the firmware-based control the following features of such columns are possible:

1. Constant SPL versus distance.

- Mid-band frequencies.
- Non-complex shaped audience areas.

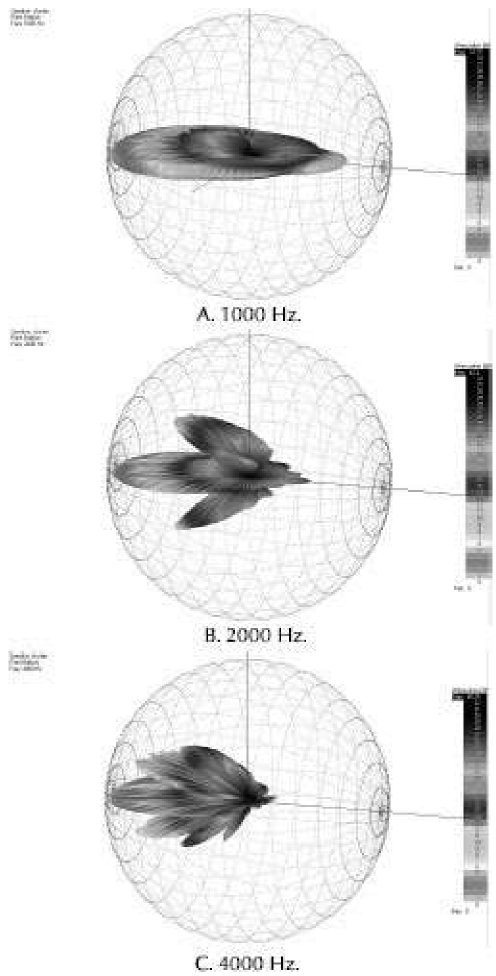


Figure 39-18. Cluster balloon presentation of Intellivox 2C in EASE.

2. The performance is optimized with the following parameters:

- Opening angle.
- Aiming angle.
- Focus distance.
- Mounting height with respect to audience area.

This solution, however, finds its limitations with a complicated audience-area layout. Moreover, certain sound level distributions can be obtained only after several corrections. This gives rise to questions like: “How can a loudspeaker array be controlled so as to create a predefined far and near field response?” One approach was made by DURAN Audio with the Digital Directivity Synthesis (DDS).[24]

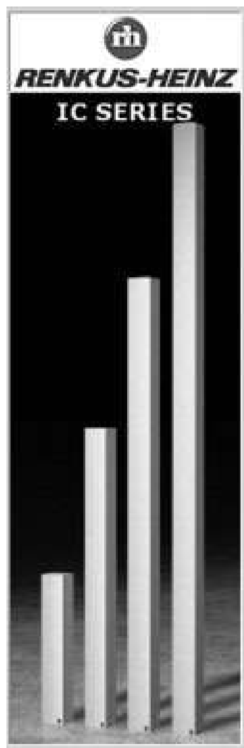


Figure 39-19. ICONYX Series from Renkus Heinz.

Here the directivity patterns of an array are adaptable to audience areas; and uniform SPL distribution also becomes possible in complex-shaped audience areas.

It stands to reason that point sources or line loudspeakers of a certain dimension can no longer exhibit this radiation behavior. Here the manufacturer has to supply along with the arrays not only the electronic driver unit for the same, but also the parameter setup algorithm. By means of attached software this algorithm will then be controlled according to the desired application. Over the last few years, a number of companies have introduced numerical optimization functions that provide advanced FIR filters in order to adapt the sound system's radiation pattern to the geometry of the venue, e.g., the software DISPLAY by Martin Audio,[25] BeamEngine by Tannoy,[26] and FIRmaker by AFMG.[27]

39.2.4 Wall Materials

To simulate the radiation behavior of sources in rooms or open spaces one needs to construct corresponding models. All boundary walls of these models need to have the corresponding acoustic properties like:

- Absorption.
- Scattering.
- Diffraction.

These properties have been discussed in Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, section 9.3.4, therefore this should not be repeated here. Instead, some specialties important to know when doing computer modeling should be added here.

39.2.4.1 Absorber Data

Absorbing behavior is known for hundreds of years, but data has been available only for 80 years. We distinguish between data measured in a reverberation chamber (Standard ISO 354)[28] or data that is angle-dependent. The latter absorption coefficient is very seldom measured and only available for special applications. For computer simulation the absorption coefficient measured in the diffuse field will be used. This coefficient is measured by the corresponding manufacturers and published in specification brochures. It is measured in octave or one-third octave bands and starts normally at 63Hz and goes up to 12kHz or even 16kHz. In most simulation programs the low end is skipped because the actual simulation routines do not cover frequency ranges below 100Hz and the highest-frequency band is quite often only 8kHz.

All this data is meanwhile published in table form and some simulation programs have more than 2000 materials from different manufacturers on board.

39.2.4.2 Scattering Data

Scattering data is not found in textbooks except for some special scattering materials or samples. Here should be mentioned the products of RPG Diffuser Systems Inc., which produce special modules with sound-scattering surfaces.[29]

On the other hand it is known that the absolute value of the scattering coefficient s is less important. The fact is that there is almost no material not scattering (s = 0) or only scattering (s = 1). The practical values for the scattering coefficients are between 0 and 1. So there are some rules of thumb to define the actual scattering coefficient in simulation software programs. Some programs give some guidance to estimate the coefficients, other programs like EASE use special BEM routines (compare to Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, Fig. 9-46) to derive the coefficient in a way as it should be measured according the proposals of Mommertz,[30] see also Standard ISO 17497-1.[31]

A scattering coefficient will never generally be available in tables (except the mentioned special module values), because the way the interior architect uses the materials in a hall affects the scattering behavior as well.

Therefore the scattering behavior of wall parts in a computer model must be determined model-specific.

39.2.4.3 Diffraction, Low-Frequency Absorption

As we will see in section 39.3.2, the computer simulation programs use different ray-tracing algorithms to calculate the impulse responses in model rooms. But these routines of using particle radiation are only valid above a certain frequency determined by

$$f_1 = K\sqrt{\frac{RT}{V}} \qquad (39\text{-}24)$$

where,
K is a constant (2000 in metric units, 11,885 in U.S. units,
RT is the apparent reverberation time in s,
V is the volume of the room in m^3 or ft^3.

For lower frequencies and especially in small rooms the particle assumption cannot be applied anymore. Here the Wave Acoustics routines are applied. An analytical solution is impossible, so numeric routines have been developed. Mainly the Finite Element Method (FEM) and the Boundary Element Method (BEM) are used. First for applying the FEM the 3D computer model must be subdivided in small volumes (Meshes), where the dimensions of the mesh correspond with the upper frequency handled by the FEM. The higher the frequency, the smaller the dimensions and the longer is any calculation time. As an example, to build a mesh in a hall of 10,000m^3 you need a mesh resolution of about 6.9×10^6 sub volumes to apply the FEM up to 500Hz. Fig. 39-20 shows a mesh grid for a church model. For the BEM only the surface must be meshed accordingly.

After the mesh is ready and that is a quite difficult job in complex room structures,[32] we need to know the impedance behavior of the single wall parts. This is again quite complex because any stiffness or mass values of the majority of the wall materials are not known. So in a first approach the impedance of the wall material can be derived from the known absorption coefficient. Now by applying the well known algorithm of the FEM the transfer function at selected receiver places may be calculated. By means of a Fourier transformation one obtains the impulse response in the time domain. By means of this method also transfer functions at receiver places may be calculated, even if the receiver is shadowed from the sending position and the direct sound was only coming by diffraction to the receiver.

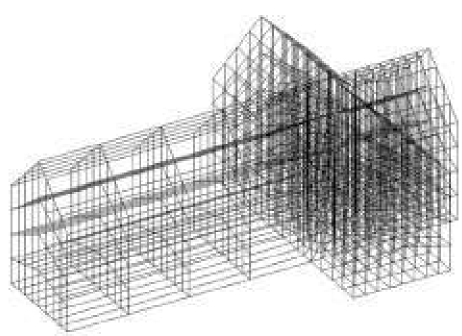


Figure 39-20. Meshed model in EASE 4.2.

This method can be used very well in small rooms below 300Hz. Under the above mentioned conditions a mesh of a control room of 135m^3 consists only of around 20,000 sub volumes, if frequencies higher than 300Hz will be neglected. This way very fast calculation results can be expected.[33]

39.2.5 Receivers and Microphone Systems

39.2.5.1 Human Ears

The properties of the human ears are explained in a lot of books about psychoacoustics, including Chapter 3, Psychoacoustics in this handbook. In simulation programs the acoustic properties of a room or the free field environment are determined by calculation of the so-called impulse response. This response is calculated using raytracing methods. For a single point in space the so-called monaural response is determined and the result supplies not only the level at this receiver place but also the frequency dependence, the angle of incidence for single reflections and the run-time delay in comparison to the first incoming signal (direct sound). Using so-called head-related-transfer functions (HRTF) measured with dummy heads or using in-ear microphones, see Fig. 39-21,[34] the monaural impulse response may be converted into a binaural one used for real-time convolution, see section 39.3.3.

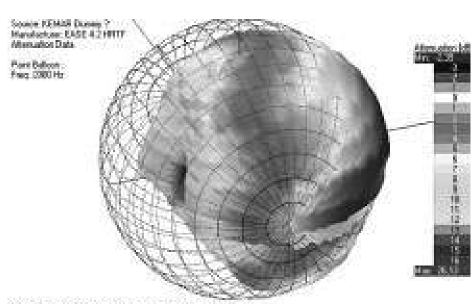


Figure 39-21. HRTF balloon of the left ear of a dummy head.

39.2.5.2 Microphones

The use of microphones in sound reinforcement systems requires observation of a number of conditions. To avoid positive acoustic feedback it is frequently necessary to keep the microphone at closer distance to the sound source so that often considerably more microphones have to be used. Moreover the live conditions demand very robust microphones.

To simulate the use of microphones, to pre-calculate the acoustic feedback threshold or to simulate enhancement systems based on electronic processing presuppose an exact knowledge of the properties of the microphone types and their connection technique is needed.

39.2.5.2.1 Basic Parameters

The microphone data are laid down in standards.[35] In this context we will consider only those data which are important for computer modeling. For further information especially regarding the types please refer to Chapter 20 Microphones.

The magnitude of the output voltage of a microphone as a function of the incident sound pressure is described by the microphone sensitivity

$$T_E = \frac{\tilde{u}}{\tilde{p}} \qquad (39\text{-}25)$$

in V/Pa or its 20-fold common logarithm, the *sensitivity level*

$$G_s = 20\log\frac{T_E}{T_0}\,\mathrm{dB} \qquad (39\text{-}26)$$

The reference sensitivity T_0 is normally specified for 1V/Pa.

Depending on the test conditions one distinguishes the following sensitivities:

- The *pressure open-circuit sensitivity* T_{Ep} as the ratio between the effective output voltage at a certain frequency and the effective sound pressure of a vertically incident sound wave.
- The *(free)-field open-circuit sensitivity* T_{Ef} which considers by special measuring conditions the pressure increase conditioned by the cross-section dimensions of the microphone.
- The *diffuse-field sensitivity* T_{Er}, which reflects the diffuse sound incident on the microphone.

39.2.5.2.2 Directivity Behavior

The dependence of the microphone voltage on the direction of incidence of the exciting sound is called *directional effect*. The following quantities are used for describing this effect:

- The *angular directivity ratio* $\Gamma(\vartheta)$ is the ratio between the (free)-field sensitivity T_{Ed} for a plane sound wave arriving under the angle ϑ to the main microphone axis and the value ascertained in the reference level (incidence angle 0°).

$$\Gamma(\vartheta) = \frac{T_{Ed}(\vartheta)}{T_{Ed}(0)} \qquad (35\text{-}27)$$

- The *angular directivity gain D* is the 20-fold common logarithm of the angular directivity ratio.
- The *coverage angle* is the angular range within which the directivity gain does not drop by more than 3dB (or 6dB, or 9dB) against the reference axis.

Apart from the quantities describing the ratio between the sensitivities of the microphone with sound incidence from various directions deviating from the main axis, it is also necessary to describe the relationship between the sensitivities with reception of a plane wave and those with diffuse excitation. With these quantities it is then possible to ascertain the suppression of the room-sound components against the direct sound of a source to be transmitted. This energy ratio is described by the following parameters:

- The front to random factor is the ratio between the electric power rendered by the microphone when excited by a plane wave from the direction of the main axis, and the power rendered by the microphone excited in a diffuse field with the same sound level and same exciting signal. If the sensitivity was measured in the direct field as T_{Ed} and in the diffuse field as $T_{E\ r}$, the front to random factor results as

$$\gamma_M = \frac{{T_{Ed}}^2}{{T_{Er}}^2}. \qquad (35\text{-}28)$$

- The front to random index is the 10-fold common logarithm of the front to random factor.

While the front to random factor of an ideal omnidirectional microphone is 1, that of an ideal cardioid microphone is 3. This means that a cardioid microphone picks up only 1/3 of the sound power of a room picked up by a comparable omnidirectional microphone at the same distance from the source. This implies for instance that with identical proportion of the sound power, the speaking distance for a cardioid microphone may be three times greater than that of an omnidirectional microphone.

39.3 Transducer Data for Acoustic Simulation

To simulate an entire acoustic system, all parts must be taken into account. Besides the room, also loudspeakers and natural sound sources as well as microphones and the human hearing system have to be considered. In this section, our main goal is to review existing practice and outline advantages and disadvantages that the user of a software program should be aware of when applying performance data for a particular sound transducer. In this regard, our intention is to talk about the simulation of transducers with respect to the electro-acoustic and room-acoustic prediction of the acoustic system as a whole. We will not be concerned with mathematical methods applied in the design process of loudspeakers or microphones. These usually provide a much higher degree of accuracy in some regards but at the same time often provide insufficient data for other simulation purposes. Specifically, for transducer design utilizing BEM/FEM-based prediction methods we refer the reader to available textbooks and publications.

39.3.1 Simulation of Loudspeakers

In computer aided-acoustic design and especially for sound reinforcement applications, the level of accuracy to which sound sources are modeled plays a crucial role. Accordingly, most simulations software packages have continuously developed their capabilities of describing loudspeakers by measurement data along with the complexity of the loudspeaker systems themselves. At the same pace, the availability and fast development of personal computers had a significant impact on acoustic measurement systems and their accuracy on the one hand and on the computing power available on the other hand.

In this sense, the measurement and simulation of loudspeaker systems can be roughly divided into two periods of time. The first period, until the late 1990s, was characterized by the use of simplified far-field data for almost any sort of loudspeaker and the assumption of a point source-like behavior. But with the advent of modern line array technology, for both tour sound and speech transmission applications, new concepts had to be developed. These methods include the use of multiple point sources as well as advanced mathematical models to image the complexity of today's loudspeaker systems. In addition to that, research was further accelerated by the broad availability of DSP platforms and the resulting need to simulate DSP controlled loudspeakers as well as the virtual disappearance of computer-based constraints, such as calculation speed and memory.

39.3.1.1 Simulation of Point Sources

39.3.1.1.1 Theoretical Background

For many years the radiation behavior of sound sources, and loudspeakers in particular, was basically described by a three-dimensional matrix containing magnitude data in a fixed spectral and spatial resolution. Starting with the late 1980s, typical data files contained directivity data for the audible octave bands, such as from 63Hz to 8kHz, and for a spherical grid with an angular spacing of 15°. Mostly, data was also assumed to be symmetric in one or two planes. With the need for higher data resolution and the limits of available PC memory and computing power changing at the same time, more advanced data formats developed eventually reaching a nowadays typical resolution of 5° angular increments for 1/3 octave frequency bands. Tables 39-1 and 39-2 show some of these typical loudspeaker data formats and their resolutions.

Now let us look at the background for this development. We express the complex sound pressure $\bar{p}$ for the time-independent propagation of a spherical wave:[36]

$$\bar{p}_{Sphere}(\vec{r}, f) = \frac{\bar{A}(\varphi, \vartheta, f)}{|\vec{r}|} \exp(-j\vec{k}|\vec{r}|) \qquad (39\text{-}29)$$

where

$\vec{r}$ is the receiver location,

f is the frequency,

$\bar{A}(\varphi, \vartheta, f)$ is the complex radiation function of the source depending on angles φ and ϑ (both being functions of $\vec{r}/|\vec{r}|$) as well as on the frequency,

$\vec{k}$ is the wave vector.

Loudspeaker measurements $\hat{A}$ happen at discrete angles φ_k, ϑ_l, and frequencies f_m, the simulation software has to interpolate between such data points to obtain a smooth response function:

$$\bar{p}_{Sim}(\vec{r}, f_m) = \frac{f_{Int}(\hat{A}(\varphi_k, \vartheta_l, f_m))}{|\vec{r}|} \exp(-j\vec{k}|\vec{r}|) \quad (39\text{-}30)$$

Here the interpolation function is represented by f_{Int}. The frequency resolution is basically given by the set of available data points of f_{m}, the angular resolution is given by the density of data points φk ϑ_1.

For a long time, most measurements were made to acquire magnitude data only, $\hat{A} = |\hat{A}|$ and $f_{Int}(\hat{A}) = |f_{Int}(\hat{A})|$. In such a case, the simulation of interaction between multiple sound sources yields a sound intensity I_{Sum}, that is derived either by power summation for incoherent sources n (located at $\vec{r}_n$):

$$I_{Sum}(\vec{r}, f_m) = \sum_n |\bar{p}_n|^2 = \sum_n \frac{[f_{Int}(\hat{A}_n(\varphi_k, \vartheta_l, f_m))]^2}{|\vec{r} - \vec{r}_n|^2} \quad (39\text{-}31)$$

or by at least considering the run time phase $\Phi_n = -\vec{k}_n(\vec{r} - \vec{r}_n)$ for coherent sources:

$$I_{Sum}(\vec{r}, f_m) = \left|\sum_n \bar{p}_n\right|^2 = \left|\sum_n \frac{f_{Int}(\hat{A}_n(\varphi_k, \vartheta_l, f_m))}{|\vec{r} - \vec{r}_n|} \exp[-j\vec{k}_n(\vec{r} - \vec{r}_n)]\right|^2 \quad (39\text{-}32)$$

This simulation model makes some significant assumptions:

1. First, the use of a spherical wave form assumes that both measurement and simulation happen in the far field of the device, that is, at a distance where the sound source (normally a surface) can be considered as a point source $\bar{p}_{Real}(\vec{r}, f) \approx \bar{p}_{Sphere}(\vec{r}, f)$.

Table 39-1. Conventional Loudspeaker Data Formats and EASE GLL Format

Data Type	EASE SPK Simple Data Table	EASE XHN Simple Data Table	GDF Simple Data Table	ULYSSES UNF Simple Data Table	CLF Simple Data Table	EASE GLL Advanced Description Language
Balloon Symmetries	Full, Half	Full	Full, Half, Quarter	Full, Half, Quarter	Full, Half, Quarter	Full, Half, Quarter
Angular Resolution	5°	5°	5° or 10°	5° or 10°	5° or 10°	1° to 90°
Frequency Resolution	1/3 Octave	1/3 Octave	1/1 or 1/3 Octave	1/1 or 1/3 Octave	1/1 or 1/3 Octave	Any
Complex Data	Yes	Yes	No	No	No	Yes
Individual Transducers	No	No	No	No	No	Yes
Filters	No	No	No	No	No	Yes
Configurable	No	No	No	No	No	Yes

Table 39-2. Measurement Parameters for Typical Balloon Resolutions

Measuring Resolution	Measuring Points	Measuring Time (for 10 seconds per measuring point)	Implemented in
2 measuring planes, 15°, symmetrical in both measuring planes	24	≈ 10 min	EASE 1.0
Measurement on sphere surface, 10°, symmetrical in the horizontal plane	325	≈ 1 h	EASE 2.1 ULYSSES 1.0/CLF
Measurement on sphere surface, 10°, no symmetry assumptions	614	≈ 1½ h	CATT-Acoustic/CLF ODEON BOSE Modeler
Measurement on sphere surface, 5°, symmetrical in the horizontal plane	1297	≈ 3½ h	CLF EASE 3.0-4.2 EASE 4.2 DLL/GLL
Measurement on sphere surface, 5°, no symmetry assumptions	2522	≈ 7 h	ULYSSES 2.82/CLF EASE 3.0-4.2 EASE 4.2 DLL/GLL
Measurement on sphere surface, 2°, no symmetry assumptions	16022	≈ 2 d	MAPP (Meyer) EASE 4.2 DLL/GLL
2 Transducer measurements on sphere surface, 10°, no symmetry assumptions	1228	3 h	EASE 4.2 GLL EASE 4.2 DLL

2. Second, it is assumed that the density of discrete data points is high enough and the frequency and angular dependency of the directivity characteristics smooth enough so that the true radiation function of the spherical wave can be approximated by $\overline{A} \approx f_{Int}\hat{A}$
3. Third, the use of magnitude-only data, assuming that $\overline{A} \approx f_{Int}\hat{A}$ requires that the point of reference during the measurement of $\hat{A}$ is chosen in a way that the true run time phase, otherwise

included with the measurements can be reconstructed by the run time phase *kr* in the model and it requires that the source-inherent phase is negligible as well, $\arg \bar{A} \approx 0$.

4. It is assumed that the concerned loudspeaker system is a fixed system that cannot be changed by the user or when its configuration is changed its performance data is not affected. The measurement data is regarded as representative for the whole range of possible applications and configurations.
5. Finally, for the use of such point sources in computations involving geometrical shadowing and raytracing calculations, the source is regarded as located at a single point and is thus either wholly visible (audible) for a receiver or not.

These assumptions have been made especially in the early 1990s in order to obtain and use loudspeaker directivity data in a practical manner. Important factors were the availability and accuracy of measurement platforms and methods, the storage size of the processed measurement data and the PC performance with regard to processor speed available to the average user of the data.

However, these assumptions have a set of drawbacks. That became most evident with the broad use of large-format touring line arrays and digitally controlled loudspeaker columns but also with the increasing use of inexpensive DSP technology employed for multi-way loudspeaker systems. Some of the issues conflicting with above points 1–5 are listed in the following.

- A large line array system of some meters height cannot be measured adequately in its far field in addition to the fact that a large number of line array applications actually happens to take place mainly in the near field. Therefore the simulation of a whole line array as a point source is not valid within reasonable error ranges.

- Another problem often encountered is insufficient angular resolution. Loudspeaker columns but also multi-way loudspeakers exhibit significant lobing behavior in the frequency ranges where multiple acoustic sources interact at similar strength. Often too coarse angular measurements fail to capture these fine structures and thus cause erroneous simulation results due to aliasing/sampling errors.
- While in many cases the phase of the sound pressure radiated by a simple loudspeaker is negligible at least if it is considered on-axis and the run time phase is compensated for, the same is not true for most real-world systems. On the one hand, for multi-way systems one cannot generally define a single point where the measured phase response vanishes for all frequencies and angular directions. This is the problem of the so-called acoustic center for a set of sound sources. In such cases the measured phase data will typically show a run time phase component that depends on angle and frequency, no matter where the point of rotation is. On the other hand, the inherent phase response plays an important role in describing the radiation behavior that is influenced by diffraction about the edges of the loudspeaker case, that is, at angles of 60° and more off-axis.
- Loudspeaker systems become increasingly configurable, so that the user can adapt them to a particular application. Typical examples include almost all touring line arrays where the directional behavior is defined mechanically by the splay angles between adjacent cabinets, and loudspeaker columns or multiway loudspeakers, where the radiation characteristics can be changed electronically by manipulating the filter settings.

- In advanced computer simulations of sound reinforcement systems in venues, geometrical calculations must be performed. This is required to obtain exact knowledge of which part of the audience might be shadowed by obstacles between the sound sources and the receivers. Geometric considerations are also needed in raytracing calculation in order to find reflections and echoes. For both processes, the reduction of a physically large loudspeaker system to a point source can lead to significant errors. Depending on the choice of the reference point for the source particular reflections might not be found or exaggerated, or a large fraction of the audience area might be seemingly shadowed by a very small object.

In addition to the above, a set of minor problems is evident too. This includes the definition of maximum power handling capabilities of multi-input systems that are represented by a single point source, the availability of case drawings to help in the mechanical design and the clear indication of the reference point that was used for the measurements.

As a result of the obvious contradictions a variety of proposed solutions emerged in the later 1990s. This development happened partially at the loudspeaker manufacturers and partially by the creators of simulation software as well. To resolve the problem of large-format loudspeaker systems a sub-division into smaller elements is required to be able to measure them and use them for prediction purposes. To properly model the coherent interaction between these elements complex measurement data, including both magnitude and phase data, is needed.

The most prominent solutions can be summarized as follows. Instead of measuring a whole system, so called far field cluster balloons were calculated based on the far field measurement of individual cabinets or groups of loudspeakers.[35] To describe individual sound sources, phase data was introduced in addition to the magnitude-only balloon data,[38,39] or mathematical models providing phase information implicitly were applied, such as minimum phase or elementary wave approaches as well as two dimensional sound sources.[18] However, these first approaches lacked generality and thus their implementation into existing simulation software packages was specific, difficult, or even impossible.

The situation was resolved first by the concept of the loudspeaker DLL (dynamic link library), which serves essentially as a programmable plug-in for simulation software.[40] Another concept, namely the GLL (generic loudspeaker library), introduced a new loudspeaker data file format that is significantly more flexible than the conventional data formats and is designed to resolve most of their apparent contradictions.[41] We will review both approaches in the next section as they have turned out to be a standardized way to model complex loudspeaker systems.

39.3.1.1.2 Practical Considerations

Improvements which are theoretically desirable must also be practically accomplished. It is clear that only reasonable measurement times can provide reliable data in an efficient manner. In practice, an angular resolution of 5° has proven to be adequate for most needs, sometimes even lower resolutions of 10° can be sufficient. Simulation software packages should be able to handle higher resolutions as well but only for special cases. This is particularly feasible when measuring durations can be reduced using multiple microphones, like 10 or 19 receivers arranged on an arc.[42] This technique requires some care in the measurement setup since all of the microphones have to be calibrated and normalized relative to each other.

Also the acquisition of both magnitude and phase data requires more care than just the measurement of magnitude-only data. However, modern impulse response acquisition platforms provide a good means to obtain complex data and a sufficient frequency resolution. The representation of a loudspeaker directivity function based on impulse response wave files is thoroughly discussed by the Working Group of the Standards Committee SC04-01 of the AES.[8] As we will present further below, the utilization of phase data in acoustic modeling has become an important factor. As an illustration, Fig. 39-22 A–D show the magnitude and phase data for a loudspeaker (UPL1 from Meyer Sound Inc.) in high resolution in both MATLAB and EASE.[43]

Additionally, it is worth mentioning, and we will give some practical guidelines in the next sections, that in order to obtain acceptable data for a point source approach measurements have to take place in the far field of that assumed point source. Like indicated previously, this may be difficult for large multi-way cabinets or column loudspeakers.

In general it must be emphasized that the computer model utilizing this loudspeaker data can only be as good as the data of the lowest quality included. Nowadays, the accuracy of the loudspeaker data is often much higher than that of the material data. Absorption and scattering coefficients are usually only known in 1/1 or 1/3 octave bands for random incidence. The user must be aware that although loudspeaker direct field predictions may be very precise, any modeling of the reflections and the diffuse sound field in the room will be limited by the available material data. Furthermore, it is not very likely that there will ever be systematic, large-scale measurements of angle-dependent complex directivity data for the reflection and scattering of sound by wall materials.

To complete this practical perspective another point of concern has to be underlined. Any data set describing the acoustic characteristics of a loudspeaker should also document important measurement parameters and conditions. In particular, the point of rotation used for the sensitivity and balloon measurements must be defined in such a data set and indicated in the case drawing as well, Fig. 39-23. Only when this reference point is known, the end user will be able to define precisely the location of the loudspeaker in the computer model and to obtain the right results.

39.3.1.2 Simulation of Complex Loudspeakers

39.3.1.2.1 Modeling by Means of a DLL Program Module

In a first step to overcome the variety of issues related to the reduction of complex loudspeakers to simple point sources the DLL approach was developed.[44,45] Technically speaking, the MS Windows dynamic link library (DLL) is a program or a set of functions that can be executed and return results. It cannot be run standalone but only as a plug-in of another software which accesses it through a pre-defined interface. The basic idea is to move the complexity of describing a sound source from the acoustic simulation program into a separate module that can be developed independently and that can contain proprietary contents. In this way, a clear cut is made between the creators for simulation software packages and the loudspeaker manufacturers who can develop product-specific DLL modules on their own.

However, acoustic prediction programs have different underlying concepts and therefore, although the DLL concept is a general philosophy, the DLL interfaces are different too. In consequence, a DLL built for one simulation platform cannot be used for another. Nevertheless, all DLL models share a similar approach to resolve the given problems. Because they can be programmed, they are essentially able to handle any kind of data and realize any kind of algorithm. If the mathematical description and/or sufficient measurement data for the loudspeaker system exists, this information can be encoded into a DLL. Given an appropriate theory for how the source radiates sound, the solution can be implemented without much compromise. Practically, the DLL provides the data describing the radiation of sound by a particular loudspeaker and the simulation software employs this data to model the interaction of the source with the room, Fig. 39-24.

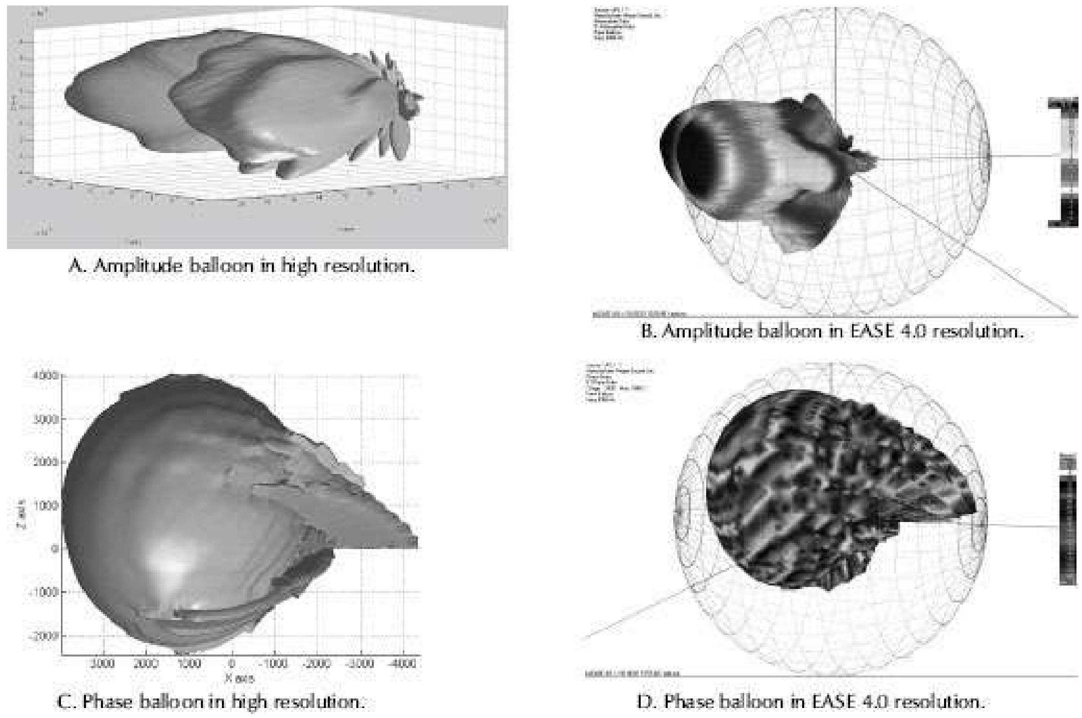


Figure 39-22. High resolution and EASE 4.0 amplitude and phase presentation.

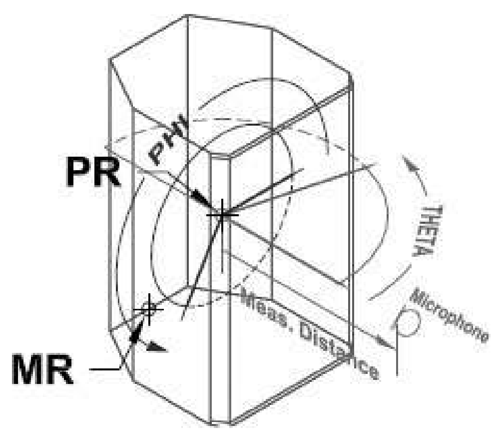


Figure 39-23. AutoCAD drawing of a loudspeaker showing the Point of Rotation PR and the case (mechanical) reference point MR.

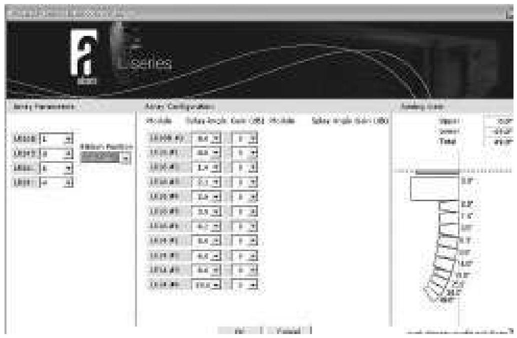


Figure 39-24. Screenshot of an EASE DLL.

Compared to conventional, tabular data formats this flexibility is unequalled. It is obvious that with a mathematical model of sufficient accuracy many of the previously discussed issues can be resolved. But at the same time, the development of a DLL module requires some programming effort. As an encoded binary file it is also proprietary to the loudspeaker manufacturer, so normally the user of simulation software and DLL plugin cannot determine how the loudspeaker system is actually modeled. Unless sufficient information is published by the creator of the DLL, the end user cannot estimate the level of prediction accuracy the model provides.

39.3.1.2.2 Modeling by Means of a GLL Data File

While trying to solve the same problems, technically the GLL concept[41] takes a different path compared to the DLL philosophy. Based on the experience with many loudspeaker manufacturing companies and the implementation of simulation and measurement software packages the generic loudspeaker library (GLL) was developed as an object-oriented description language to define the acoustic, mechanical, and electronic properties of loudspeaker systems, Table 39-1. Since for each physical entity the GLL language has a representation in the software domain, there is no need to make artificial assumptions in order to comply with rigid, reduced data structures. Basically, in the GLL philosophy every sound radiating object should be modeled as such and every interaction possible between engineer and loudspeaker in the real world should be imaged in the software domain. In this picture, transducers, filters, cabinets, rigging structures and a whole array or cluster are present in the GLL with their essential properties and parameters, Fig. 39-25.

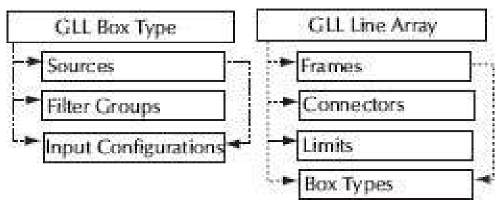


Figure 39-25. Some elementary objects of the GLL description language.

Typically, the GLL model of a loudspeaker consists of one or multiple sound sources, each with its own location, orientation, directivity, and sensitivity data. These sources can be simple point sources but also spatially extended sources, such as lines, pistons, etc. In addition to that, a complex directivity balloon based on high-resolution impulse response or complex frequency response data on a spherical grid describes the radiation behavior. With the sources representing the acoustic outputs of the loudspeaker on the one hand, the builder of a GLL defines the electronic inputs of the loudspeaker on the other hand. A filter matrix provides the logic to combine inputs and outputs, see the example in Fig. 39-26.

It can include multiple sets of filters, including IIR and FIR filters, crossover and equalization filters. The loudspeaker box is mechanically characterized by means of a case drawing and data for center of mass calculations. Boxes can be combined into arrays and clusters. Available configurations are predetermined by the loudspeaker manufacturer according to the functions available to the end user. Additional mechanical elements such as frames and connectors allow specifying exactly which configuration possibilities exist.

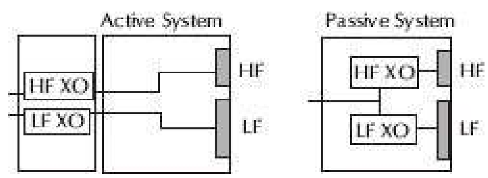


Figure 39-26. Typical GLL input configurations and filters for a two-way system.

Once all the data is assembled, the GLL is compiled into a locked, distributable file, see Fig. 39-27. In fact, the end user of a compiled GLL can only see the loudspeaker system as he would see it in the real world. The user can apply filters to the electronic inputs of the loudspeaker and he can calculate (= measure) the acoustic output of the loudspeaker. He can look at the loudspeaker case as well. When modeling arrays he may change the arrangement of boxes as allowed by the manufacturer.

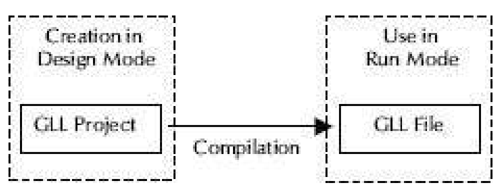


Figure 39-27. Compilation to create a GLL.

It is obvious that the GLL format provides a natural, straight-forward way to describe loudspeaker systems. By means of a GLL model any active and passive multiway loudspeakers, digitally controlled column loudspeakers, line arrays, or loudspeaker clusters can be accurately represented with regard to their acoustic, electronic, and mechanical properties. Nonetheless the GLL model will fail due to its very nature, when artificial algorithms are to be implemented which do not have a counterpart in the real world.

39.3.1.2.3 Background of Simulation and Measurement

This section reviews simulation methods and measurement requirements as well as their theoretical basis with respect to both DLL and GLL modeling approaches.

1. Resolving the Far-Field Problem.

One of the primary points to address in the simulation of the acoustic sources is the correct application of the data for the near field and far field. In the previous section we have emphasized that many loudspeakers and loudspeaker systems employed in the field today are actually used mainly in their near field, that is, at distances where the system cannot be approximated by a single point source with a distance-independent directivity pattern. Because of their size these systems can hardly be measured as a whole in their far field, anyway. However, measurements at a near field distance are only valid for use at that distance and not beyond, see Eq. 39-23.

Principally, there are two solutions to that. On the one hand, one can try to model the system as what it is, namely a spatially extended source. It could be characterized mathematically by an ideal straight or curved line source with some correction factors derived from measurements. On the other hand, already for the purposes of practical assembly, transport, and maintenance almost all large-format loudspeakers are composed of individual elements.

For example, a touring line array is built of multiple cabinets each of which house multiple transducers. Thus it seems natural that the line array is described primarily by its components and its overall radiation characteristics are derived from that. In consequence, representing the significantly smaller elements individually as point sources, now the measurement and the simulation only have to happen in the far field of the respective element. Coherent superposition of the sound waves radiated by these elements will then yield the correct behavior of the entire system for both near and far field.

2. Acquisition and Interpolation of Complex Data

The issue of using complex data instead of magnitude-only data is closely related to finding an accurate way to interpolate data points over angle and frequency properly for both magnitude and phase data. In return, using complex data on the level of individual elements eliminates the need for higher precision when measuring and interpolating data on the level of the loudspeaker system as a whole.

Critical Frequency. First, let us review the error that we make when measuring a loudspeaker directivity balloon about a given point of rotation (POR). Problems usually arise from the fact that one or several sources are slightly off-set from the POR and thus the measured data suffers a systematic error. For a given setup, Fig. 39-28, we can estimate the relative error for the magnitude data $\delta|\hat{A}|$ for large measuring distances.[46]

$$\delta|\hat{A}| = 1 - \frac{x^2}{2d^2} + \frac{x}{d}\sin\vartheta \qquad (39\text{-}33)$$

where,

x is the distance between the POR and the concerned acoustic source,

d is the measurement distance between the POR and the microphone (with $d >> x$),

ϑ is the measurement angle between the microphone and the loudspeaker axis.

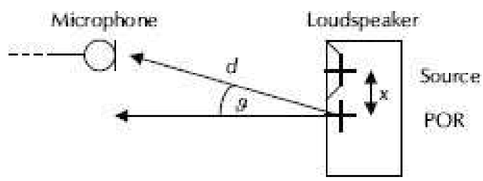


Figure 39-28. Typical setup for loudspeaker measurements.

The error is maximal for measurements where the connecting line between microphone and POR passes through both POR and acoustic source, here at an angle of ϑ = 90°. Nevertheless, for all practical cases the error is largely negligible. For example, typical values of x = 0.1m and d = 4m yield an error of only 0.2dB.

When describing loudspeakers by magnitude data only, the phase is neglected completely. To simulate the interaction between coherent sources, often the run time phase calculated from the time of flight between POR and receiver is used instead. As stated earlier, this assumes that the inherent phase response of the system is negligible and that there is an approximate so-called acoustic center where the run time phase vanishes and which must be used as the POR. For this measurement situation the systematic error in the phase data, $\delta\Phi = \delta\arg\hat{A}$, can be calculated as well.[46] For large measuring distances d, see Fig. 39-28, it is given by

$$\delta\Phi = \frac{2\pi}{\lambda}x|\sin\vartheta|$$

for magnitude-only data ($\arg \hat{A} \approx 0$)
where,
λ denotes the wavelength.

In contrast, by acquiring phase data in addition to the magnitude data, this offset error can be minimized:

$$\delta\Phi = \frac{2\pi}{\lambda}\frac{x^2}{2d}(1 - \sin^2\vartheta) \qquad (39\text{-}34)$$

for complex data $\arg \hat{A} <> 0$

Hence, by using phase data the error is reduced by an order of magnitude. In practice, it is most useful to define a maximum acceptable phase error, such as $\delta\Phi_{Crit} = \pi/4$ and use that to derive an upper (critical) frequency limit based on the measurement setup. Fig. 39-29 shows this critical frequency f_{Crit} as a function of the distance between POR and acoustic source.

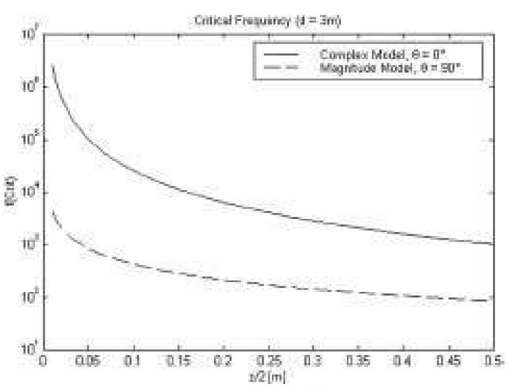


Figure 39-29. Critical frequency f_{Crit} for magnitude data (- -) and complex data (-) as a function of the distance × = z/2 between POR and acoustic source, at a measuring distance of 3m and a maximal phase error of 45°.

We emphasize that the use of phase data does not only reduce the error in the directivity data but it also largely eliminates the need to define, find and use the so-called acoustic center, the imaginary origin of the far field spherical wave front.

Local Phase Interpolation. Once complex directivity data is available for a loudspeaker, the next step is to define an appropriate interpolation function for the discrete set of data points to image the continuous sound field of the source in the real-world $(f_{Int}(\hat{A}) \rightarrow \bar{A})$. An algorithm will have to work for both magnitude and phase data in both domains: frequency and space. While averaging, smoothing and interpolating magnitude data is usually a straight-forward task, the same is not true for phase data. Due to the mathematical nature of phase, its data values are located on a circle. Accordingly, when phase is mapped to a linear scale the interpolation has to take wrapping into account. In this regard, a variety of methods has been proposed, such as use of group delay, unwrapped phase, or the so-called continuous phase representation. Although these methods have their advantages, it could be shown that generally none of them is appropriate for full-sphere radiation data of a real loudspeaker.[47]

Alternatively, a method named local phase interpolation can be applied successfully when some care is taken about the resolution of the underlying data. This method essentially interpolates phase on a local scale rather than globally. For example, let the phase average of two data points *i* and *j* be defined as

$$\langle \Phi \rangle = \frac{1}{2}\Phi_i + \frac{1}{2}\Phi_j \qquad (39\text{-}35)$$

Then, it is assumed that the corresponding phase data points are all located within a certain range:

$$|\Phi_i - \Phi_j| < \frac{\pi}{2} \qquad (39\text{-}36)$$

In this respect, i and j may represent two angular data points ϑ_i and ϑ_j or two frequencies f_1 and f_2.

Also, the averaging or interpolation function may involve more than two points.

Note that in the above case we have assumed that for calculating the absolute difference the maximum possible difference is π. This can always be accomplished by shifting the involved phase values by multiples of 2π relative to each other.

From the condition above we can derive requirements directly for the measurement. Assuming that the phase response will be usually dominated by a run time phase component due to one or several acoustic sources being located away from the POR, conditions for the spatial and spectral density of data points can be computed.[47] With respect to frequency one obtains

$$x_{crit} \approx \frac{c}{4\Delta f} \qquad (39\text{-}37)$$

where,
Δf denotes the frequency resolution,
c the speed of sound.

Given these parameters, x_{Crit} is the maximal distance allowed between POR and acoustic source at the given frequency resolution. With regard to angle one finds analogously

$$x_{crit} \approx \frac{c}{4f\sin(\Delta\vartheta)} \qquad (39\text{-}38)$$

where,

f is the frequency,
$\Delta\vartheta$ is the angular resolution.

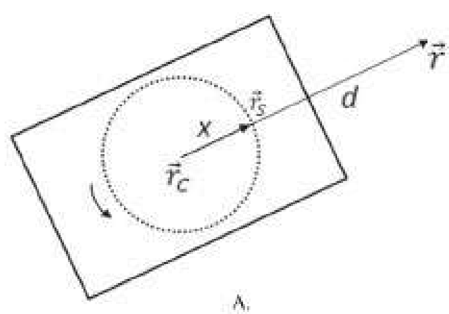
As an example, these limits correspond roughly to a measurement setup where the acoustic source is not farther away than ca. 0.15m from the POR. Phase data points will be close enough up to a frequency of 8kHz, if the frequency resolution is at least 1/12 octave (or $\Delta\phi \approx 475\text{Hz}$) according to Eq. 39-36 and the angular resolution is at least 5°.

The frequency condition is illustrated by Fig. 39.30A. The slope of the phase response versus frequency is steepest when the acoustic source $\vec{r}_s$ is located on the line connecting the point of rotation $\vec{r}_c$ and the microphone $\vec{r}$. Accordingly, Eq. 39.35 determines the required frequency resolution Δf based on the distance $x = x_{crit}$. Similarly, the angular condition is indicated in Fig 39.30B. Here the phase difference between two adjacent angular data points with the differential angle $\Delta\vartheta$ will be greatest when the line connecting the acoustic source $\vec{r}_s$ and the point of rotation $\vec{r}_c$ is perpendicular to the measurement axis. Based on the distance $x = x_{crit}$ this yields the required angular resolution according to Eq. 39.36.

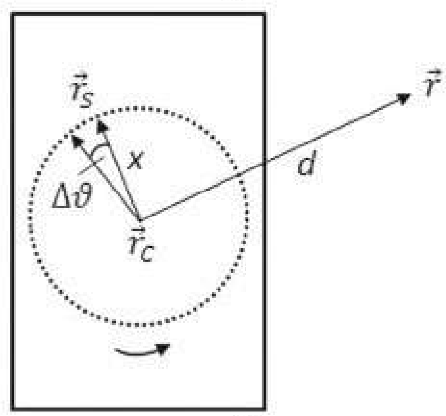
Such conditions are well within of what is possible with modern measurement platforms.

Data Acquisition. Although loudspeaker performance data and directivity patterns have been measured for several decades, no definitive standard has emerged from that practice. Also for some years, the AES standards committees try to unify the variety of existing methods and concepts to reach some commonly accepted measuring recommendations.

The accurate measurement of loudspeaker polar data is one of the issues of the on-going discussion. Especially the acquisition of complex frequency response data asks for significantly higher accuracy in the measurement setup and better control of the environment than the measurement of magnitude-only data. To determine the exact phase response of the loudspeaker under test relative to the POR, it is inevitable to measure and compensate the measuring distance as well as the environmental conditions that influence the propagation of sound along that path. For example, to be exact within a quarter of a wavelength at 8kHz, all distance measurements must be accurate within less than a centimeter of length. Although this is not a trivial task, professional acoustic laboratories have been built at the factories of manufacturers, at universities, and by independent service providers.[48] As a result, today many loudspeaker systems are measured using measurement platforms that can provide high-resolution impulse response or complex frequency response data.



A.



B.

Figure 39-30. A. Phase difference between adjacent frequency data points is maximal when the source is located on the line through reference point and measurement point. B. Phase difference between adjacent angular data points is maximal when the source is close to the perpendicular of the line through reference point and measurement point.

But it is important to note that gathering measurement data as described above only slightly increases the overall effort. To build a measurement setup capable of acquiring complex balloon data means a high initial effort, but with respect to automated polar measurements the subsequent measuring durations are the same as for magnitude-only data. The measurement of the individual components of a loudspeaker cabinet or array is obviously connected with longer measurement times. However, in many cases the angular resolution for a transducer measurement may be lower than for the full multi-way device because its directivity behavior is much smoother. In the same manner the frequency resolution can be chosen adequately. Finally, the acquisition of phase data also means that the so-called acoustic center does not have to be determined in a timeconsuming procedure. Mounting the loudspeaker for a measurement is thus much simpler. The measurement of different transducers of the same loudspeaker does not require remounting the device anymore, as well. Additionally, as we will show below, loudspeaker designers and manufacturers gain direct benefits from advanced measurement data, such as directivity prediction, crossover design, and verification capabilities.

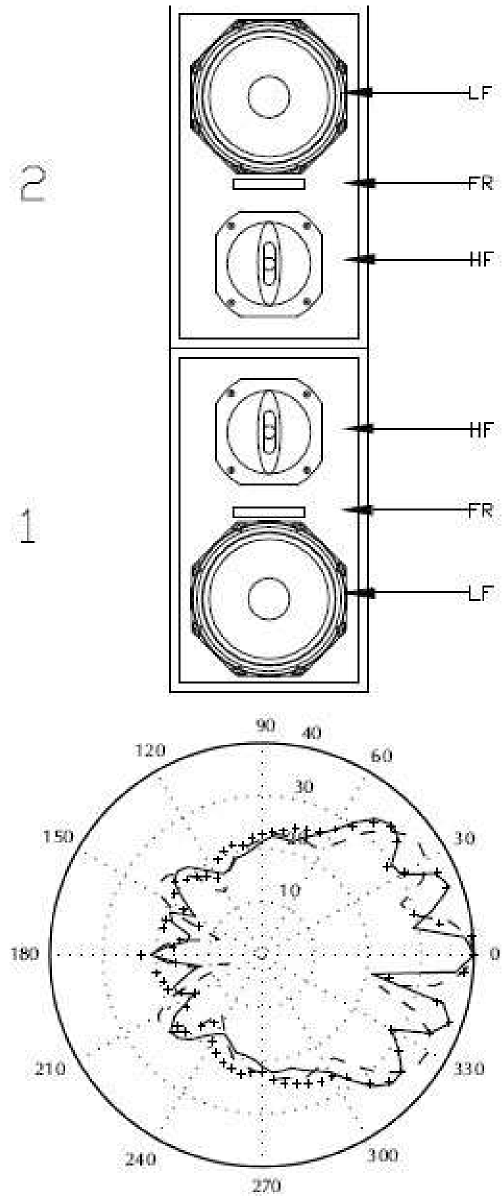
Figs. 39-31, 39-32, and 39-33 show some of the advantages gained by using complex data for individual components. Fig. 39-31 shows a comparison of measurement versus prediction for a stacked configuration of two two-way loudspeakers, arranged horn to horn (HF-HF). Its vertical directivity pattern at 1kHz is displayed in Fig. 39-31, measured data (+) and calculation based on complex data (-) are in good agreement, calculations with magnitude-only data (– –), provide erroneous results. In this case, the port of the loudspeaker (FR) was chosen to be the POR. A similar discrepancy between measurement and prediction using magnitude-only data can be seen in the arrangement of woofer to woofer (LF – LF), Fig. 39-32. To illustrate the sampling problems described before, Fig. 39-33 shows the same configuration at 4kHz. Here measurements (+) can only be imaged properly by a computation at angular increments of 2.5° (– –) using individual components measured at 5°). Computations or measurements at a too coarse resolution of 5° (– –) fail completely to describe the properties of the system when being interpolated.

Due to the complexity of establishing an accurate and phase-stable measurement setup, a set of alternative approaches is practiced. This includes, in particular, the modeling of the wave front radiated at the loudspeaker by elementary point sources according to the Huygens principle. Other models are based on the idea of deriving the missing phase response from the magnitude response, such as by the minimum phase assumption. Some of these implementations work quite well for a subset of applications, such as in the vertical domain or within some opening angle relative to the loudspeaker's axis. But generally these ideal models lack the means to depict the sound radiating properties of the loudspeaker in those domains where it is not so well behaved and analytically treatable.

3. Configurable Loudspeakers

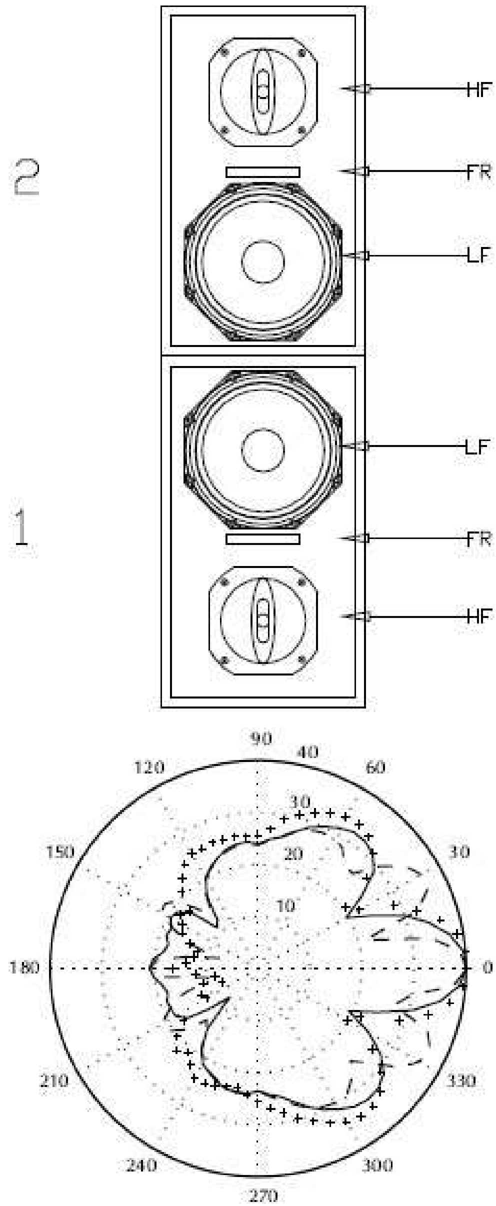
In the previous sections an overview about the crucial parts of modeling modern loudspeaker systems was given. In turn, the acquisition of complex directivity data for individual components creates the basis for another step towards resolving apparent problems on the software side. It allows including system configurability, both electronic and mechanical.

Filter settings of active and passive loudspeakers can now be taken into account in a straight forward way. We can describe the complex radiation function more precisely by including the electronic input $\bar{U}(f)$ into the system, the sensitivity of the transducer $\bar{\eta}(f)$, and the filter configuration $\bar{h}(f)$ of the system:



Measured data (+), complex data (solid) and magnitude-only data (dashed) at 1kHz and 1/3-octave bandwidth.

Figure 39-31. Comparison of measurement and prediction for the HF-HF configuration of two two-way loudspeakers (a), measured data (+), complex data (-) and magnitude-only data (– –) at 1kHz and 1/3 octave bandwidth (b).



Measured data (+), complex data (solid) and magnitude-only data (dashed) at 1 kHz and 1/3-octave bandwidth.

Figure 39-32. Comparison of measurement and prediction for the LF-LF configuration of two two-way loudspeakers, a, measured data (+), complex data (-) and magnitude-only data (–) at 1kHz and 1/3 octave bandwidth.

$$\overline{A}(\varphi, \vartheta, f) = \overline{\Gamma}(\varphi, \vartheta, f)\overline{\eta}(f)\overline{h}(f)\overline{U}(f) \qquad (39\text{-}39)$$

where,

$\Gamma(\varphi, \vartheta, f)$ denotes the angle- and frequency-dependent directivity ratio.

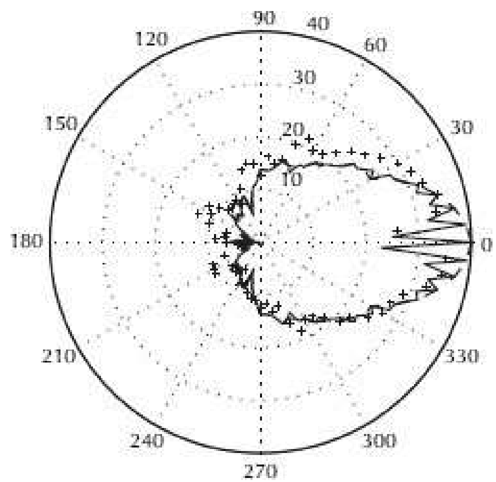


Figure 39-33. Comparison of measurement and prediction for the LF-LF configuration of two two-way loudspeakers, measured data at 5° angular resolution (+), complex data at 5° angular resolution (- -) and complex data (-) at 2.5° angular resolution, at 1kHz and 1/3rd octave bandwidth.

Correspondingly, the coherent pressure sum of several components of a system is expressed by:

$$\bar{p}_{Sum}(\vec{r}, f) = \tag{39-40}$$

$$\sum_n \frac{\bar{\bar{\Gamma}}_n(\varphi, \vartheta, f)\bar{\eta}_n(f)\bar{h}_n(f)\bar{U}_n(f)}{|\vec{r} - \vec{r}_n|} \exp(-jk_n|\vec{r} - \vec{r}_n|) \tag{39-41}$$

This formulation relates principally to Eq. 39-6 with the equalities of $E_K \sim |\bar{\eta}_n(f)\bar{h}_n(f)|$ and $P \sim |\bar{U}_n(f)|^2$. The loudspeaker properties $\Gamma_n(\varphi, \vartheta, f)$ and $\bar{\eta}_n(f)$ will normally be measured and the parameters $\bar{h}_n(f)$ and $\bar{h}(f)$ are defined by the manufacturer or end user. As a result, this concept allows one to model the full response of a multi-component system under consideration of the given filter settings, may it be a multiway loudspeaker, a digitally steered column, or a concert sound line array. Evidently the general filter transfer function $\vec{h}(f)$ can represent IIR filters for equalization as well as FIR filters for beam steering such as with FIRmaker.[37] Of course, the effect of changing crossover parameters on the directivity chaacteristics can be calculated also.[49] An example is shown in Fig. 39-34.

A second step can now be taken as well. The mechanical variability of touring line arrays or clusters can be considered by defining $\vec{r}_n$, either directly by its coordinates or indirectly as a function of user-defined parameters, such as the mounting height of the system and the splay angles between individual cabinets.

4. Shadowing and Ray-Tracing

It has been pointed out earlier, that for a large-format loudspeaker system the use of a single point as the origin for raytracing- or particle-based methods is not adequate. On the other hand, it is not practical to use all individual acoustic sources as origins for the raytracing process, given available computing power and the geometrical accuracy of the model. But that is not necessary anyway, since the raytracing algorithm can be run for subsets or groups of acoustic sources. Therefore representative points have to be found, so-called virtual center points, that can be used as particle sources, Fig. 39-35.

Typical lower frequency limits for the particle model and the level of detail in common room models suggest raytracing sources to be spaced apart by about 0.5 to 1m. In many cases this corresponds to one raytracing origin per loudspeaker cabinet. While this method of virtual center points is significantly more accurate than using a single source of rays for the whole array, it is still viable with respect to the required computational performance, compare Fig. 39-36A and B.

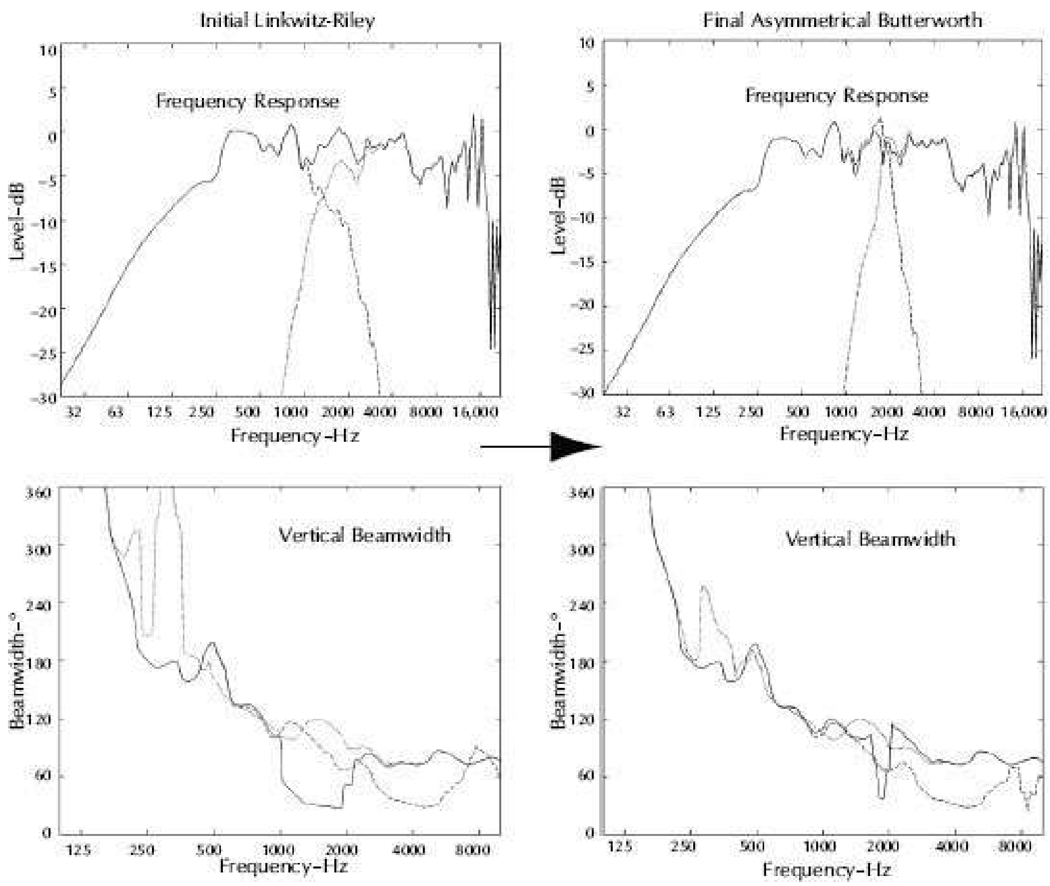


Figure 39-34. Directivity optimization with the prediction software EASE SpeakerLab. Left column shows frequency response and vertical beamwidth of a two-way loudspeaker for initial crossover filter settings, right column shows optimized frequency response and vertical beamwidth. LF unit (- –), HF unit (· ·), full-range (-).

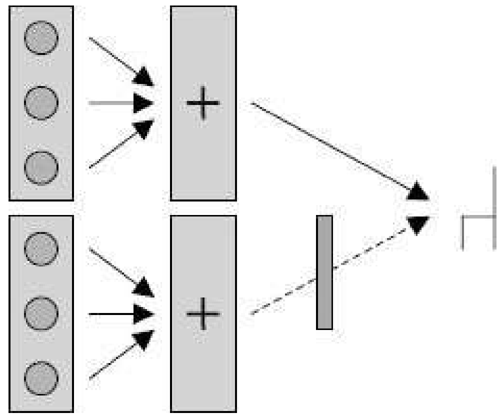


Figure 39-35. Schematic display of shadowing/raytracing using virtual center points. Loudspeaker boxes are indicated by gray rectangles, sources indicated by gray circles, virtual centers by black crosses. Only the virtual center points are used for visibility tests.

5. Additional Notes

Some other problems are also automatically resolved by modeling the components of a loudspeaker system separately. For example, the definition of maximum power handling capabilities becomes straightforward. Each component can be described individually by its maximum input level and possibly the frequency response of the test signal, preferably broadband pink noise. In this respect also the focus of the pro-audio community increasingly shifts from sometimes obscure maximum power values as defined by the loudspeaker manufacturer towards the specification of maximum voltage as the entity that is directly measured and applied in modern constant-voltage amplifiers.

Finally, one should be aware of the errors made in advanced modeling approaches like the GLL or DLL. It is clear that the acquisition of complex data requires more care and thus engineers will initially see significant measuring errors, especially with respect to the repeatability of measurements. By refining the measurement setup, using latest measurement technology and employing data averaging as well as symmetry assumptions, the data acquisition can usually be improved by an order of magnitude. In addition, it must be emphasized that the variation between samples of the same loudspeaker model may be larger than the measuring error. However, this depends strongly on the manufacturer and its level of quality control.

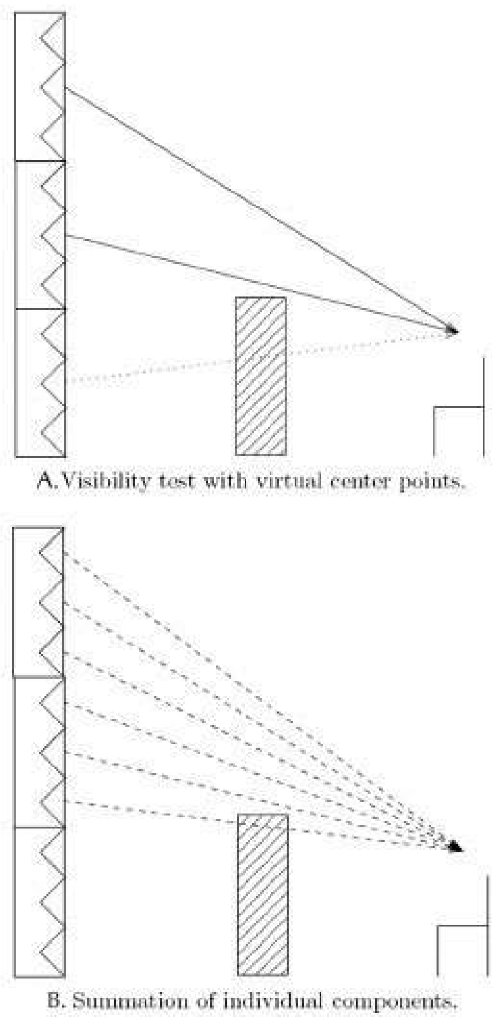


Figure 39-36. Geometrical visibility tests. Triangles indicate individual sound sources; sources are arranged in groups indicated by rectangles

From the point of view of the simulation software, the best known practices should be assumed. There is not much sense in limiting the capabilities of an acoustic simulation package because of the quality of the most inexpensive loudspeaker boxes. Like for the geometrical and acoustic model of room, the "garbage in – garbage out" principle holds true for the sound system part of the room as well and the user must be aware of that.

39.3.2 Receiver Simulation

For a complete acoustic model, the acoustic receivers must also be considered. Most important for auralization purposes is to account for the characteristics of the human head and how it influences the sound that reaches the inner ear. Often, simulation software packages also allow utilizing microphone directivity data, in order to be able to image real-world measurement. However, it must be stated that in general the correct implementation of electroacoustic receivers has not nearly received the same level of attention as the sources.

39.3.2.1 Simulation of the Human Head

Central to incorporating the characteristics of the human head into the simulation results and thus preparing them for final auralization purposes is the head-related transfer function. Typically, this is a data set that consists of two directivity balloons, one for the left ear and a second one for the right ear. Each of them describes, usually by means of complex data, how the human head and the outer part of the ear change the incoming sound waves as they arrive at the ear. It is critical for a satisfactory binaural auralization that the signal for each ear is weighted with an appropriate angle- and frequency-dependent directivity function.

The acquisition of measurement data for the human head is not a trivial matter. Since real human heads cannot be measured directly, a so-called dummy head has to be built or in case of a human head in-ear microphones have to be used (see section 39.1.5.1 Human Ears). Each ear of a dummy or a real head is equipped with a microphone. Balloon measurements are made similar to loudspeaker balloon measurements, only that the locations of source and receiver are inversed and a stereo set of data files is obtained.[34]

Recent research[50] has shown that the inclusion of the human torso into the HRTF also has significant effect on the quality of the binaural reproduction. Even more so, auralization results of highest quality can be obtained utilizing a head-tracking system and a set of HRTF balloons, where each pair of balloons describes the transfer function for the left and right ear for a particular angular position of the human head relative to the human body. This data can then be employed to auralize impulse responses of either a measured or simulated environment with speech and music contents.

39.3.2.2 Simulation of Microphones

The need for inclusion of microphones in acoustic simulation software has several reasons. On the one hand, to be able to compare measurements with computational results, the frequency response and the directivity characteristics of the microphone have to be taken into account. On the other hand, the possibility to simulate either recording or reinforcement of a talker or musician is of practical interest too. For example, by varying the location and orientation of the pick-up microphones the coverage can be optimized. Finally, by including microphones it becomes possible to simulate the entire chain of sound reinforcement, from the source over the microphone to the loudspeaker and back to the microphone. Only this enables the prediction of feedback and to estimate the potential gain before feedback.

However, the acquisition and distribution of microphone data must still be considered in its infancy. Available data consists largely of octave-based magnitude-only data that assumes axial symmetry. Measurement techniques vary significantly among microphone manufacturers and measuring conditions, such as the measurement distance, are not standardized and often not even documented. Therefore most users of simulation programs do not consider implementing microphone data into their models, or if so, they use generic data based on ideal directional behavior, like cardioid or omnidirectional patterns.

There are several more issues that inhibit the widespread acquisition, acceptance, and use of microphone data.

- First, especially the measurement distance is important with respect to the acquisition of the data and its application in the software domain. A lot of microphones exhibit the so-called proximity effect, that is, the property that their frequency response and directivity function change depending on the shape of the incident wave front. This effect is most visible if the acoustic source is within a few meters range of the microphone and thus the wave front cannot be considered as plane anymore.
- Secondly, we described earlier with respect to loudspeaker data that it is important to preserve configurability also in the software domain. In this regard, switchable multi-pattern microphones have to be taken into account when developing a fully descriptive

data model.

- The use of combined microphones is also wide spread. In particular, multichannel receivers, such as dummy heads, coincidence recording microphones, or B-format receivers, need to find an appropriate representation in the simulation software.
- Another issue of concern is the acquisition of phase data. The impact of neglecting the phase of the loudspeaker on the simulation of its performance is well known. But not much research has happened in that respect regarding microphones. Nevertheless, it is clear that under special circumstances like in feedback situations or for the electronic combination of microphone signals (e.g., two active microphones on lecterns) phase plays an important role.
- Finally, of course it must be stated that the usability of microphone data has its limitations depending on the application of the particular model. Compared to installation microphones typical handheld microphones have different properties. The data that is needed and that can be acquired may differ accordingly.

Recently an advanced data model was proposed that is able to resolve many of the issues listed above.[51] Basically, it suggests to use a similar approach like for the loudspeaker description language (GLL) introduced earlier, namely to describe receiver systems in a generalized, object-oriented way. This means especially that:

- Microphone data files should at least include far-field data (plane wave assumption), but can also contain proximity data for various near-field distances.
- A microphone model can consist of multiple receivers, that is,

acoustic inputs, and can have multiple channels (electronic outputs).

- A switchable microphone should be represented by a set of corresponding data subsets.
- Impulse response or complex frequency response data should be utilized to describe the sensitivity and the directional properties of the microphone as appropriate.

As an example for an import function in the new EASE Microphone Database software still under construction see Fig. 39-37.

39.4 Tools of Simulation

Today, an acoustic CAD program must be able to predict all needed acoustic measures exactly enough. A 100% forecast is certainly impossible but the results of a computer simulation must come close to reality (errors generally about or less than 30%). Then it becomes also possible that the acoustic behavior of a facility can be made audible by a so-called auralization. (You will listen to sound events just “performed” by means of the computer.) The following will give a short introduction of the possibilities of computer simulation today.

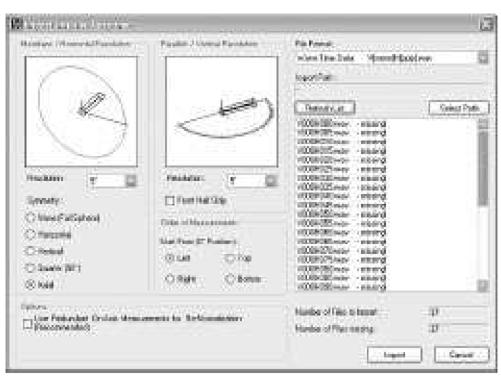


Figure 39-37. Import routine for measured microphone data.

39.4.1 Room Acoustic Simulation

39.4.1.1 Statistical Approach

Based on simple room data and the associated surface absorption coefficients a computer program is able to calculate the reverberation time according the Sabine and Norris-Eyring equations, see Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, section 9.2.1.1. On the other side measured values must be usable directly in such a program. Calculation of the Early Decay Time (EDT) should be possible too.

A comprehensive database of country-specific and international wall materials and their absorption coefficients is part of the program. This database should be accessible to allow the user to import and enter data from other textbook sources or measurements. Because most of the needed scattering coefficients are not available in textbooks a computer program should allow deriving values even by rules of thumb.

A set of frequency-dependent "target" reverberation times should be available for entering into the simulation program so that the room-model calculated (or real-world measured) reverberation times (RT) can be compared with the "target" values. The program should then indicate (for each selected frequency band) the calculated (or measured) RT vs. the "target" RT and list the number of excess or deficient RTs for each band relative to the "target" values within a range of tolerances, Fig. 39-38.

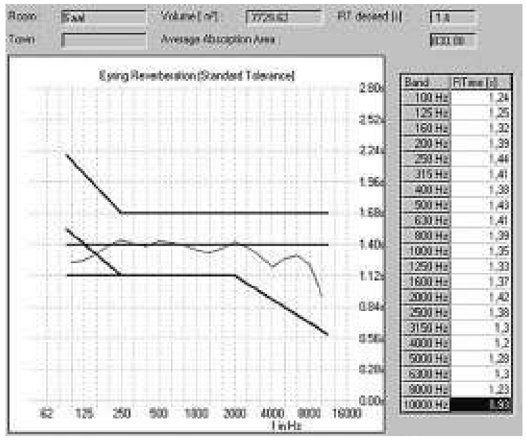


Figure 39-38. RT chart with tolerance range.

The graph of RTs should allow plotting multiple RT values within a single graph, so as to show the impact of various audience sizes, proposed and/or alternative room treatments, etc., on the RT. An option must allow plotting a grayed or dashed area as the desirable range of reverberation times for a particular project, against which the measured or calculated RT values can be referenced.

39.4.1.2 Objective Room-Acoustic Measures

The simplest way to obtain objective measures is to use the direct sound of one or more sources and calculate the reverberation level of the room by means of the reverberation time equations assuming the room follows a statistically even distributed sound decay (homogeneous, isotropic diffuse sound field, that is, the RT is constant over the room). From these calculations it is possible to derive the direct sound and the diffuse-sound levels and consequently a range of objective acoustic parameters, see Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, section 9.1. It goes without saying that this requires the acoustical conditions of the room to show a statistically regular behavior (frequency response of the reverberation time that is independent of the location considered in the room). In practice, however, such behavior will hardly be found. For this reason one tends to qualify such data as having only a preliminary guideline character and to have them confirmed by additional detailed investigations.

39.4.2 Ray-Tracing or Image Modeling Approach

39.4.2.1 Preliminary Remark

There are several ways to calculate the Impulse Response of a radiated sound event. The widest-known method is the Image Source Algorithm. Worth mentioning at this point are also the Ray-trace Method which was first known in optics, and other special procedures like Cone Tracing or Pyramid Tracing. Nowadays these procedures are more often than not used in a combined form as so-called Hybrid Procedures.

39.4.2.2 Image Modeling

With image modeling there is a source and a receiving point selected. Then a deterministic search of all image sound sources of different orders is started to compute the impulse response, Fig. 39-39.

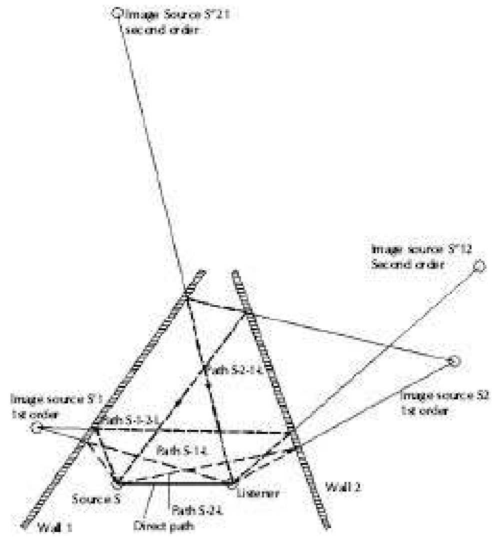


Figure 39-39. Ray calculation with Image model algorithm.

In the Image Modeling method a receiving point is used instead of a counting balloon (in contrast to classical Ray-Tracing). Frequency response and interference effects (including phase investigations) are also easily calculated.

This method is very time-consuming and the calculation time is proportional to N^i with: N = number of model walls and i = order of wall bounces.

So one gets usable results for models with N <50 and i < +6. For larger models and more complicated investigations the next method is more advantageous.

39.4.2.3 Ray-Tracing

In contrast to image modeling, here the path of a single sound particle radiated under a random angle into the room along a ray is followed. All surfaces are checked to find the reflection points (with or without absorption or diffusion). At each such point the particle energy is reduced according to the wall's absorption characteristics. Then the particle is reflected back into the room either under a specular angle (geometric reflection) or under a random angle (scattered sound) depending on the scattering coefficient. The tracing of the single ray is terminated when the remaining sound energy has decreased to a certain level or when the particle hits an appropriately arranged counting balloon with a finite diameter, typically at the location of a listener in the room, Fig. 39-40.

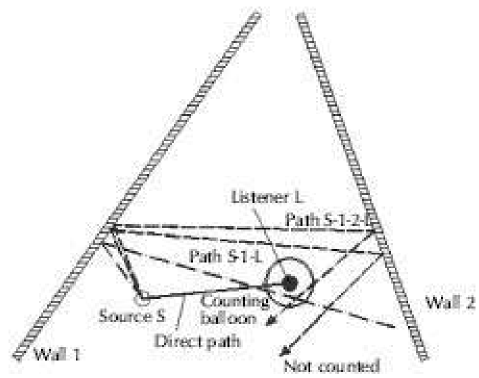


Figure 39-40. Ray calculation with Ray-Tracing algorithm.

Phase considerations are not possible directly, but can be derived if an image model routine is run that retraces the last ray after intercepting the counting balloon or if the ray retains the information about its sequence of reflection points throughout the process.

This method runs significantly faster and the calculation time is only proportional to the number *N* of the model walls. Ray-tracing methods can be even faster, if they are based on logarithmic search for the intersection points (~ logN).

39.4.2.4 Cone Tracing

This method is used in various CAD programs. Its advantage is the directed ray radiation over the different room angles, Fig. 39-41.

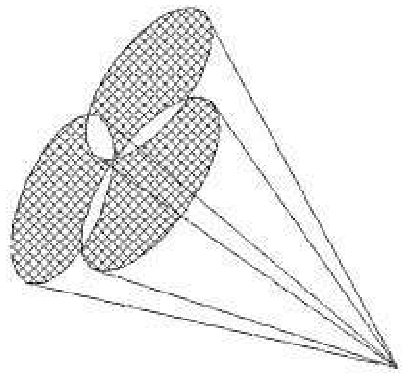


Figure 39-41. Ray radiation in cones.

Because of these cones fast ray calculations can proceed. The fact that the cones do not cover the source "sphere" surface completely turns out to be a disadvantage. It is necessary to overlap adjacent cones and an algorithm is required to avoid multiple detections or to "weight" the energy so that the multiple contributions produce (on average) the correct sound level. Some famous conical beam tracers are known,[52,53] implementing different techniques to correct this point.

39.4.2.5 Pyramid Tracing

This method was introduced by Farina in the program "RAMSETE" in 1995.[54]

Farina demonstrated that the Pyramid Beams do not suffer from the cone-trace overlap, as adjacent pyramids cover perfectly the source "sphere", Fig. 39-42.

Originally a subdivision of the surface in triangles was made by subsequent subdivisions of the 8 "octants" of the sphere. According to Farina "this way the number of pyramids generated can be any power of 2, and all of them have almost the same base area, giving a nearly isotropic sound source."

39.4.2.6 Room-Acoustic Analysis Module AURA

To illustrate these methods, as an example the new Hybrid Raytracing algorithm AURA will be explained in more detail in the following.

The AURA algorithm[55,56,57] calculates the transfer function of a room for a given receiver point using the active sound sources. For this purpose a hybrid model is employed that uses an exact image-based raytracing model for early specular reflections and an energy-based ray-tracing model for late and scattered reflections. The transition between the two models is determined by a fixed reflection order.

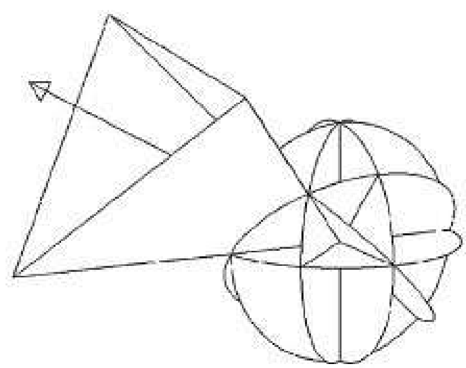


Figure 39-42. Ray radiation in pyramids.

The ray-tracing model utilizes a probabilistic particle approach and can therefore be understood as a Monte-Carlo model. At first, the sound source emits a particle in a randomly selected direction with a given energy. The particle is then traced through the room until it either hits a boundary or a receiver or its time of flight reaches the user-defined cut-off time. When the particle hits a boundary it is attenuated according to the surface material and its direction is adjusted according to the reflection law. An essential assumption of this Monte-Carlo approach is that attenuation due to air or surface reflections is taken into account as a reduction of particle energy while the propagation loss over distance is indirectly covered by the reduced detection probability for individual particles with increasing distance and fixed receiver sizes.

Per receiver and simulated frequency a so-called echogram is created which contains energy bins linearly spaced in time. When a receiver is hit, the energy of the detected particle is added to the bin that corresponds to the time of flight. Also, as a separate step, the contributions from the image source model are included. The particle model accounts for scattering in a probabilistic way. Whenever a particle hits a surface, the material absorption part is subtracted from its energy. Then, a random number is generated and depending on the scattering factor the particle is either reflected geometrically or it is scattered under a random angle based on a Lambert distribution. After that the particle is traced until it hits a receiver or a wall again.

For room acoustic models brute-force ray-tracing, that is, testing all model walls or wall triangles for intersection, is often impractical since computation time scales linearly with the number of triangles. Improved performance is obtained by structuring triangle data such that each ray is tested for intersection only with a subset of triangles. Current methods are based on two main strategies: hierarchical bounding volumes (HBV)[58] and space partitioning.[59] In the former case, a hierarchy of simple bounding volumes (such as spheres) is constructed, where a particular volume may include either a number of smaller child-volumes or actual triangles. A ray is tested for intersection starting at the top of the hierarchy, such that a particular child-volume is only tested if the parent was hit. The cost of ray-bounding volume intersection is small, and the resulting computation scaling with the number of triangles is approximately logarithmic. In space partitioning schemes, the physical space where the triangles reside is partitioned into smaller cells or so-called voxels. Rays are followed through adjacent voxels and tested only against triangles pertaining to those voxels. The partitioning may be uniform or more complex, e.g., hierarchical, adaptive, etc.

Previous studies indicate that no particular ray-tracing acceleration structure is obviously the most efficient, since the total computation cost depends both on algorithm and hardware implementation.[60] Whereas highly refined hierarchical acceleration schemes may require less intersection tests, the associated data structures are nonuniform (i.e., hard to parallelize), involve traversal of nonlocal data structures, and as such are less suitable for cache and vector processing optimizations as available on modern processors and graphics cards. On the other hand, space partitioning methods, in particular those involving simple data structures like uniform grids, are more suitable to efficient implementation on vector processing elements.

In AURA a uniform grid ray-tracing algorithm is implemented similar to Amanatides and Woo.[61] A 3D uniform grid is assigned to the simulation box and each triangle is associated with every cell having a common interior point with it. The grid spacing in every direction is determined automatically via an empirical formula: the number of cells on each axis is proportional to the square root of the total number of triangles and to the box length along the axis divided by the average box dimension (since in general the triangles form approximately a 2D shell, such a formula matches the average cell dimension to the average triangle dimension). Up to 64 cells per axis are allowed, in order to limit memory requirements. Given a ray specified by an origin and direction vector, a fast grid traversal algorithm computes the next grid cell intersected by the ray. Each triangle associated with this grid cell is then tested for intersection with the ray. No particular optimization is done to avoid duplicate ray-triangle tests when one triangle spans multiple voxels. Thus a ray-triangle intersection is only considered if it occurs within the boundaries of the current cell. The grid traversal continues until a hit point is found or the ray exits the simulation box.

39.4.2.7 Features of All These Methods

All of the ray-tracing or image modeling methods that calculate impulse responses have to take into account the directivity of the sound sources and the absorptive and scattering characteristics of the surfaces encountered en route from the source to the receiving point.

The design program must allow the user to designate specific surfaces/planes as being reflective or non-reflective. This will make it possible to simulate not only sound-reflecting walls, but also simplified floor planes, i. e., which in reality are complex shapes such as seating areas or orchestra stages or pits. At present these methods use statistical absorption factors that are readily available instead of angle-dependent ones, (for the latter no sources are available in textbooks), as well as some scattering factors estimated by rule of thumb and/or specially measured scattering factors. The diffraction behavior is still in the academic stage and some program approaches are using FEM or BEM methods[32,62], see also section 39.1.4.2. and Chapter 9, *Room-Acoustical Fundamentals for Auditoriums and Concert Halls*, Fig. 9-46. Additionally the dissipation of sound energy in air, i.e., the frequency-dependent air attenuation must be considered too.

A “library” of potential “natural” sound sources must be available, such as the human voice and various orchestral instruments/sections to go along with the electroacoustic sources/loudspeakers that should include the sound power level and directivity of these sources/loudspeakers. As a result of all these calculations you get impulse responses or energy-time curves as shown in the following figures.

The program CATT-Acoustic[44] shows the complete echogram with all input data (room, loudspeaker, listener position, frequency) and presents all resulting room acoustic measures this way, Fig. 39-43. With EASE and AURA it looks different, Fig. 39-44.

The calculated energy time curve should be able to be “stepped through” reflection by reflection, with the appropriate “rays” and surfaces being highlighted to indicate the ray’s path and the surfaces it encounters en route from source to receiver, Fig. 39-45. The software should indicate median/lateral/horizontal positioning of energy arrivals (and relative magnitude as well) at the receiver’s location, Fig. 39-46.

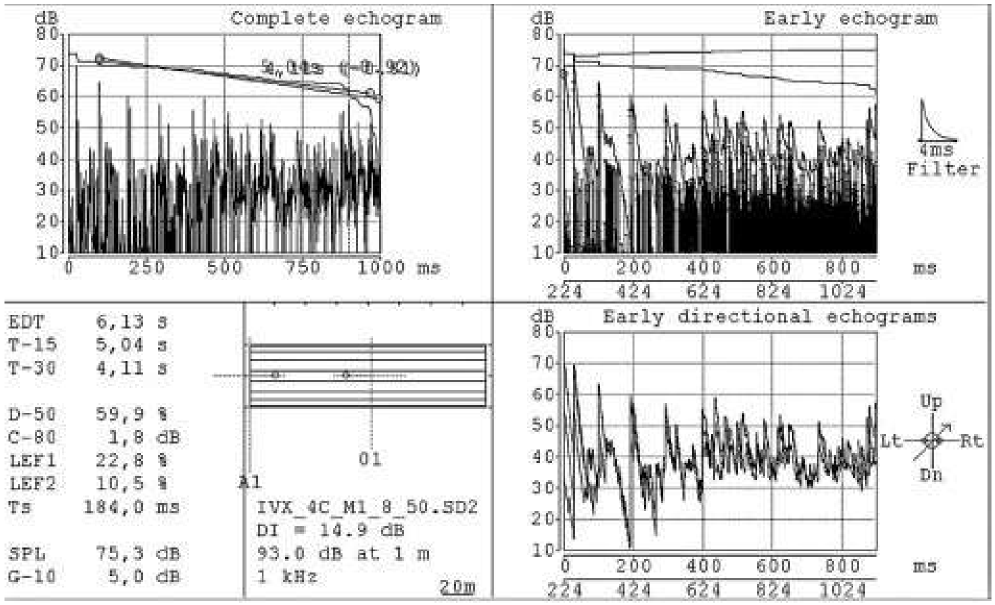


Figure 39-43. Echo and data plot in CATT acoustics.

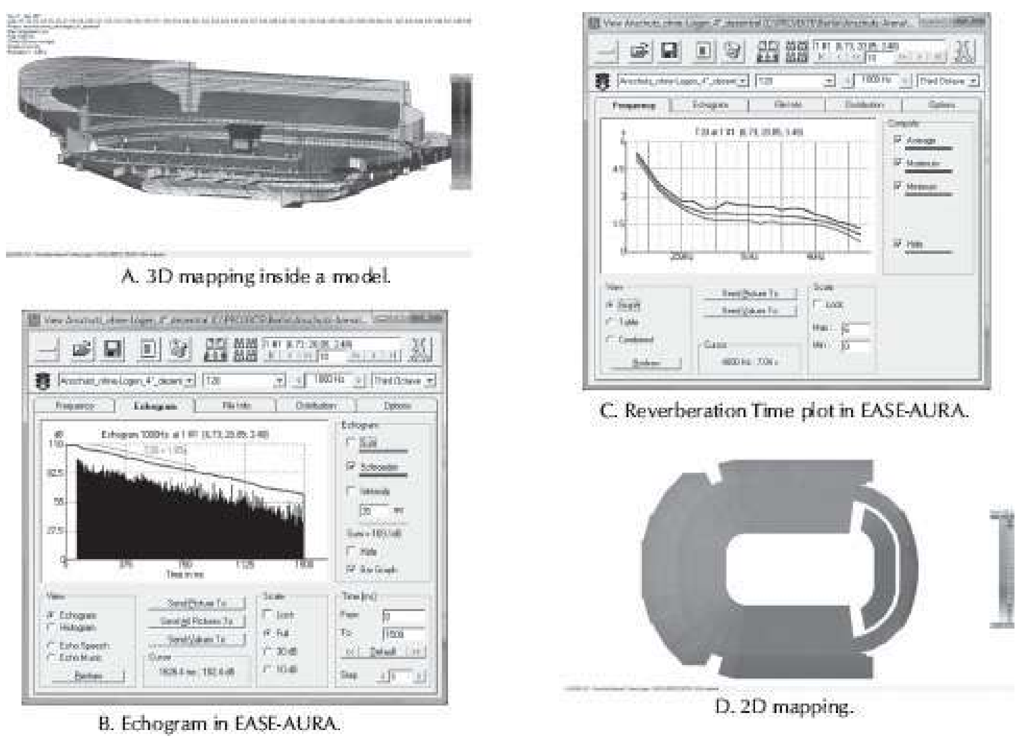
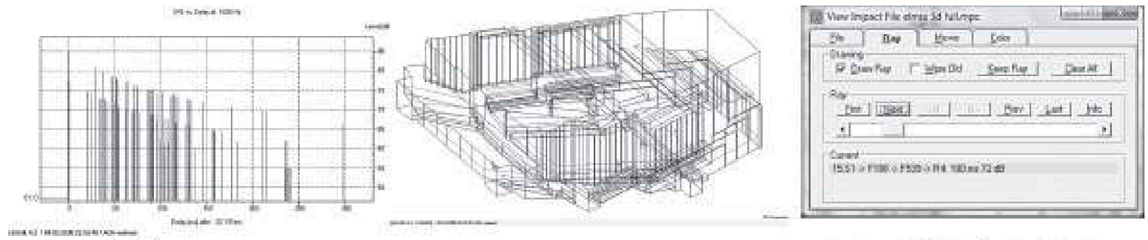


Figure 39-44. Echo and data plot in EASE 4.3



A. Reflectogram. B. Ray visualization in an EASE model. C. Control folder for visualization.

Figure 39-45. Reflectogram and ray visualization in EASE 4.3.

Additionally, a simulation program should provide the capability to calculate early/late energy ratios. It is important to be able to set the early/late transition time and also to select the cutoff time for the late energy integral, Fig. 39-47.

The software's ray-tracing or image modeling method of deriving an energy-time curve should provide the ability to indicate interaural cross-correlation (IACC) as well as lateral energy coefficient predictions at specified listener positions.

39.4.3 Auralization

The simulation program must have the ability to transfer the calculated impulse response curve to a post processing routine which will be used to auralize the room time/energy data with anechoic music or speech source material. Of course the routine must generate a binaural data file in WAV-format, Fig. 39-48 or other computer sound file format in common use. In addition to binaural data it is increasingly more interesting to use spatially enveloping loudspeaker arrangements for auralization purposes. Here multichannel data is given, e.g., in B-format, Fig. 39-49.

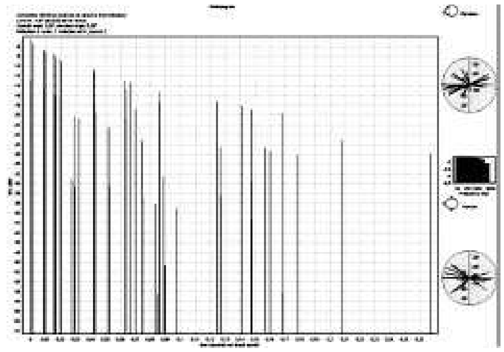


Figure 39-46. Reflectogram and Impact hedgehogs in ODEON 9.0.

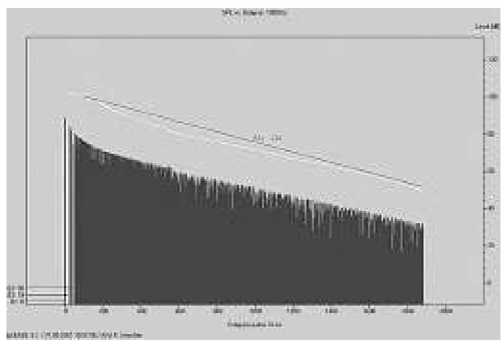


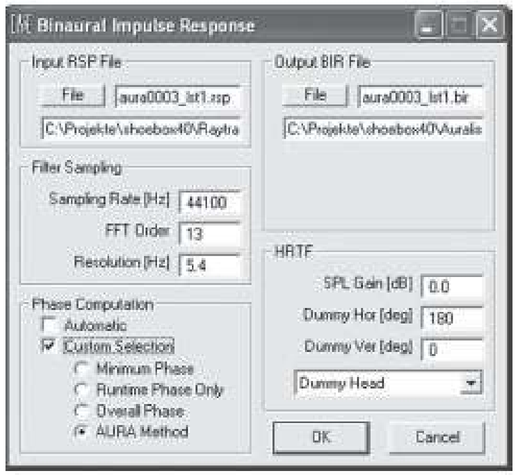
Figure 39-47. Reflectogram with tail and Schroeder plot.

39.4.4 Sound Design

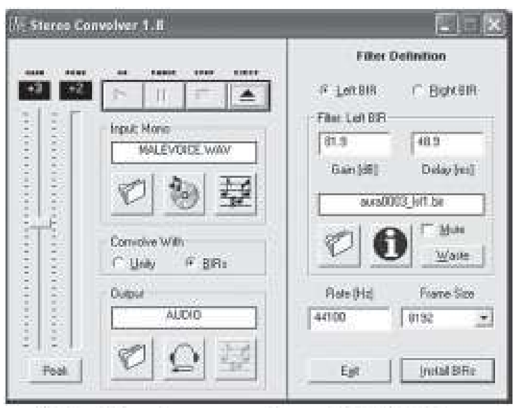
39.4.4.1 Aiming

Aiming the individual loudspeakers is an important operation ensuring the proper spatial arrangement and orientation of the sound reinforcement systems. Once the corresponding room or open-air model is at hand and the mechanical and acoustical data of the loudspeaker systems are exactly known, these systems are approximately positioned and then one may right away begin with the fine tuning of the same. A modern simulation program uses a kind of isobeam/isobar method to initially aim the loudspeakers, preferably utilizing the −3 dB, −6dB or −9dB contours.

A. Calculated binaural impulse responses.



B. Convolution with HRTF files.



C. Real time stereo convolver unit in EASE 4.3.

Figure 39-48. Auralization.

Fig. 39-50 shows various types of projection of the −3, −6, and −9dB curves into the room. On audience areas one can then also see superposed aiming curves for multiple loudspeakers, Fig. 39.51.

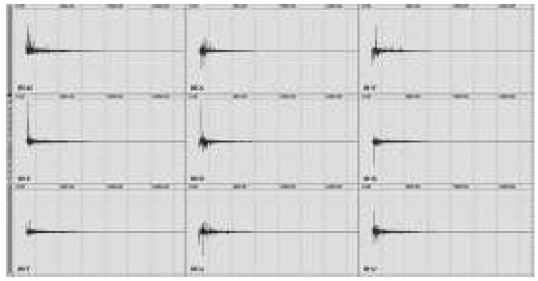


Figure 39-49. B-Format files of second order derived for Arup SoundLab.

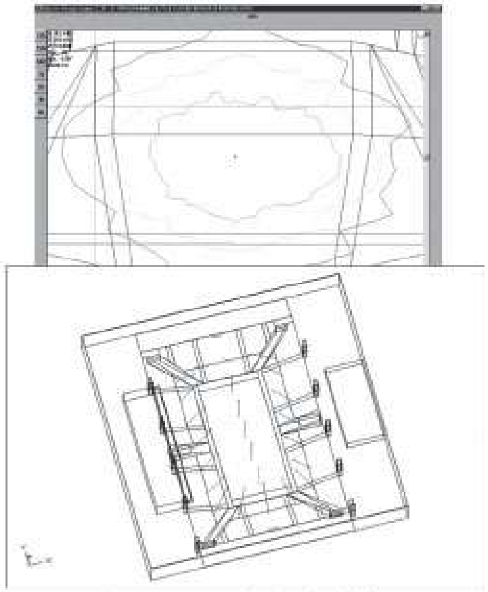
39.4.4.2 SPL Calculations

After the loudspeakers have been correctly aimed, one may begin to calculate the sound-level conditions attainable by these. The first results are given for the Direct Sound Pressure level (*SPL*). As long as we predict a good direct sound coverage over the listener area we have also to expect perfect intelligibility numbers, of course under the condition that the reverberation level is not too high.

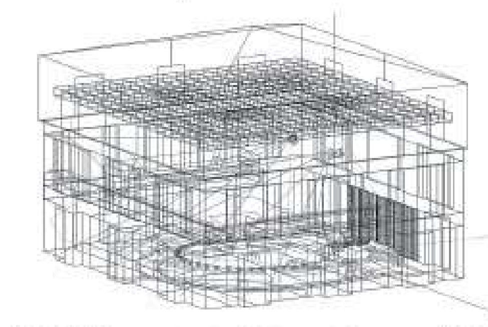
A Complex Summation (Phase conditions including travel-time differences should be included) has to be used as the standard method of calculating the Direct *SPL*. This method is exact for a planar wave, but only an approximation for the superposition of waves with different propagation directions. But anyway the complex sound pressure components of different coherent sources must first be added and afterwards squared to obtain SPL numbers. In so-called DLL or GLL approaches one always calculates the complex sum of all sources in the array.

Today simulation programs are usually still only analyzing programs, capable of calculating which levels can be obtained by which loudspeakers and under which acoustical conditions. But questions are more and more asked the other way around. New design features should query the user for a desired average *SPL* of the system, and automatically adjust the power provided to each loudspeaker (with a warning when the power required exceeds the capabilities for the loudspeaker), based upon the desired *SPL* of the design, the sensitivity and directivity of the loudspeaker, the distance of throw, and the number of loudspeakers. This presupposes, of course, new algorithms that in most of the simulation programs are just being developed, e.g., like FIRmaker by AFMG.[27]

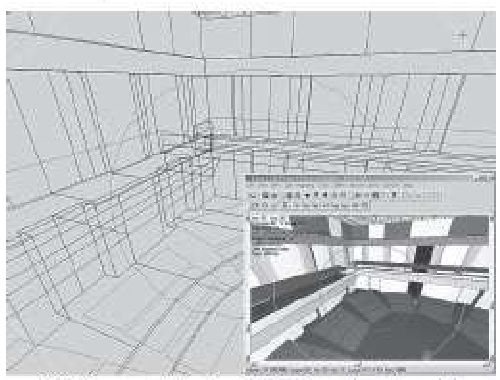
In Fig. 39-52 the level-time-frequency-behavior of a loudspeaker cluster at a chosen listener seat in a room is shown by a simulated waterfall diagram.



A. 3D presentation in ULYSSES 2.3



B. 3D presentation in EASE 4.0 wireframe model.



C. 3D presentation in EASE 4.0 rendered model.

Figure 39-50. 3D Aiming presentations in simulation programs.

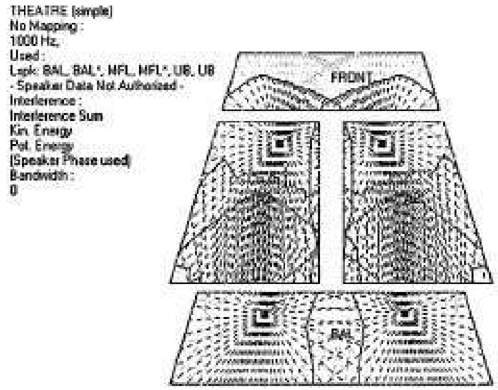


Figure 39-51. 2D aiming mapping in EASE 4.3.

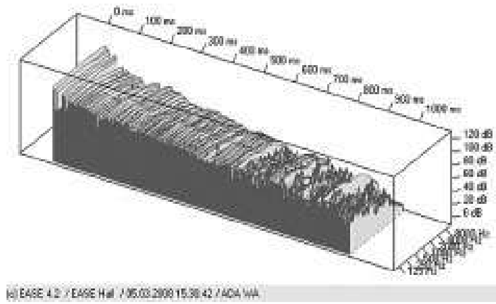


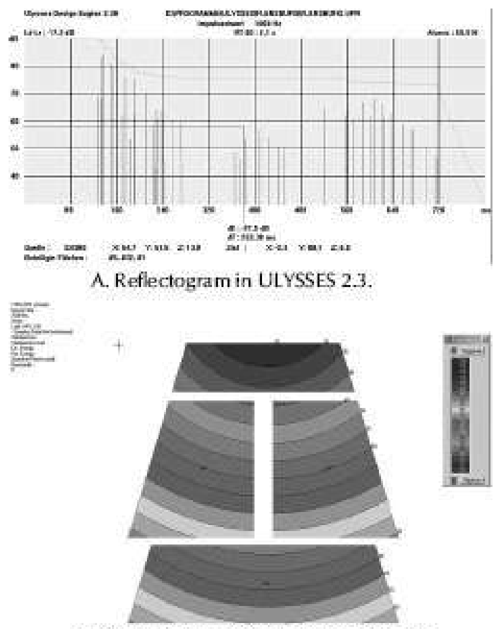
Figure 39-52. Waterfall presentation in EASE 4.2.

The target of all the efforts is to cover the whole audience area(s) evenly with musically pleasing and intelligible sound, while providing sound pressure levels suitable for the intended purpose.

All simulation programs are today widely lacking the algorithms required for computing the acoustical feedback. These, however, should soon become available, since the already available microphone data are similar in structure to those of loudspeakers, see section 39.2.1.2. Then it will be possible to calculate the maximum and nominal acoustical gain based upon the parameters of microphone/loudspeaker/listener/talker, which take into account the number of microphones and/or the arrangement of loudspeakers.

39.4.4.3 Time Arrivals, Alignment

A graph of time-arrivals (direct, direct + reflected, reflected only) should allow the user to show the first energy arrival as required by the design, to adjust a signal delay loudspeaker to bring the loudspeakers into synchronicity and to realize an acoustic localization of an amplified source (via distance and the HAAS effect), see Figs. 39-53A and B.



A. Reflectogram in ULYSSES 2.3.

B. Delay pattern of first arrival in EASE 4.3.

Figure 39-53. Delay presentations in simulation programs.

Matters are often complicated by special requirements such as localization, stereo imaging, etc. Simulation programs allow determining the first wave front as well as calculating initial time delay gaps or echo detections (c.f. in this respect Figs. 39-54A to C).

Predicted array lobing patterns of "arrayable" loudspeakers should be displayed by simulation programs, with the ability to provide signal delay and/or move the appropriate loudspeakers to attempt to bring the array into acoustic alignment. A program today will have the ability to provide signal delay to the individual loudspeakers to align them in time. The corresponding sound pressure calculations will take into account either measured phase data for the individual loudspeakers or the run-time phase if phase differences among the components can be neglected.[63]

Fig. 39-55A and B show the frequency response of nonaligned and aligned loudspeaker groups, simulated by EASE.

39.4.4.4 Mapping, Single-Point Investigations

Once the aiming, power setting, and alignments are completed, the program should provide a colored visual coverage map of the predicted sound system performance. This coverage map must take into account the properties of the loudspeakers as well as the impact of reflecting or shadowing planes, and provide as a minimum the following displays:

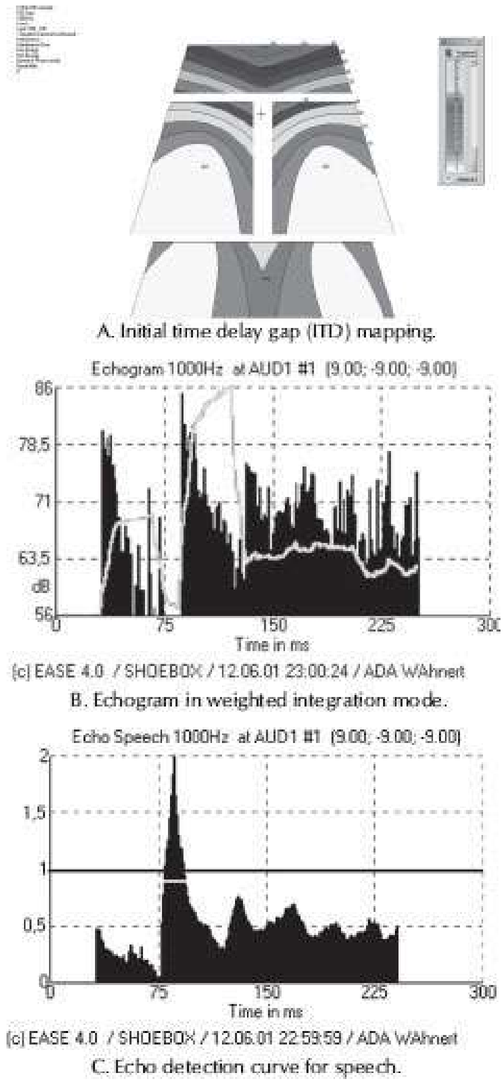
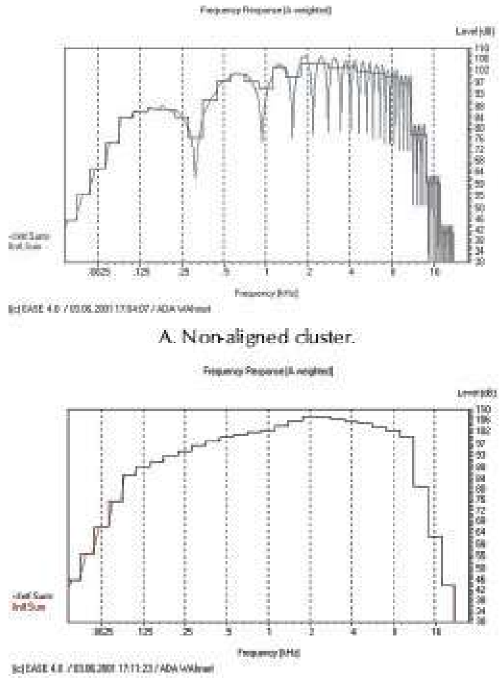


Figure 39-54. Echo detection in EASE 4.3.

- Predicted sound pressure level, viewed at octave or 1/3 octave band frequencies, and at an average of these frequencies, Figs. 39-56A–C.



A. Non-aligned cluster.

B. Aligned cluster.

Figure 39-55. Frequency response of loudspeaker cluster.

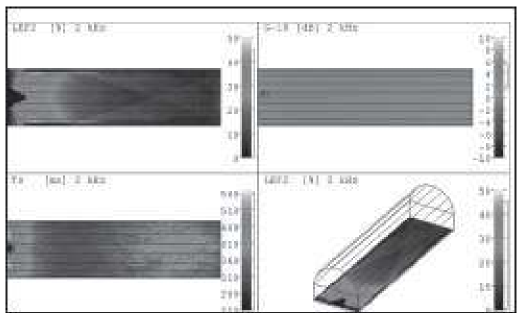
- Predicted intelligibility values (in the 2kHz octave band, or the weighted average of 500Hz to 4kHz octave band data), listed in STI or RASTI values, Figs. 39-57A and B.
- Predicted acoustic measures (for octave or 1/3 octave band frequencies), listed in C80, C50, %Alcons, Center time, Strength, or other values according to ISO standard 3382, compare Fig. 39-58A and B.

39.5 Verification of the Simulation Results

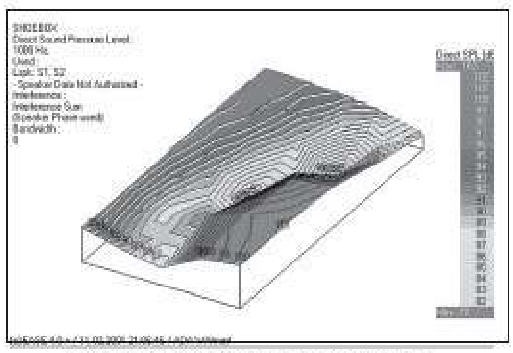
After the simulation, the practical design and the installation it is important to check the results and to compare it with the prediction. For this purpose tools are developed during the last 20 years:

- The most famous TEF 10, 12, and 20 by Crown (later Gold Line).

- MLSSA by DRA Laboratories.
- SMAART by SIA Soft.
- WinMLS by Morset Sound Development.
- DIRAC by Brüel&Kjær.
- SpectraLAB by Sound Technology Inc.
- EASERA by AFMG Berlin and SysTune by AFMG Berlin too.



A. 2D presentation in CATT acoustics.



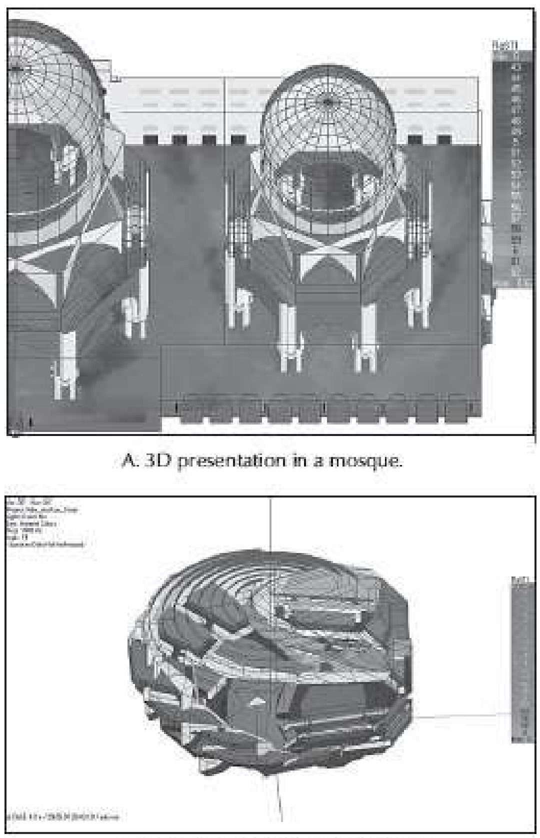
B. Narrow band presentation in EASE 4.0.

C. Broad band presentation in ULYSSES 2.3.

Figure 39-56. SPL mapping in simulation programs.

All measurements with predefined excitation signals generally utilize two or more ports. The input port of the system under test (DUT) is fed with an excitation signal, generated by the analyzer. Fig. 39-59 shows the blockdiagram for a modern software-based 4-port measurement tool including the needed AD/DA converter.

At first the unprocessed output of the DUT (raw data) is recorded and stored on the PC hard disk. Based on this original data set the corresponding processing algorithm including band pass filters or time windows can be used multiple times with different parameters to look at the parts of interest.



A. 3D presentation in a mosque.

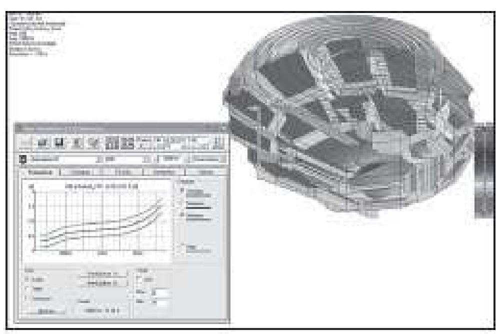
B. 3D presentation in a concert hall.

Figure 39-57. RaSTI presentations in simulation programs.

A simple way to calculate the transfer function H(ω) from the recorded raw data is to divide the measured frequency response *Y*(ω) by the frequency response of the signal *X*(ω) (or by a reference response that was previously measured). The impulse response $h(t)$ can then be computed using the inverse Fourier transform.

Until now it is common to utilize a static measuring procedure where the impulse response is derived in a separate step after every acoustic measurement. In contrast, a newly developed, dynamic method allows us to measure room acoustic impulse responses RIR in an efficient manner and to analyze this way the acoustic properties of an investigated acoustics space very user-friendly, i.e., in real time,[64] see Fig. 39-60.

Determining the impulse response by Real Time Deconvolution means in this respect that gathering the acoustic source signals and calculating the impulse response data is a simultaneous and continuous process. The dynamically derived RIR is, because of a number of optimized post-processing steps, qualitatively absolutely equivalent to a statically derived RIR and may have typical lengths of 4–10s.


A. C80 presentation.

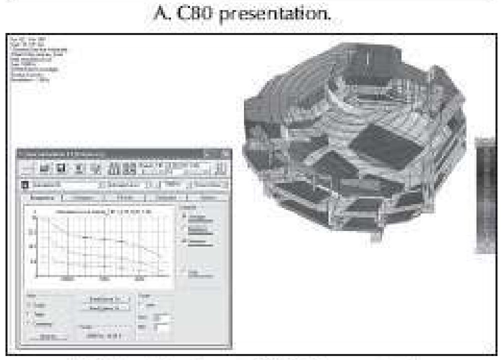

B. Articulation Loss AlCONS presentation.

Figure 39-58. AURA presentations in EASE 4.3.

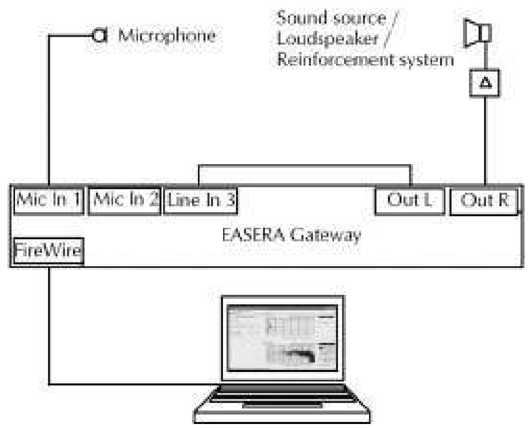


Figure 39-59. Block diagram of a modern software-based 4- port measurement tool including the needed AD/DA convertor.

The transformation between the frequency and time range is linear and of full length, analogous to the static procedure. Averaging can be likewise used to suppress the noise.

The dynamically derived RIR is because of a number of optimized postprocessing steps qualitatively absolutely equivalent to a statically derived RIR and may have typical lengths of 4–10s.

The transformation between the frequency and time range is linear and of full length, analogous to the static procedure. Averaging can be likewise used to suppress the noise.

The real time ability of the measuring system is based on very high refresh rates for the calculation of results and their display and analysis (approximately 10/s).

One can understand such a measuring system also as an "oscilloscope for room impulse responses." Possible changes of the acoustic behavior may be seen immediately and directly.

In Fig. 39-60, the excitation is done with noise, sweep or MLS. In a live situation this will be quite often annoying and cannot be done under all circumstances. So the next step is to use running music or speech signals as excitation signals and this way to derive impulse responses too. Fig. 39-61 shows a block diagram for such a tool usable with natural signals like music or speech and in Fig. 39-62 the graphic user interface of SysTune[65] is shown, a tool that allows such kind of measurements.

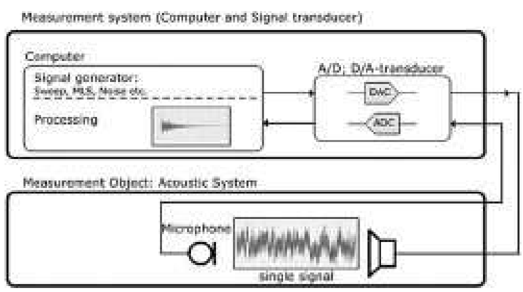


Figure 39-60. Static type of FFT-based measurements.

Once the IR has been computed in either a dynamic or static manner, electroacoustic and room-acoustic measures can be derived from it, such as the RT, D/R ratio, C50, or STI. These values can then be compared with the modeling results.

When performing such a comparison, it is always necessary to estimate the errors on each side, measurement and simulation, quantitatively in order to determine the significance of the deviations. The agreement between the results will depend on the degree to which measurement and model can provide reliable results.

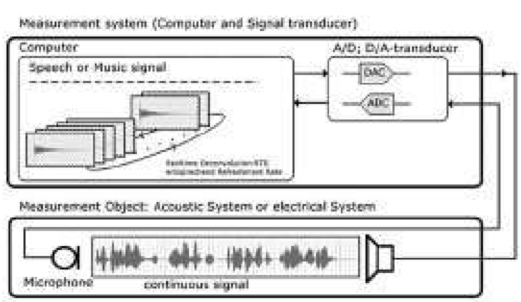


Figure 39-61. Dynamic (continuous) type of FFT-based measurements.

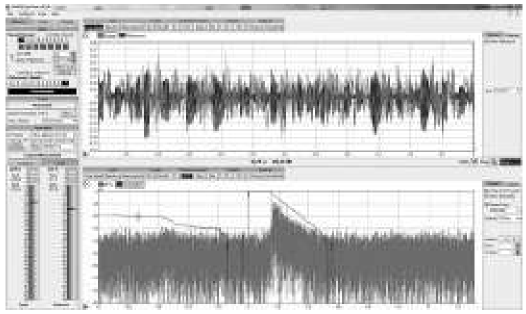


Figure 39-62. Graphic user interface of SysTune.

References

1. EASE 4.2 for Windows, User's Manual & Tutorial, Renkus-Heinz Inc., March 2007.

2. W. Reichardt, A. Kussev, Ein- und Ausschwingvorgang von Musikinstrumenten und integrierte Hüllkurven ganzer Musikinstrumentengruppen eines Orchesters (Initial and decay transients of complete instrumental groups and envelopes of an orchestra). Zeitschrift für elektrische Informations- und Energietechnik 3 (1973) 2, pp. 73–88.

3. *Taschenbuch der Akustik* (Handbook of acoustics) (Editor: W. Fasold, W. Kraak, W. Schirmer,). Verlag Technik Berlin, 1984.

4. H.-P. Tennhardt, Modellmessverfahren für Balanceuntersuchungen bei Musikdarbietungen am Beispiel der Projektierung des Großen Saales im Neuen Gewandhaus Leipzig (Model measuring procedures for balance investigations of music performances demonstrated on the example of the great hall of the Neues Gewandhaus Leipzig). *Acustica* 56 (1984) Nr. 2 (Oktober 1984), pp. 126- 135.

5. V.O. Knudsen, *Architectural Acoustics*. John Wiley and Sons, Inc. New

York, 1932.

6. Meyer, J.: *Akustik und musikalische Aufführungspraxis* (Acoustics and music presentation practice). Verlag Erwin Bochinsky Frankfurt am Main, 1999.

7. S. Pelzer, M. Pollow, M. Vorländer, Auralization of a virtual orchestra using directivities of measured symphonic instruments. *Proc. of the 11th Congr. on Acoustics*, Montreal 2013.

8. AES-2-R (specification of loudspeaker components used in professional audio and sound reinforcement) and AES56–2008 (Sound source modeling-loudspeaker polar radiation measurements).

9. Former Standard DIN 45570.

10. IEC 60268-5 Ed. 2003 Sound system equipment Part 5: Loudspeakers.

11. W. Ahnert, F. Steffen, *Sound Reinforcement Engineering, Fundamentals and Practice*. London: E&FN Spon 2000, USA and Canada, New York: Routledge 2000.

12. OIRT Empfehlung 55/1 Technische Parameter von Studio-Abhöreinrichtungen (Matinkylä 1985).

13. H. Stenzel, Über die Richtwirkung von Schallstrahlern, Elektrische Nachrichten-Technik, Band 4 (1927), Seite 239.

14. H. Stenzel, *Leitfaden zur Berechnung von Schallvorgängen*, Verlag von Julius Springer, Berlin 1939.

15. H. F. Olson, *Elements of Acoustical Engineering*, D. van Nostrand Company, 2nd edition, New York 1947.

16. M. S. Ureda, Wave Field Synthesis with Horn Arrays, 100th AES

Convention Copenhagen/Denmark May 1996, preprint No. 4144.

17. H. Benecke, S. Sawade, *Strahlergruppen in der Beschallungstechnik*. Sonderdruck Funkpraxis 1951.

18. C. Heil, Sound Fields Radiated by Multiple Sound Sources Arrays, 92nd AES Convention, Vienna, 1992 March 24–27, preprint No. 3269.

19. Duran-Audio BV White paper: Modelling the directivity of DSP controlled loudspeaker arrays, Zaltbommel/Holland, June 2000, under www.duran-audio.nl.

20. R. Heinz, "DSP-Driven Vertical Arrays," 2006, white paper, www.rh.com.

21. http://www.ateis-international.com/index.php.

22. D. Gunness, Touring Line Array Technical Issues, White paper, EAW, August 2000, under www.eaw.com

23. http://www.meyersound.com/document-library:cal_brochure_e.pdf.

24. G.W.J. Van Beuningen, E.W. Start, Digital directivity synthesis of DSP controlled loudspeaker arrays – A new concept, DAGA 2001, Hamburg-Harburg.

25. http://www.martin-audio.com.

26. http://www.tannoypro.com.

27. http://firmaker.afmg.eu.

28. ISO 354:2003: *Acoustics*—Measurement of Sound Absorption in a Reverberation Room.

29. T. J. Cox, and P. D'Antonio, "*Acoustic Absorbers and Diffusers: Theory, design and application*," Taylor & Francis 2009.

30. E. Mommertz, Determination of scattering coefficients from the reflection directivity of architectural surfaces. Applied Acoustics 60 (2000) 201–203; see also: Vorländer, M.; Mommertz, E.: Definition and measurement of random-incidence scattering coefficients, Applied Acoustics 60 (2000) 187–199.

31. ISO 17497-1: 2004 Acoustics – Sound-scattering properties of surfaces – Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room.

32. M. Bansal, Wave Theory based Extensions to Standard Room-acoustic Particle Models, PhD Thesis, Institute of Technical Acoustics, Technical University Berlin, 2008,

33. H. Schmalle, D. Noy, S. Feistel, G. Hauser, W. Ahnert, and J. Storyk. "Accurate Acoustic Modeling of Small Rooms," presented at the 131st Convention of the Audio Engineering Society, *J. Audio Eng. Soc.*, convention paper 8457 (2011 Oct.).

34. Kemar-Head: http://www.gras.dk/00012/00330/.

35. IEC 60268-4: Ed. 3(2003), Sound system equipment. Part 4: Microphones

36. L. Kinsler, A. Frey, A. Coppens, J. Sanders, Fundamentals of Acoustics, 4th ed. (Wiley, New York, 2000).

37. W. Ahnert, S. Feistel, "Cluster Design with EASE for Windows," presented at the 106th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 47, p. 527 (1999 June), convention paper 4926.

38. J. Baird, P. Meyer, "The Analysis, Interaction, and Measurement of

Loudspeaker Far-Field Polar Patterns", presented at the 106th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.*, vol. 47, June 1999.

39. W. Ahnert, S. Feistel, J. Baird, P. Meyer, "Accurate Electroacoustic Prediction Utilizing the Complex Frequency Response of Far-Field Polar Measurements," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 357 (2000 Apr.), convention paper 5129.

40. Dynamic Link Library: http://msdn2.microsoft.com/en-us/library/ms682589.aspx.

41. S. Feistel, W. Ahnert, S. Bock, "New Data Format to Describe Complex Sound Sources," presented at the 119th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 53, pp. 1239, 1240 (2005 Dec.), convention paper 6631: GLL format specification, http://www.ada-acousticdesign.de, http://www.sda.de.

42. Ch. Hughes, in cooperation with NWAA Labs (R. Sauro), White paper: How Accurate is Your Directivity Data?, http://www.excelsior-audio.com, http://Publications.html, 2005–10.

43. W. Ahnert, C. Bourillet, S. Feistel, Phase Presentation in the Acoustic Design Program EASE. Presented at the 110th AES Convention, May 2001, special print under http://www.ada-acousticdesign.de.

44. CATT-Acoustic software, www.catt.se.

45. EASE software, http://ease.afmg.eu.

46. S. Feistel, W. Ahnert, "The Significance of Phase Data for the Acoustic Prediction of Combinations of Sound Sources," presented at the 119th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 53, p. 1240 (2005 Dec.), convention paper 6632.

47. S. Feistel, W. Ahnert, “Modeling of Loudspeaker Systems Using High-Resolution Data“, *J. Audio Eng. Soc.*, vol. 55, pp. 571–597 (2007 July/August).

48. Anselm Goertz: www.anselmgoertz.de; Pat Brown: www.etcinc.us; Ron Sauro: www.nwaalabs.com.

49. S. Feistel, W. Ahnert, C. Hughes, B. Olson, “Simulating the Directivity Behavior of Loudspeakers with Crossover Filters,” presented at the 123rd Convention of the Audio Engineering Society, New York City, NY, 2007 October 5-8, convention paper 7254.

50. C. Moldrzyk, S. Feistel, W. Ahnert, “Verfahren zur binauralen Wiedergabe akustischer Signale,” Patent No. DE 10 2006 018 490.4.

51. AFMG Measurement Workshop, 123rd AES, New York 2007.

52. B. I. Dalenbäck, Verification of Prediction Based on Randomized Tail-Corrected Cone-Tracing and Array Modeling, 137th ASA/2nd EAA Berlin (March 1999).

53. G M. Naylor, ODEON – Another hybrid room acoustical model, Applied Acoustics Vol. 38 no. 2–4, page 131 (1993).

54. A. Farina, RAMSETE – A New Pyramid Tracer for Medium and Large Scale, Proceedings of EURO-NOISE 95 Conference, Lyon 21–23 March 1995.

55. M. Vorländer, Simulation of the transient and steady state sound propagation in rooms using a new combined sound particle – image source algorithm, *J. Acoust. Soc. Am.*, 86, 172.

56. O. Schmitz, S. Feistel, W. Ahnert, M. Vorländer, Merging Software for Sound Reinforcement Systems and for Room Acoustics. Presented at the 110th AES Convention, 2001, May 12-15, Amsterdam, Preprint No. 5352.

57. S. Feistel, W. Ahnert, A. Miron, O. Schmitz, Improved Methods for Calculating Impulse Responses with EASE 4.2 AURA, 19th International Congress on Acoustics, Madrid, September 2007.

58. S.M. Rubin, T. Whitted, *Computer Graphics* 14(3), 1980.

59. A. Fujimoto, T. Tanaka, K. Iwata, *IEEE Computer Graphics and Applications* 6(4), 1986.

60. V. Havran, J. Prikryl, W. Purgathover, Tech. Rep. TR-186-2-00-14, Institute of Computer Graphics, Vienna University of Technology, 2000.

61. J. Amanatides, A. Woo, Eurographics'87 Conference Proceedings, 1987.

62. ODEON software, version 10.0, www.odeon.dk.

63. M. S. Ureda, Line Arrays: Theory and Applications, 110th AES Convention March 12-15 Amsterdam 2001, preprint No. 5304.

64. EASERA software, http://www.Easera.com.

65. SysTune software, http://SysTune.AFMG.eu.

Chapter 40

Designing for Speech Intelligibility

by Peter Mapp

40.1 Introduction

The fundamental purpose of a paging, announcement, voice alarm, or speech reinforcement system is to deliver intelligible speech to the listener. A surprising number of systems, however, fail to achieve this basic goal. There can be many reasons for this, ranging from inadequate signal-to-noise ratio to poor room acoustics or inappropriate choice or location of the loudspeaker. It is the job of the sound system designer to be aware of these factors and take them into account when designing a sound system and selecting devices to provide the degree of intelligibility required. In order to do this, however, an understanding of the basic factors that affect speech intelligibility and the way we hear speech is required. This chapter therefore begins by taking a look at the nature of the speech signal and how we hear it before discussing design strategies and ways of optimizing system design and performance. Current methods of assessing and measuring intelligibility are then also discussed together with comments on their practical limitations.

40.2 Parameters Affecting Speech Intelligibility

Although sound quality and speech intelligibility are inextricably linked, they are not the same thing. For example it is quite possible to have a poor sounding system that is highly intelligible (e.g., the frequency response limited and resonant re-entrant horn) or alternatively a high-quality system that is virtually unintelligible (e.g., a hi-fi loudspeaker in an aircraft hangar). Similarly a common mistake, often made when discussing intelligibility, is to confuse audibility with clarity. Just because a sound is audible does not mean to say that it is intelligible. Audibility relates to the ability of a listener to physically be able to hear a sound, whereas clarity describes the ability to detect the structure of the sound. In the case of speech, this means hearing the consonants and vowels correctly in order to identify the words and sentence structure and so give the speech sounds intelligible meaning.

40.3 The Nature of Speech

A speech signal involves the dimensions of sound pressure, time, and frequency. Fig. 40-1 shows some typical speech waveforms representing the numbers "one" "two," and "three." The waveforms are highly complex, with amplitudes and frequency contents that change almost millisecond by millisecond. Consonant sounds typically have durations of around 65ms and vowels 100ms. The duration of syllables is typically 300–400ms whereas complete words are about 600–900ms in length dependent on their complexity and rate of speech. When speech is transmitted into a reverberant space, local reflections and the general reverberation distort the speech waveform by smearing it in time. The reverberant tail of one syllable or word can overhang the start of the next and so mask it, thereby reducing the potential clarity and intelligibility, Fig. 40-2. Equally if the background noise level is high or more accurately if the speech signal-to-noise ratio is too low, then again parts of words or syllables become lost and intelligibility deteriorates, Fig. 40-3. There are many other factors that can affect the potential intelligibility and perceived clarity of a speech signal, the most important are summarized below.

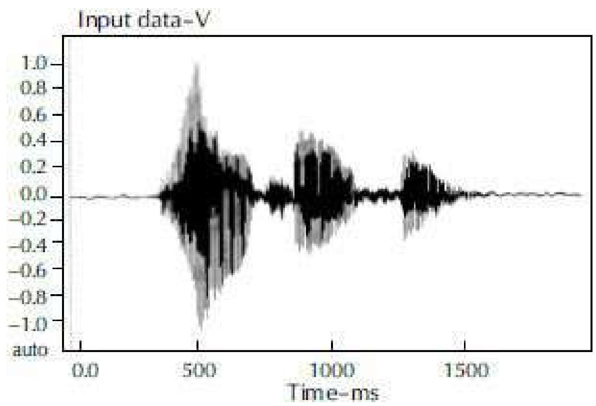


Figure 40-1. Anechoic speech waveforms for the numbers "one," "two," and "three."

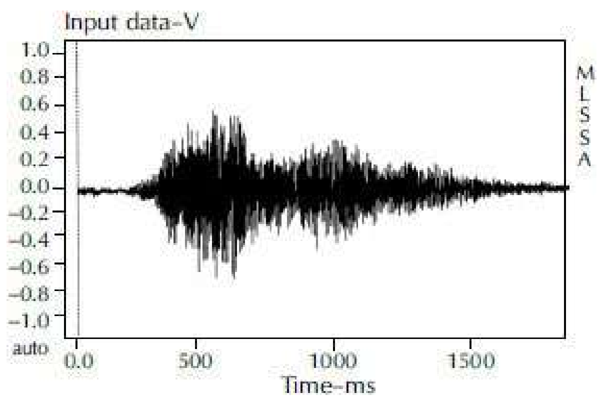


Figure 40-2. Speech waveforms (as Fig. 40-1) but with reverberation (RT = 2.4s). The way one word runs into the next can clearly be seen, but with concentration the individual words can still be understood.

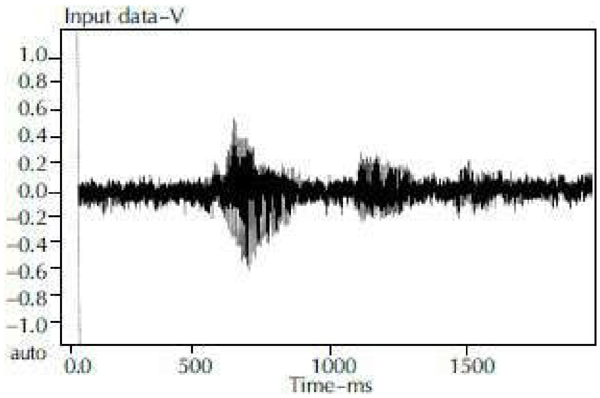


Figure 40-3. Speech waveforms (as Fig. 40-1) in background noise. The noise masks much of the waveform detail but the speech remains intelligible.

40.4 Factors Affecting Sound System Intelligibility

40.4.1 Primary Factors

- Sound system bandwidth and frequency response.
- Loudness and signal-to-noise ratio.
- Room reverberation time (RT).
- Volume and size and shape of the space.
- Distance from the listener to a loudspeaker.
- Directivity of the loudspeaker.
- The number of loudspeakers operating within the space.
- The direct to reverberant ratio** (this is directly dependent upon the previous 5 factors).
  - * Talker annunciation/rate of delivery.
  - * Listener acuity.

** Strictly speaking a more complex characteristic than the simple D/R ratio should be used. Better correlation with perceived intelligibility is obtained by using the ratio of the direct sound and early reflected energy to late reflected sound energy and reverberation. This may be termed C50 or C35 depending upon the split time used to delineate between the useful and deleterious sound arrivals.

40.4.2 Secondary Factors

- System distortion (e.g., harmonic or intermodulation).
- System equalization.
- Uniformity of coverage.
- Presence of very early reflections (<1–2 ms).
- Sound focusing or presence of late or isolated higher-level reflections (>70ms).
- Direction of sound arriving at the listener.
- Direction of any interfering noise.
  * Gender of talker.
  * Vocabulary and context of speech information.
  * Talker microphone technique.

The bulleted parameters marked with a bullet (*) are building or system related, while those marked with an asterisk (*) relate to human factors outside the direct control of the system itself.

How each of the above factors affects the potential intelligibility of a sound system is discussed below together with ways that a system designer can minimize the deleterious effects and optimize the desirable characteristics.

40.5 System Frequency Response and Bandwidth

Speech covers the frequency range from approximately 100Hz–8kHz, although there are also higher harmonics affecting the overall sound quality and timbre extending up to 12kHz and above. Fig. 40-4 shows an averaged speech spectrum with the relative frequency contributions in octave bands. Maximum speech energy occurs over the approximate range 200–600Hz—i.e., in the 250Hz and 500Hz octave bands, and falls off rapidly at about 6dB per octave at higher frequencies as can be seen in Fig. 40-4.

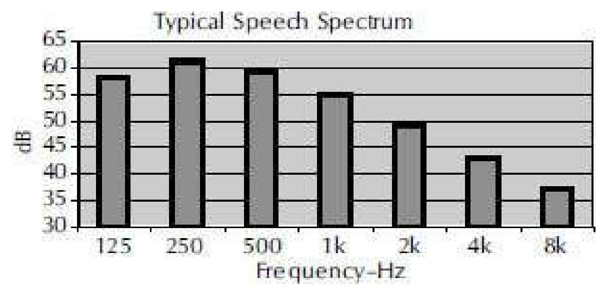


Figure 40-4. Average speech spectrum (octave band resolution).

The lower frequencies correspond to the vowel sounds whereas the weaker upper frequencies correspond to the consonants. The contributions to speech intelligibility, however, do not follow the same pattern—indeed quite the reverse. Fig. 40-5 shows the relative octave band percentage contributions to intelligibility. Here we can clearly see that most intelligibility is concentrated in the 2kHz and 4kHz bands, these contributing approximately 30% and 25%, respectively, while the 1kHz octave contributes a further 20%. These three bands therefore provide over 75% of the available spectral intelligibility content. (It should be noted that these contributions are not absolute but depend on how the intelligibility experiment is conducted and the speech materials employed).

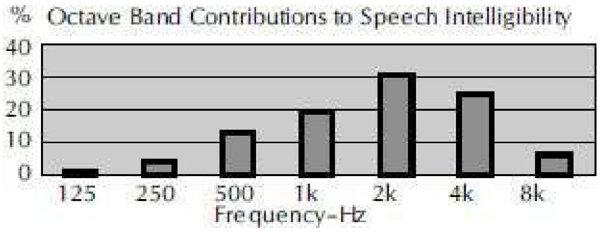


Figure 40-5. Octave band percentage contributions to speech intelligibility.

Whereas the range 300–3000Hz has been shown to be adequate for telephone intelligibility, a wider range is generally required for sound system use—particularly under more difficult acoustic conditions. This effect is shown in Fig. 40-6. This contrasts the results of telephone (monophonic listening) with some recent research carried out by the author in a reverberant space (RT = 1.5 s). The upper curve after Fletcher (1929) shows that the contribution to intelligibility hardly increases beyond 4kHz, while the lower curve, made on a system in a real space (binaurally) shows improvements occurring up to 10kHz. The need for an extended bandwidth can therefore immediately be seen. Limited bandwidth should not be a problem with modern sound system equipment and loudspeakers. However, there are some notable exceptions. These include:

1. Inexpensive poor-quality microphones.
2. Some re-entrant horn loudspeakers (or CD horn drivers used without equalization).
3. Some inexpensive digital message stores.
4. Miniature, special purpose loudspeakers.

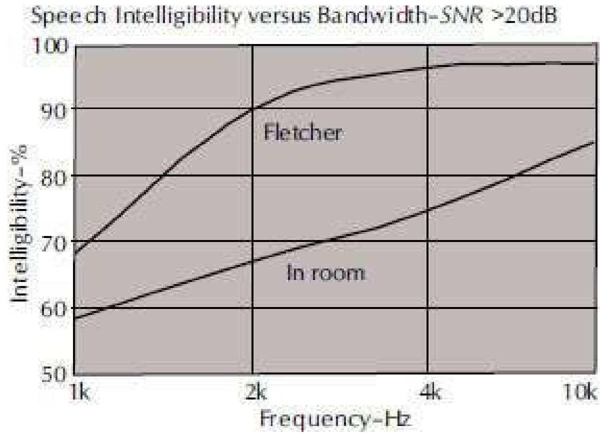


Figure 40-6. Effect of frequency bandwidth on speech intelligibility. Upper curve—monophonic listening (after Fletcher). Lower curve—binaural listening (after Mapp).

Many potentially adequate sound systems are often let down by employing a cheap or restricted bandwidth microphone at the front end of the system. In the author's experience, even on a basic paging system employing restricted bandwidth loudspeakers—e.g., re-entrant horns—the difference between a microphone with a reasonably wide and well-controlled frequency response can always be readily identified over one with a restricted response, even if it exceeds the response of the loudspeakers themselves. Rubbish in equals rubbish out is certainly the case here. However, when operating under high-background noise conditions, a compromise may need to be reached between optimal frequency response and optimal noise rejection, as the two parameters are often divergent.

Apart from component equalization (or the lack of it) by far the most common problems associated with system frequency response stem from either loudspeaker/boundary-room effects or interactions between closely spaced (multiple) loudspeakers. Fig. 40-7 shows the effect of positioning a high-quality monitor loudspeaker with an impeccably flat response close to a boundary wall. As can be seen the response is now far from flat!

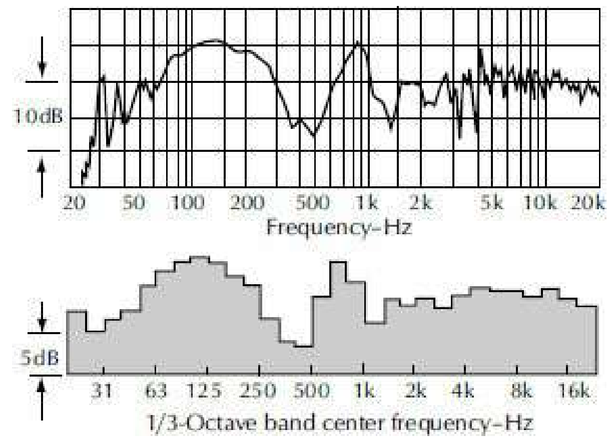
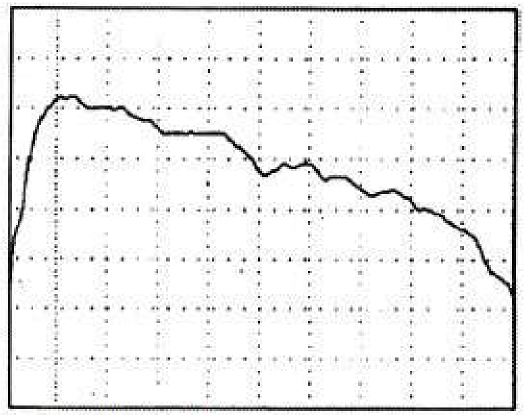


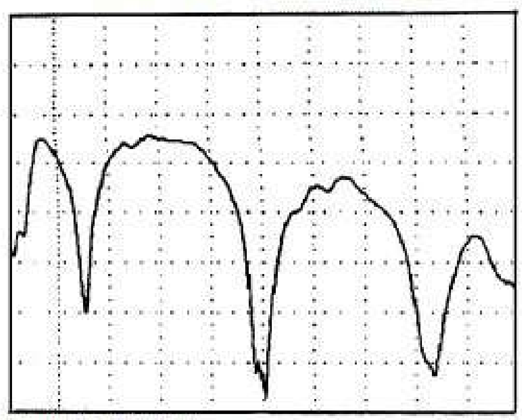
Figure 40-7. Effect of local boundary interaction on loudspeaker frequency response.

Equalization alone cannot correct for this problem. Reduction of the peaks is possible but the notches in the response cannot be equalized out as they are caused by complex phase interactions that cannot be corrected by means of frequency filtering. Interaction between loudspeakers is a common problem in cluster design, where the radiated wave fronts can suffer from missynchronization due to different acoustic path lengths occurring — e.g., due to differences between acoustic centers. Fig. 40-8 shows a typical interaction problem).



Vertical: 6 dB/division
Horizontal: 50.33 Hz – 10,001.20 Hz
Resolution: 5.3674E + 02 Hz
Both horns: Near throw delayed 300μs

A. Loudspeakers in synchronization.



Vertical: 6dB/division
Horizontal: 50.33 Hz – 10,001.20Hz
Resolution: 5.3674E + 01Hz
Both horns: No delay

B. Loudspeakers out of synchronization by 300μs.

Figure 40-8. Frequency response of two loudspeakers. Upper curve shows effect when the sound arrivals are synchronized, lower curve shows effect of 300 ps missynchronization. Courtesy Don and Carolyn Davis.

Here the frequency response of two horn loudspeakers is shown. In the upper curve, the sound arrivals are synchronized and hence add constructively. However, in the lower curve, the horns are missynchronized by just 300μs. A series of sharp comb filters occur. Not only is useful speech information lost in the extensive series of nulls but also the polar radiation pattern is often undesirably affected as shown in Fig. 40-9. The resultant lobes may not only result in certain frequencies not being transmitted to the listeners, but lobes may also be created that can cause undesirable reflections to occur. These may cause either additional unwanted excitation of the reverberant field or cause the generation of late reflections (echoes) that may damage intelligibility.

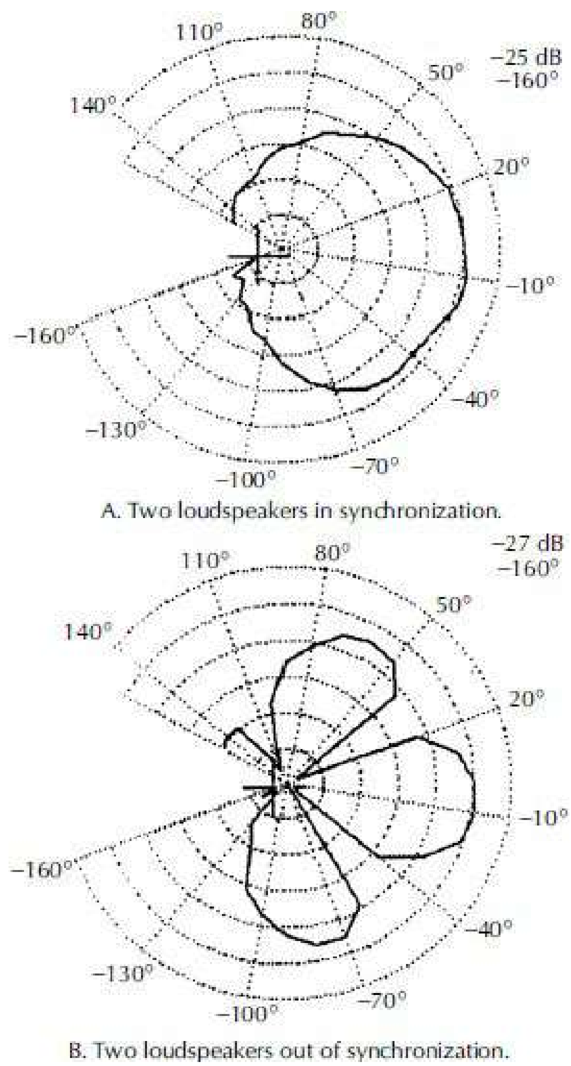


Figure 40-9. Polar response of two loudspeakers in and out of synchronization.

Fig. 40-10 shows the corresponding ETC reflection sequence for the horns in a reverberant space in and out of synchronization. Note the increased excitation of the reverberant field when out of synchronization.

The lobes caused my missynchronization, apart from potentially reducing intelligibility, may also reduce system feedback margin, either by directly radiating sound back to a live microphone or by causing a strong early reflection to occur back into the microphone.

40.6 Loudness and Signal-to-Noise Ratio

The sound level produced by a sound system must be adequate for the intended listeners to be able to hear it comfortably. If the level is too low, many people, particularly the elderly or those suffering even a mild hearing loss, may miss certain words or strain to hear, even under quiet conditions. Although normal face to face conversation may take place at around 60dBA, regularly listeners demand higher sound pressure levels from sound systems, with 70–75dBA being typical for conference systems even when under quiet listening conditions.

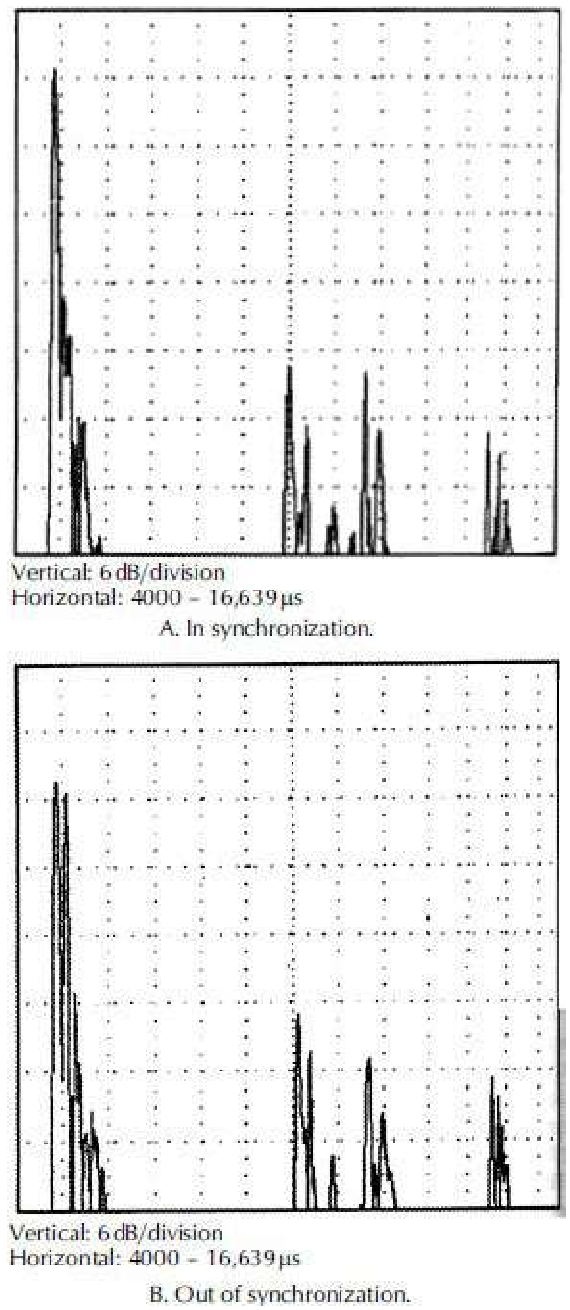


Figure 40-10. ETC curves of the two loudspeakers shown in Fig. 40-9.

In noisy situations, it is essential that a good *SNR* is achieved. Various rules of thumb have been developed over the years. As a general minimum, 6dBA is required and at least 10dBA should be aimed for. Above 15dBA there is still some improvement to be had, but the law of diminishing returns sets in for most practical systems.

There is some disagreement among the generally accepted reference data. Fig. 40-11, for example, shows the general relationship between *SNR* and intelligibility. As the curve shows, this is an essentially linear relationship. In practice, the improvement curve flattens out at high signal-to-noise ratios—though this is highly dependent on the test conditions. This fact is shown in Fig. 40-12, which compares the results of a number of studies, using different test conditions and signals.

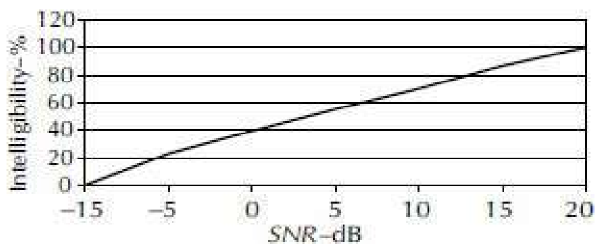


Figure 40-11. Effect of *SNR* on speech intelligibility.

The curve, for example, shows that for more difficult listening tasks, the greater the *SNR* has to be in order to achieve good intelligibility. Fig. 40-13 shows the effect of *SNR* on the %Alcons intelligibility scale. Here, the improvement can be seen to clearly flatten out above 25dB SNR. Under high noise conditions, such a *SNR* could demand excessively high SPLs and caution must be exercised.

Where noise is a particular problem, a full spectral analysis should be performed. Ideally this should be in terms of 1/3-octaves but for many applications 1/1-octave band analysis will be adequate and certainly more informative than a single dBA value. Fig. 40-14 shows such an analysis.

In the upper curve, which depicts a positive *SNR*, it can be seen that the speech signal is greater than the noise over each of the octave band frequencies. However, in the lower curve, it can be seen that at high frequencies the noise exceeds the desired speech signal. The overall effect on potential intelligibility can be calculated by looking at the individual octave band *SNRs* and then weighting and summing them in accordance to their relative contributions as shown earlier in Fig. 40-5.

This is the basis of the Articulation Index (AI), which is a good measure for determining the effects of noise on speech—either over single channel transmission lines such as telephone or radio communications or over PA systems in low but noisy spaces. The AI method is not able to take account of room reverberation or reflections. AI has now been replaced by SII (see section 40.14.2.4) but it is still a useful method for evaluating the effectiveness of speech privacy and sound masking systems).

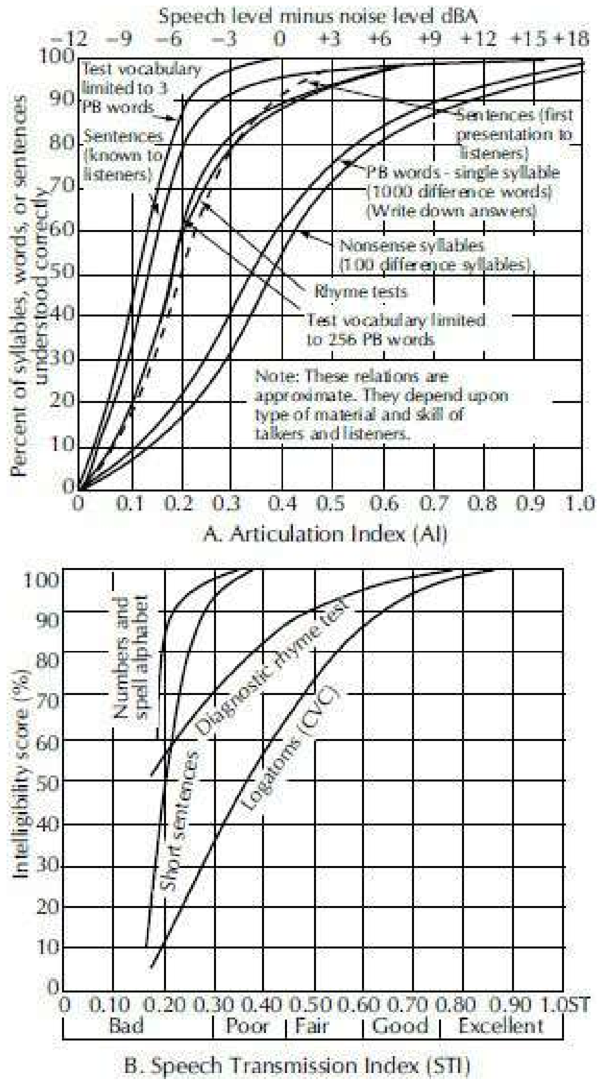


Figure 40-12. Comparison of speech intelligibility test formats as a function of Articulation Index (AI) and Speech Transmission Index (STI).

In many situations, the background noise may not be steady but vary over time. This is particularly the case in many industrial complexes or transportation concourses. Spectator sports also can exhibit highly variable crowd noise levels dependent on the action at any given time. Fig. 40-15 shows a typical noise profile for an underground train station. Peaks of 90dBA plus were recorded as the trains moved in and out of the platforms. A PA system would therefore need to generate at least 96–100dBA in order to achieve an appropriate *SNR* at these times.

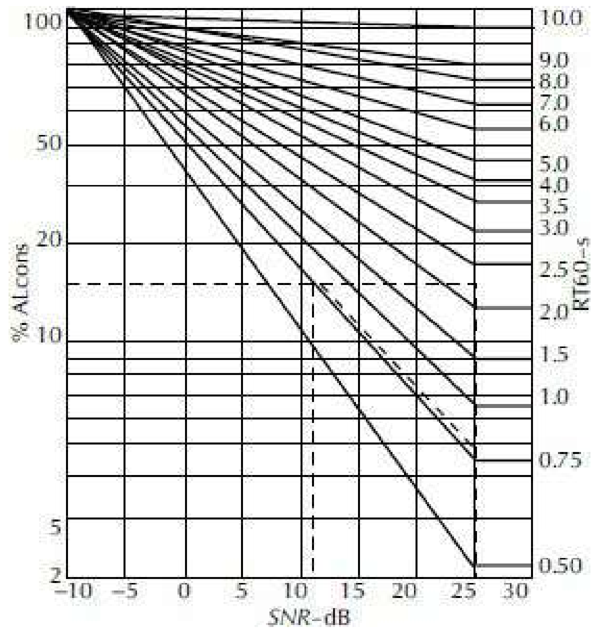


Figure 40-13. Effect of signal to noise ratio on *%Alcons*.

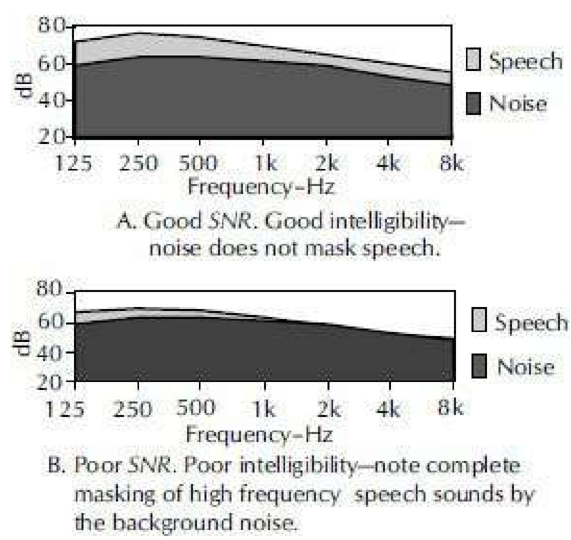


Figure 40-14. Spectral analysis of speech-to-noise ratio (one octave band resolution).

Noise sensing and automatic level control are essential under such conditions, otherwise during the relatively quiet periods when ambient levels drop down to around only 66dBA, significant startle may be caused by such high-level announcements. (A better solution is to store announcements and wait for the regularly occurring quieter periods rather than trying to compete with the background noise all the time.) It should be noted that at sound pressure levels above approximately 80–85dB, speech intelligibility will generally begin to decrease as the sound level increases.

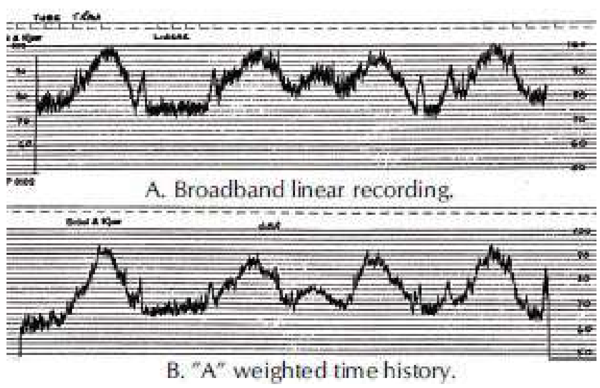


Figure 40-15. Noise-time history profile of underground trains entering and leaving station.

Spectator sports can also create wildly fluctuating noise levels. Again if possible, announcements should be made during the quieter periods, the levels of which can be best determined by a statistical analysis of the crowd behavior at the particular venue in question. Fig. 40-16 shows part of the time history for a soccer match. Note that peak values in excess of 110dBA can occur.

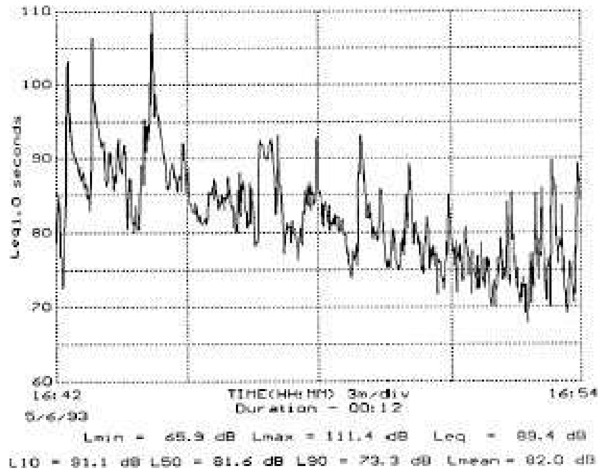


Figure 40-16. Noise-time history analysis of soccer game crowd noise—note short-term variability and peak of 111dB as compared to average of 82 dB.

It must not be forgotten that any noise occurring at the microphone itself will reduce the perceived *SNR*—indeed this is directly additive to the *SNR* at the listener's position. At least 20dBA should be aimed for and preferably >25dBA. A number of techniques can be employed to achieve this, including:

- Close talking/noise canceling microphones.
- Use of highly directional microphones (e.g., gun microphones or adaptive arrays).
- Providing a noise hood or preferably by locating the microphone in a suitable quiet room or enclosure.
- Digital, noise canceling and processing can also be used in extreme conditions to improve the *SNR*.

40.7 Reverberation Time and Direct-to-Reverberant Ratios

Just as noise can mask speech signals so too can excessive reverberation. However, unlike the simpler case of *SNR*, the way in which the direct-to-reverberant (D/R) ratio affects speech intelligibility is not constant but depends on the room reverberation time, the level of the reverberant sound field and on the nature of the speech itself.

The effect is illustrated in Fig. 40-17A–C. The upper trace is the speech waveform of the word *back*. The word starts suddenly with the relatively loud “ba” sound. This is followed some 300ms later by the consonant “ck” sound. Typically the “ck” sound will be 20dB–25dB lower in amplitude than the “ba” sound.

With short reverberation times—e.g., 0.6s—the “ba” sound has time to die away before the start of the “ck” sound. Assuming a 300ms gap, the “ba” will have decayed by around 30dB and will not mask the later “ck.” However, if the reverberation time increases to 1s and if the reverberant level in the room is sufficiently high (i.e., a low *Q* device is used), then the “ba” sound will have only decayed by approximately 18dB and will completely mask the “ck” sound by 8dB to 13dB. It will therefore not be possible to understand the word *back* or distinguish it from similar words such as *bat, bad, bath*, or *bass* since the all important consonant region will be lost. However, when used in the context of a sentence or phrase, the word may well be worked out by the listener from the context. Further increasing the reverberation time (or reverberant level) will further increase the degree of masking.

Not all reverberation, however, should necessarily be considered to be a bad thing, a degree of reverberation is essential to aid speech transmission and to aid the talker by returning some of the sound energy back to him or her. This enables subconscious self-monitoring of their speech signal to occur and so feed back information about the room and projected level. The room reverberation and early reflections will not only increase the perceived loudness of the speech, thereby acting to reduce the vocal effort and potential fatigue for the talker, but also provide a more subjectively acceptable atmosphere for the listeners. (No one would want to live in an anechoic chamber.) However, as we have seen the balance between too much or not enough reverberation is a relatively fine one.

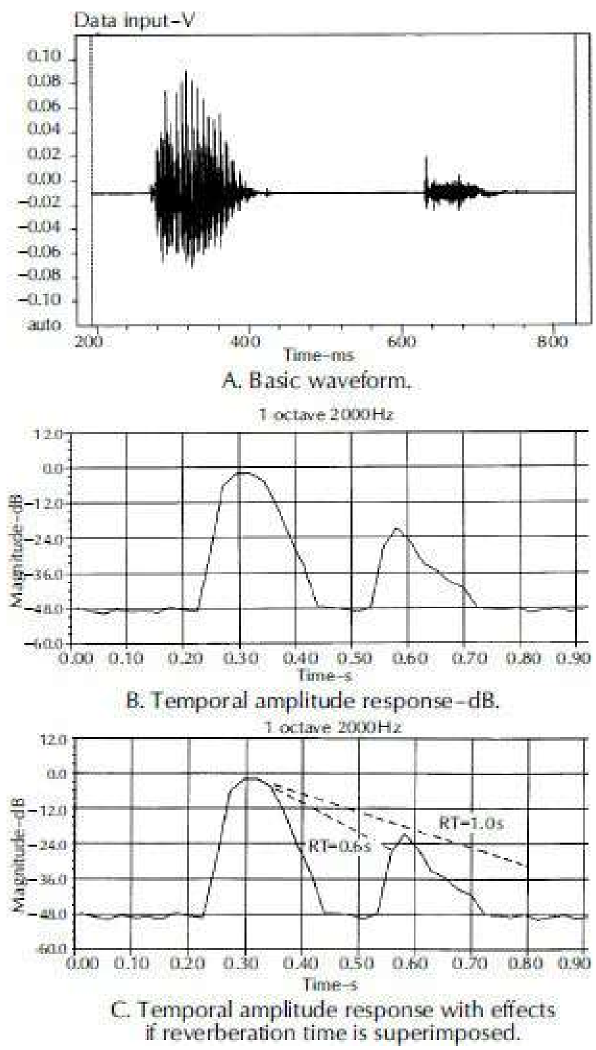


Figure 40-17. Waveform of the word *back*.

The sound field in a large space can be highly complex.

Statistically, it can be divided into two basic components, the direct field and the reverberant field. However, from the point of view of subjective impression and speech intelligibility the sound field needs to be further subdivided to produce four distinct components. These are:

1. Direct Sound—that directly from source to listener.
2. Early Reflections—arriving at the listener approximately 35–50ms.
3. Late Reflections—arriving at the listener approximately 50–100ms later (though discrete reflections can also be later than this).
4. Reverberation—high density of reflections arriving after approximately 100ms.

Fig. 40-18 summarizes the sound field components discussed above.

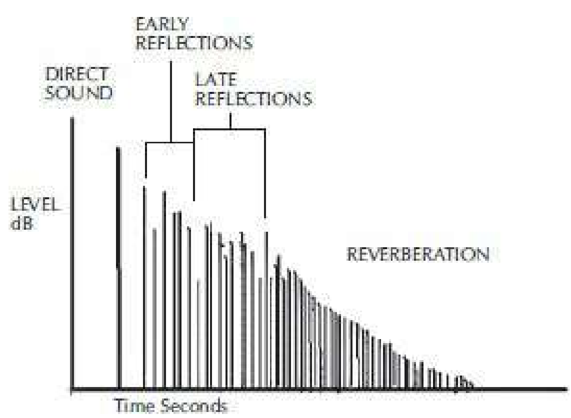


Figure 40-18. Sound field components.

To the above list one could also add “Early Early” reflections—those occurring within 1–5ms. (If specular in nature, these generally cause comb filtering and sound coloration to occur.

Reflections of 1–2ms are particularly troublesome as they can cause deep notches in the frequency response to occur around 2kHz and thereby reduce intelligibility by attenuating the primary speech intelligibility frequency region.)

Opinion as to how the direct sound and early reflections integrate is currently somewhat divided. Many believe that reflections occurring up to around 35–50ms after the direct sound fully integrates with it, provided that they have a similar spectrum. This causes an increase in perceived loudness to occur, which under noisy conditions can increase the effective *SNR* and hence intelligibility. Under quieter listening conditions, however, the case is not quite so clear, with factors including spectral content and direction of reflection becoming increasing important. Equally some research suggests that the integration time may be frequency dependent but generally around 35ms for speech signals. However, there is general agreement that later arriving reflections (>50ms) act such as to degrade intelligibility with increasing effect as the arrival time delay increases.

Sound arriving after approximately 100ms generally signals the start of the reverberant field though strong discrete reflections arriving after 60ms or so will be heard as discrete echoes. It is the ratio of direct + early reflections to late reflections and reverberation that determines the potential intelligibility in a reverberant space (assuming that other effects such as background noise and frequency response considerations are neglected). As a rule, positive ratios are desirable but rarely achieved in reality, though there are exceptions.

This is demonstrated in Figs. 40-19 and 40-20. Fig. 40-19 shows the energy time curve (ETC) sound arrival analysis for a highly directional (high Q) loudspeaker in a large reverberant church (RT = 2.7s at 2kHz). The D/R ratio at the measuring position (approximately 2/3 way back) is 8.7dB resulting in a high degree of intelligibility. Other intelligibility measures taken from the same TEF data (see section 40.13 on measuring intelligibility) are:

- *%Alcons* 4.2%.
- Equivalent RaSTI 0.68.
- C50 9.9dB.

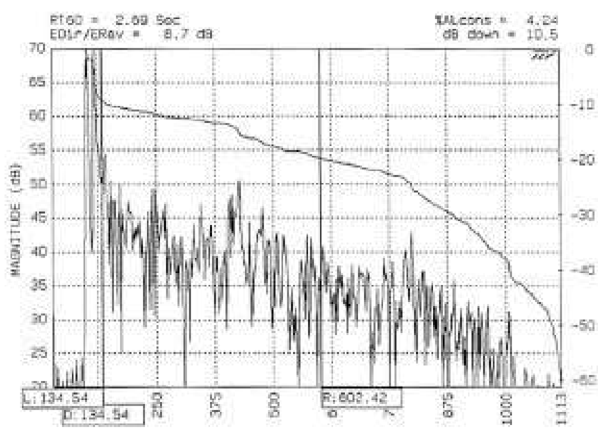


Figure 40-19. ETC of a high *Q* (highly directional) loudspeaker in reverberant church.

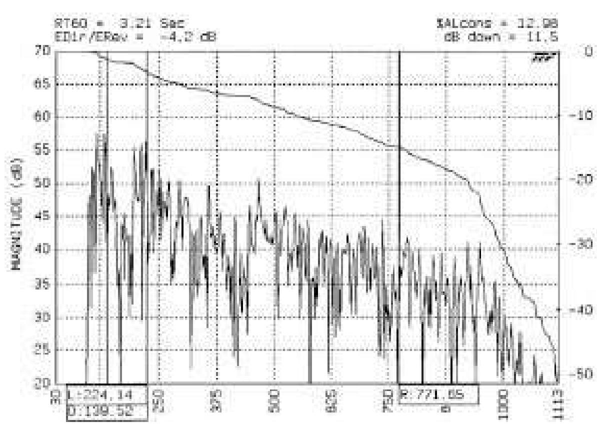


Figure 40-20. ETC of low Q (omnidirectional) loudspeaker in a reverberant church.

An opportunity to exchange the high *Q* device for an almost omnidirectional, low *Q* loudspeaker was taken and found to have a profound effect on the perceived intelligibility and the resulting ETC. This is shown in Fig. 40-20, which presents an obviously very different curve and pattern of sound arrivals. Clearly there is far more excitation of the reflected and reverberant sound fields. The D/R ratio is now −4dB (a degradation of some 12dB) and other computed data is:

- *%Alcons* is now only 13%.
- C50 has been reduced to −3.6dB.
- Equivalent RaSTI to 0.48.

All the indicators and even a visual inspection of the graphs show there to be a significant reduction in the potential intelligibility.

While visual inspection of an ETC can be very enlightening, it can also at times be misleading. Take for example the curve shown in Fig. 40-21. At first glance this resembles the ETC for the low Q device shown above and might suggest low intelligibility since no clear direct sound component is visible. However, densely distributed ceiling loudspeaker systems in a controlled environment do not work in the same way as point source systems in large spaces. In the former, the object is to provide a dense, short path length sound arrival sequence, from multiple nearby sources. The early reflection density will be high and in well-controlled rooms, the later arriving reflections and reverberant field will be attenuated. This results in smooth coverage and high intelligibility. In the case shown in Fig. 40-21, the RT was 1.2s and the resulting C50 was +2.6dB and the RaSTI was 0.68, results both indicating high intelligibility, which indeed was the case.

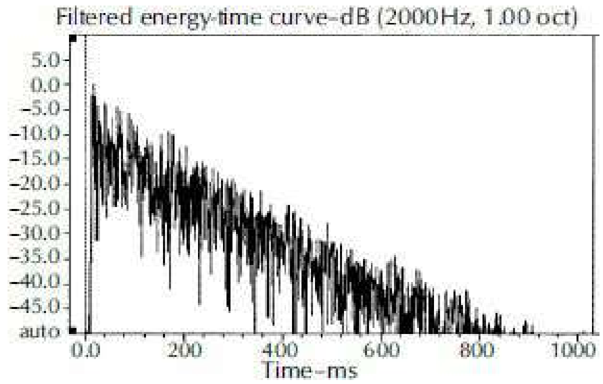


Figure 40-21. ETC of distributed ceiling loudspeaker system in an acoustically well-controlled room.

It should not be forgotten that whereas it may well be possible to produce high intelligibility in a localized area, even in a highly reverberant space, extending the coverage to a greater area will always result in reduced intelligibility at this point, as the number of required sources (additional loudspeakers) to accomplish the task increases. This is primarily due to the resulting increase in acoustic power fed into the reverberant field (i.e., increase in reverberant sound level) often referred to as the loudspeaker system *n* factor.

40.7.1 Intelligibility Prediction—Statistical Methods

While it is relatively trivial to accurately calculate the direct and reverberant sound field components by means of traditional statistical acoustics, it is not possible to accurately estimate, on a statistical basis, the early and late reflection fields. (To do this requires a computer model of the space and ray-tracing/reflection analysis program.)

Prior to such techniques being available, a number of statistically based intelligibility prediction methods based on calculation of the direct and reverberant fields were developed and are still useful in order to provide a quick ball park review of a design or idea. They have greater accuracy when applied to center cluster or point source systems as opposed to distributed loudspeaker systems (particularly high-density distributed systems).

The best known equation is that of Peutz as later modified by Klein and is the articulation loss of consonants equation (*%Alcons*). Peutz related intelligibility to a loss of information. For a loudspeaker-based system in a reverberant room, the following factors are involved:

- Loudspeaker directivity (Q).
- Quantity of loudspeakers operating in the space (n).
- Reverberation time (RT).
- Distance between listener and loudspeaker (D).
- Volume of the space (V).

$$\%Alcons = \frac{200^{*}D^{2}(RT^{2})(n+1)}{QV} \qquad (40\text{-}1)$$

* use 656 for US units

The %Alcons scale is unusual in that the smaller the number, the better the intelligibility. From Eq. 40-1 it can be seen that the intelligibility in a reverberant space is in fact proportional to the *volume* of the space and the *directivity* (Q) of the loudspeaker, (i.e., increasing either of these parameters while maintaining the others constant will improve the intelligibility). From the equation it can also be seen that intelligibility is inversely proportional to the squares of *reverberation time* and *distance* between the listener and the loudspeaker.

The equation was subsequently modified to take account of talker articulation and the effect that an absorbing surface has on the area covered by the loudspeakers.

$$Alcons = \frac{200^* D^2 (T_{60}^{\ 2})(n+1)}{QVma} + K \qquad (40\text{-}2)$$

* use 656 for American units

where,

m is the critical distance modifier, taking into account higher than average absorption of the floor with an audience, for example,

m is (1 – *ay*/(1 – *ac*) where *a* is the average absorption coefficient, *ac* is the absorption in the area covered by the loudspeaker,

k is the listener/talker correction constant typically 1–3, but for poor listeners/talkers can increase to 12.5%.

Peutz found that the limit for successful communication was around 15% Alcons. From 10 to 5% intelligibility is generally rated as good and below 5% the intelligibility can be regarded as excellent. A limiting condition

$$Alcons = 9T + k \qquad (40\text{-}3)$$

was also found to occur by Peutz.

Although not immediately obvious from the equations, they are effectively calculating the direct-to-reverberant ratio. By rearranging the equation, the effect of the direct-to-reverberant ratio on %A*lcons* can be plotted with respect to reverberation time. This is shown in Fig. 40-22. From the figure, the potential intelligibility can be directly read from the graph as a function of D/R and reverberation time. (By reference to Fig. 40-13 the effect of background noise *SNR* can also be incorporated.)

The Peutz equations assume that the octave band centered at

2kHz is the most important in determining intelligibility and uses the values for the direct level, reverberation time, and *Q* to be measured in this band. There is also an assumption that there are no audible echoes and that the room or space supports a statistical sound field being free of other acoustic anomalies such as sound focusing.

In the mid-1980s Peutz redefined the *%Alcons* equations and presented them in terms of direct and reverberant levels and background noise level.

$$\%Alcons = 100(10^{-2(A+BC)-ABC}) + 0.015 \quad (40\text{-}4)$$

where,

$$A = -0.32\log\left[\frac{L_R + L_N}{10L_D + L_R + L_N}\right]$$

$$\text{for } A \geq 1, \text{ let } A = 1$$

$$B = -0.32\log\left[\frac{L_N}{10L_R + L_N}\right]$$

$$\text{for } B \geq 1, \text{ let } B = 1$$

$$C = -0.50\log\left(\frac{RT}{12}\right)$$

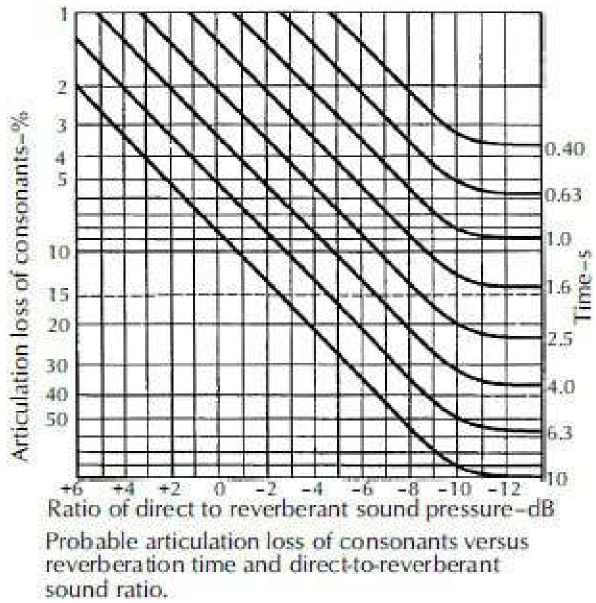


Figure 40-22. Effect of direct-to-reverberant ratio as a function of RT on *%Alcons*.

The *%Alcons* equations work reasonably well with single point or center cluster systems or even split clusters, however, with distributed systems (especially high-density ceiling systems for example) determining the (*n* + 1) factor becomes extremely difficult, as it is difficult to apportion what percentage of the radiation from adjacent or semi-adjacent speakers is actually contributing to the direct field and early fields and which is contributing to the reverberant.

To a certain extent this is made easier in the more complex or long form version as a straight apportionment factor can be applied, though some considerable skill in doing this is required. Because the *%Alcons* equations do not effectively account for the early or late reflected energy, their accuracy needs to be treated with some caution. Furthermore, the method and equations are based on statistical acoustics, which at low reverberation times (e.g., <1.5 s) in itself becomes less accurate. The %Alcons method does not take into account the contributions from other speech frequency bands—as it was derived for natural speech purposes where these relationships are reasonably constant.

40.7.2 Intelligibility and Reverberation Time

Although, as we have seen, there is a lot more to intelligibility than reverberation alone, knowing the reverberation time of a space is a good starting point for a system design and immediately allows the potential difficulty of the task to be quantified. Some general rules of thumb can be applied in this context as seen in Table. 40-1.

Table 40-1. Effect of Reverberation Time

RT	Results
< 1s	Excellent intelligibility should be obtained.
1.0–1.2s	Excellent to good intelligibility should be achieved.
1.2–1.5s	Good intelligibility should be achieved though loudspeaker type and location become important.
>1.5s	Careful design required (loudspeaker selection and spacing).
1.7s	Limit for good intelligibility in large spaces (distributed systems)—e.g., shopping malls, airport terminals.
>1.7s	Directional loudspeaker required (churches, multipurpose auditoriums, and highly reflective spaces).
>2s	Very careful design required. High-quality directional loudspeaker required. Intelligibility may have limitations (Concert halls, churches, treated sports halls/arenas.)
>2.5s	Intelligibility will have limitations. Highly directional loudspeaker required. Large (stone built) churches, sports halls, arenas, atriums, enclosed railway stations, and transportation

	terminals.
>4s	Very large churches, cathedrals, mosques, large and untreated atria, aircraft hangars, untreated enclosed ice sports arenas/stadiums. Highly directional loudspeakers required and located as close to the listener as possible.

When designing or setting up systems for use in reverberant and reflective environments, the main rule to follow is, "Aim the loudspeakers at the listeners and keep as much sound as possible off the walls and ceiling." This automatically partially maximizes the direct-to-reverberant ratio, though in practice it may not be quite so simple. The introduction of active and phased line arrays has had a huge impact on the potential intelligibility that now can be achieved in reverberant and highly reverberant spaces. Arrays of up to 5m (—16ft) are readily available and can produce remarkable intelligibility at distances of over 20-30m even in 10s plus reverberation time environments. The use of music line arrays has also led to a significant improvement in music/vocal clarity in arenas and concert halls. Whereas the intelligibility form a point or low *Q* source effectively reduces as square of the distance, this is not the case for a well-designed/installed line array. An example of this is Fig. 40-23 where it can readily be seen that the intelligibility (as measured using the Speech Transmission Index—STI) remains virtually constant over a distance of 30m in a highly reverberant cathedral (RT = 4s).

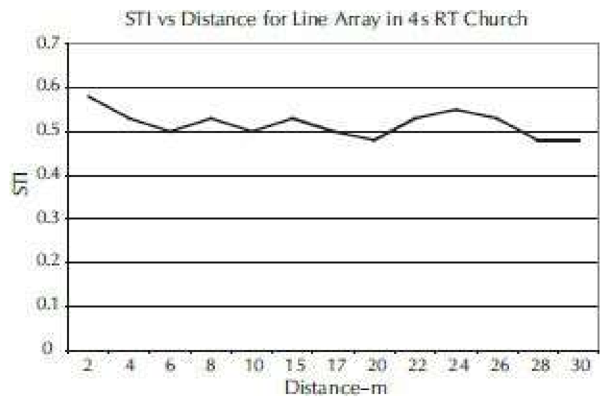


Figure 40-23. Intelligibility versus distance for a four meter line array in a reverberant church.

40.8 Some Further Effects of Echoes and Late Reflections

As already noted, speech signals arriving within 35ms of the direct sound generally integrate with the direct sound and aid intelligibility. In most sound system applications and particularly in distributed loudspeaker systems, a considerable number of early reflections and sound arrivals will occur at a given listening position. These can provide a useful bridging effect (sequential masking) which can extend the useful arrival time to perhaps 50ms. The way in which single or discrete reflections affect intelligibility has been studied by a number of researchers—perhaps the best known being Haas.

Haas found that under certain conditions, delayed sounds (reflections) arriving after an initial direct sound could in fact be louder than the direct sound without affecting the apparent localization of the source. This is often termed the *Haas effect*. Haas also found that later arriving sounds may or may not be perceived as echoes depending on their delay time and relative level. These findings are of significant importance to sound system design and enable, for example, delayed infill loudspeakers to be used to aid intelligibility in many applications ranging from balcony infills in auditoria and pew back systems in churches to large venue rear fill loudspeakers. If the acoustic conditions allow, then improved intelligibility and sound clarity can be achieved without loss of localization.

Fig. 40-24 presents a set of echo disturbance curves produced by Haas and shows the sensitivity to disturbance by echoes or secondary sounds at various levels and delay times.

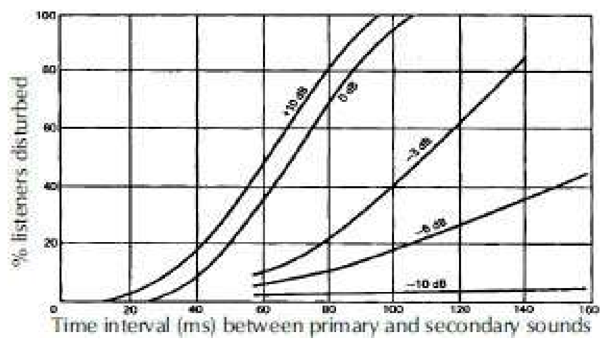


Figure 40-24. Echo disturbance as a function of delay and level (after Haas).

Fig. 40-25 shows a curve of echo perception for various delay times and levels (dotted curve) and indicates that delayed sounds become readily discernible at delays in excess of 35ms (e.g., at 50ms delay), a single reflection or secondary signal has to be more than 10dB lower before it becomes imperceptible and has to be more than 20dB lower at 100ms. The solid curve in Fig. 40-25 shows when a delayed sound will be perceived as a separate sound source and ceases to be integrated with the direct sound.

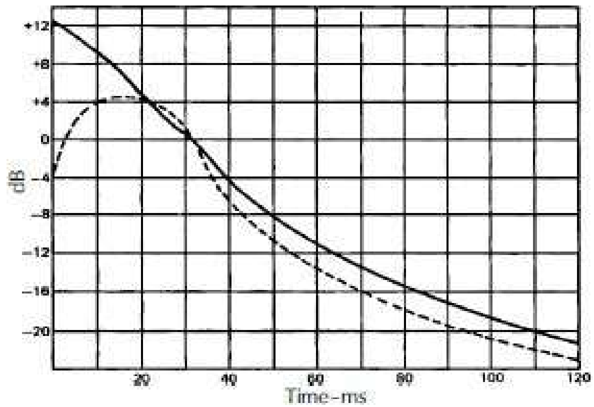


Figure 40-25. Echo perception as a function of delay time and level (after Meyer and Shodder).

Although potentially annoying, echoes may not degrade intelligibility as much as is generally thought. Fig. 40-26, based on work by Peutz, shows the reduction in *%Alcons* caused by discrete sound arrivals or echoes. The curve starts at just under 2% as this was the residual loss due to the particular talker and listener group taking part in the experiment. As the figure shows, the single reflections typically only caused an additional loss of around 2–3%.

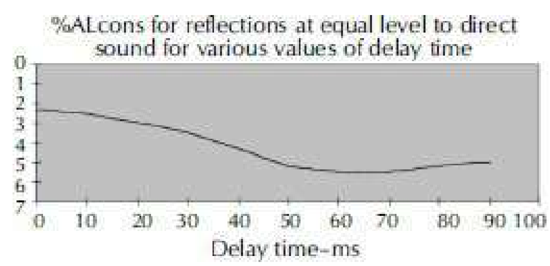


Figure 40-26. Effect of echoes on *%Alcons* (after Peutz).

However, typically more complex systems operating in reverberant spaces can often give rise to the creation of groups of late reflections which, anecdotally at least, would appear to be rather more detrimental. Fig. 40-27 shows the ETC measured on the stage of a 1000 seat concert hall auditorium. A small group of prominent late reflections is clearly visible.

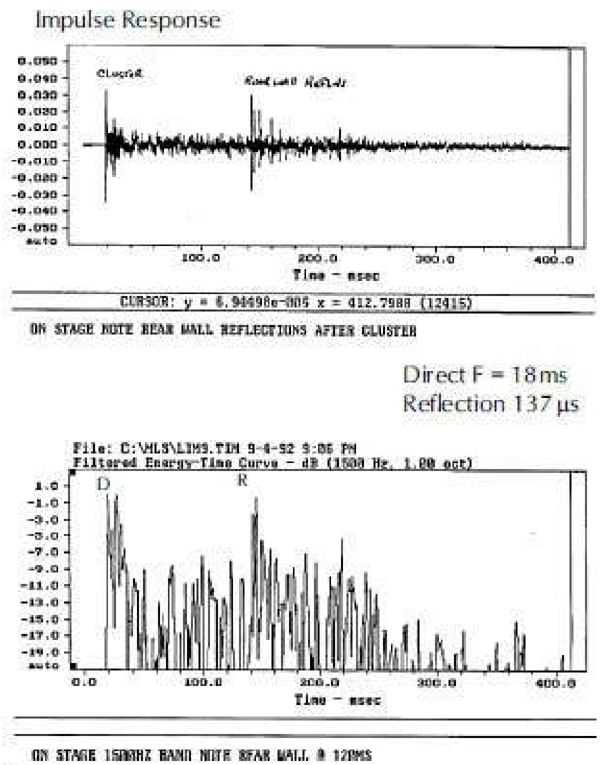


Figure 40-27. Impulse response and ETC of late reflection from auditorium rear wall to stage.

The reflections arrive some 120ms after the direct sound and are less than 0.5dB lower and would therefore be expected to be a significant problem. The cause was sound from the center cluster rebounding off the acoustically untreated rear wall and returning to the stage. This was not only clearly audible but also extremely annoying to anyone using the system when speaking from the stage—the coverage and intelligibility throughout the audience area, however, were extremely good. The problem, although clearly caused by the sound system, was in fact, not the fault of the system, but rather the lack of appropriate acoustic treatment on the rear wall. Sound from the cluster had to strike the wall in order to cover the rear rows of seating. Although this was realized at the design stage and appropriate treatment arranged, in the event this was not installed and an extremely annoying echo resulted. (Later installation of the specified treatment solved the problem, which shows how important it is to properly integrate systems and acoustics.)

Another interesting problem found in the same auditorium during initial setting up of the system is shown in Fig. 40-28. Again a group of late reflections is clearly visible. A strong reflection occurred 42ms after the direct sound just 1.9dB down and the later group arrived 191ms after the direct and 4.5dB down.

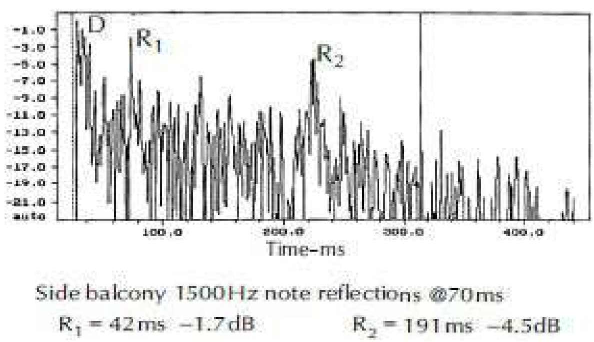


Figure 40-28. ETC showing late reflections in auditorium causing blurring of sound and loss of intelligibility.

Perhaps surprisingly, the effect of these reflections was not to create a distinct echo but rather to cause a general loss of intelligibility and blurring of the sound. In other nearby seats, the intelligibility was good and measured 0.70 STI but in the seats where the intelligibility was poor the STI was 0.53. Although significantly lower than 0.7, a value of 0.53 would still appear to be too high in relation to the subjective impression obtained. However, Houtgast and Steeneken specifically warn against the use of STI for assessing situations with obvious echoes or strong reflections. Identifying the problem however, would not have been possible without the ability to see the ETC.

40.9 Uniformity of Coverage

It is essential when designing systems to work in noisy and/or reverberant spaces to insure that the direct sound level is as uniform as practical. For example, while a 6dB variation (±3dB) may be acceptable under good acoustic conditions, such a variation in a reverberant space can lead to intelligibility variations of 20–40%. A 40% degradation of clarity under such conditions is usually unacceptable. For the case of noise alone, the variation would be at least a 20% reduction in potential intelligibility—though this will be dependent upon the spectrum of the noise. The off-axis performance of a selected loudspeaker is therefore of critical importance—a smooth and well-controlled response being a highly desirable feature.

Where listeners are free to move about—e.g., on a concourse or in a shopping mall—it may be possible to have a greater variation in coverage and hence intelligibility. However, with a seated audience or spectators in an enclosed space, it is essential to minimize seat to seat variations. In critical applications, variations in coverage may need to be held within 3dB in the 2kHz and 4kHz octave bands. This is a stringent and often costly requirement. To put this into perspective consider the following example: assume a given space has an RT of 2.5 s. Calculation shows that on-axis to the loudspeaker at a given distance gives a value of 0.52STI (10%*Alcons*)—an acceptable value. However going off-axis or to a position where the direct sound reduces by just 3 dB will result in a predicted STI of 0.40 (%*Alcons* of 20%)—an unacceptable value, see Fig. 40-22. This shows that it is vital to remember off-axis positions as well as the on-axis ones when carrying intelligibility predictions and system designs. Particularly when it is considered that in many applications, the potential intelligibility will be further degraded by the presence of background noise—even when it is not the primary factor.

40.10 Computer Modeling and Intelligibility Prediction

Computer modeling and the current state of the art are discussed in depth in Chapter 13, *Acoustical Modeling and Auralization* and Chapter 39, *Computer-Aided Sound System Design* and so will only be briefly mentioned here. The ability to accurately predict the direct and reverberant sound fields and compute the complex reflection sequences that occur at any given point are truly remarkable advances in sound system design. As we have seen, calculation of intelligibility from the statistical sound fields alone is not sufficiently accurate for today's needs—particularly with respect to distributed sound systems. The computation of the reflection sequence and hence the impulse response at a point allows far more complex analyses to be carried out including predictions of the early-to-late sound field ratios and the direct calculation of STI. (It should be noted that some of the current simpler programs and many of the earlier prediction programs, although purportedly providing a prediction of STI, in fact base this on a statistical *%Alcons* calculation and convert the resulting value to RaSTI. The accuracy of the result value is therefore highly questionable.)

Some programs, however, are capable of highly accurate prediction, particularly as the precision of the loudspeaker data increases to 1/3-octave bandwidths and 10° or better angular resolution. Also as the computing power continually increases, greater reflection sequence lengths and orders can be more practically accommodated and hence more accurate reflection field data can be calculated. The main restriction currently is not the mathematical accuracy of the model itself, but the time and effort required to build it in the first place. For many schemes this is often not economically viable so some form of simple prediction routine, to at least insure that the proposed system will achieve roughly the right order of magnitude of intelligibility, is still required.

40.11 Equalization

It is surprising how many sound systems are still installed either with no or totally inadequate equalization facilities. Yet the major variations in frequency response (both perceived and measured) that systems exhibit when normally installed can have significant effect on the resultant intelligibility and clarity. Equally many systems after they have been equalized often sound worse than they did before.

This is primarily due to a lack of understanding on behalf of the person carrying out the task. There would appear to have been very little research carried out on the effects of equalization on intelligibility. The author has noted improvements of up to 15–20% on some systems, but otherwise the improvements that can be gained are not well publicized.

There are probably about eight main causes of the frequency response anomalies generally observed prior to equalizing a sound system. Assuming that the loudspeaker(s) has a reasonably flat and well-controlled response to begin with these are:

1. Local boundary interactions, Fig. 40-7.
2. Mutual coupling or interference between loudspeakers.

3. Missynchronization of units in a cluster.
4. Incorrectly acoustically loaded loudspeaker, (e.g., a ceiling loudspeaker in too small a back box and/or a coupled cavity).
5. Irregular (poorly balanced) sound power characteristic interacting with reverberation and reflection characteristics of the space.
6. Inadequate coverage, resulting in dominant reverberant sound off-axis.
7. Excitation of dominant room modes (Eigen tones). (These may not appear as large irregularities in the frequency response but subjectively can be very audible and intrusive.)
8. Comparison for high-frequency losses caused by long cable runs or excess atmospheric absorption.
9. Obscuring the loudspeakers by structures or painting over ceiling loudspeaker grills.

To these may be added abnormal or deficient room acoustics particularly if exhibiting strong reflections or focusing.

Equalization is a thorny subject, with many different views being expressed as to how it should be carried out and what it can and cannot achieve. Suffice it to say that equalization can make a significant improvement to both the intelligibility and clarity of a sound system.

In some cases the improvements are dramatic—particularly when considering not so much the intelligibility per se but associated factors such as ease of listening and fatigue. The essential point is that there is no one universal curve or equalization technique that suits all systems all of the time.

Two examples of this are given below. Fig. 40-29 shows the curves before and after equalization of a distributed loudspeaker system in a highly reverberant church. The anechoic response of the loudspeakers in question is reasonably flat and well extended at high frequencies. Because the measurement (listening) position is beyond the critical distance, the reverberant field dominates and it is the total acoustic power radiated into the space that determines the overall response.

The power response of the loudspeaker in question is not flat but falls off with increasing frequency. (This is the normal trend for cone-based devices but some exhibit sharper roll-offs than others.) This, coupled with the longer reverberation time at lower frequencies due to the heavy stone construction of the building, results in an overemphasis at low and lower midfrequencies. The peak at around 400Hz is due to a combination of power response, mutual coupling of loudspeakers, and boundary interaction effects. The resultant response causes considerable loss of potential intelligibility as high-frequency consonants are lost. Equalizing the system as shown by the solid curve improved the clarity and intelligibility significantly resulting in an improvement of some 15%.

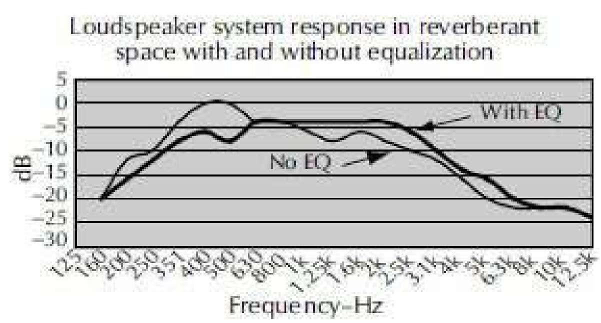


Figure 40-29. Frequency response of a sound system in reverberant church before and after equalization.

Fig. 40-30 shows a widely quoted equalization curve for speech systems. This has been found to work well for distributed systems in reverberant spaces, but it is only a guideline and should not be regarded too rigorously. Loudspeakers that have a better balanced power response that more closely follows the on-axis frequency response will exhibit less high-frequency roll-off and will generally allow a more extended high-frequency equalization curve. It is important to ensure that the Direct Sound from the loudspeakers is relatively flat to begin with, then adjust the resultant in room response with careful equalization that does not overly affect this sound component.

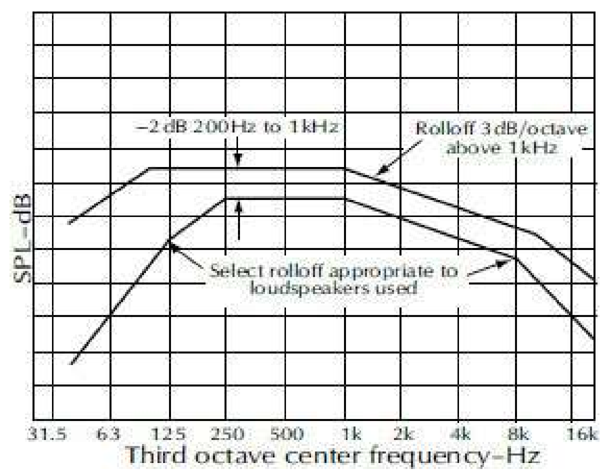


Figure 40-30. Typical response guideline curve for speech reinforcement systems.

An example of this is shown in Fig. 40-31. This is the response of a distributed loudspeaker system employing two-way enclosures in a reflective but well-controlled acoustic environment. In this case, rolling off the high-frequency response would be wholly inappropriate and would degrade the clarity of the system.

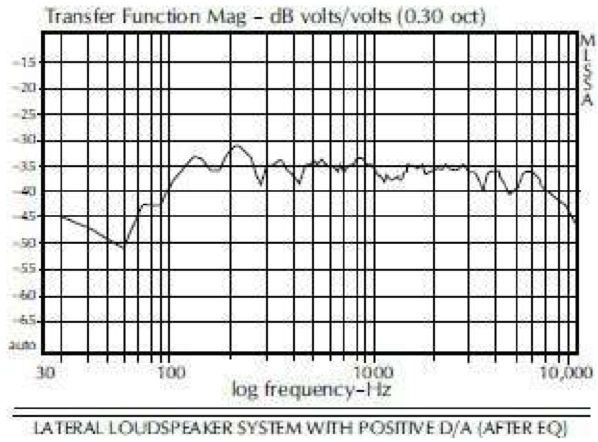


Figure 40-31. Frequency response curve of distributed two-way loudspeaker system in reflective but well-controlled acoustic space.

Adding bass to a sound system may make it sound impressive but will do nothing for the clarity and intelligibility. Indeed, in general, such an approach will actually reduce the intelligibility and clarity particularly in reverberant spaces. Where music as well as speech need to be played through a system, different paths with different equalization settings should be employed so that the different requirements of each signal can be optimized.

40.12 Talker Articulation and Rate of Delivery

Whereas the sound system designer has some control or at least influence over many of the physical parameters that affect the potential intelligibility of a sound system, an area where no such control exists is that of the person using the microphone. Some talkers naturally articulate better than others and so the resultant broadcast announcements are also inherently clearer.

However, it must not be forgotten that even good talkers cause some loss of potential intelligibility. Peutz, for example, found that good talkers produced 2–3% additional *Alcons* loss over and above that caused by the system and local environment. Poor talkers can produce additional losses of up to 12.5%. It is therefore important to design in some element of safety margin into a sound system in order to compensate for such potential losses.

The rate at which a person speaks over a sound system is also an important factor—particularly in reverberant spaces. Considerable improvement in intelligibility can be achieved by making announcements at a slightly slower than normal rate in acoustically difficult environments such as large churches, empty arenas, gymnasiums, or other untreated venues.

Training announcers or users on how to use the system and how to speak into a microphone can make a significant improvement. The need for proper training can not be overstated but is frequently an area that is often ignored. Prerecorded messages loaded into high-quality, wide bandwidth digital stores can overcome certain aspects of the problem.

For highly reverberant spaces, the speech rate needs to be slowed down from the normal rate of speaking—e.g., from around five syllables per second down to about three syllables per second. This can be very difficult to do under normal operating conditions but carefully rehearsed, slower recordings can be very effective. Equally, the author has found that feeding back a slightly delayed or reverberated signal of the person speaking (e.g., via headphones or an earpiece) can be a very effective way of slowing down the rate of speech—though this has to be carefully controlled and set up, as too much delay can become off putting and counterproductive.

Research has shown that intelligibility is improved when the lips of the talker can be seen. At low levels of intelligibility (e.g., 0.3 to 0.4 AI [Articulation Index]) visual contact can produce improvements of up to 50%. Even with reasonably good intelligibility (e.g., 0.7 to 0.8 AI) improvements of up to 10% have been observed. This suggests that paging and emergency voice alarm systems may have a more difficult task than speech reinforcement systems where additional visual cues are generally also present.

40.13 Summary of Intelligibility Optimization Techniques

The following tips should hopefully prove useful in optimizing sound system intelligibility or act as a catalyst for other ideas and design strategies. Although some would appear very basic, it is remarkable how many systems could be improved with just a minor adjustment or simple redesign.

- Aim the loudspeakers at the listeners and keep as much sound off the walls and ceiling—particularly in reverberant spaces or where long path echoes can be created.
- Provide a direct line of sight between the loudspeaker and listener and ensure the acoustic output from the loudspeaker is not obstructed.
- Minimize the distance between the loudspeaker(s) and listener.
- Insure adequate system bandwidth, extending from a minimum of 250Hz to 6kHz and preferably >8–10kHz.
- Avoid frequency response anomalies and correct unavoidable peaks with appropriate equalization.
- Try to avoid mounting loudspeakers in corners.
- Avoid long path delays (>45ms). Use electronic signal delays to overcome such problems where loudspeaker spacing >20ft/6m

(30ft/9m max).

- Use directional loudspeakers in reverberant spaces to optimize potential D/R ratios. (Use models exhibiting smoothly controlled and reasonably flat or a gently sloping power response if possible.)
- Minimize direct field coverage variations. Remember that variations of as little as 3dB can be detrimental in highly reverberant spaces.
- Insure speech *SNR* is at least 6dBA and preferably >10dBA.
- Use automatic noise level sensing and gain adjustment to optimize SNR where background noise is variable.
- Provide a quiet area or refuge for the announcement microphone or use a good quality and effective noise canceling microphone with good frequency response.
- Insure that the microphone user is properly trained and understands the need not to go off mic and to speak clearly and slowly in reverberant environments.
- Repeat important messages.
- In very difficult environments, use simple vocabulary and message formats. Consider use of high-quality specially annunciated prerecorded messages.
- Consider making improvements to the acoustic environment. Do not design the sound system in isolation. Remember, the acoustical environment will impose limitations on the performance of any sound system.

40.14 Intelligibility Criteria and Measurement

A number of intelligibility criteria and rating and assessment methods have already been noted in earlier sections. Here they are treated in a rather more comprehensive overview. However as each technique is quite complex, readers are referred to the bibliography at the end of this chapter to obtain more detailed information.

It is obviously important to be able to specify the desired degree of intelligibility required either for a particular purpose or so that it can be objectively specified for a given project or system. The need then also automatically follows that there has to be a corresponding method of measuring and assessing that a given criterion has been met. Intelligibility measurement and assessment techniques can be divided into two broad categories. These are:

1. Subject based measures—employing a panel of listeners and using a variety of speech-based test materials.
2. Objective acoustic measures of a parameter or parameters that correlate with some aspect of perception.

Subject-based measures include writing down word scores, sentence recognition, modified rhyme tests, and logotom recognition. Objective acoustic measures include broadband and weighted *SNR*, Articulation Index, Speech Interference Level (SIL and PSIL), direct-to-reverberant measures (including TEF *%Alcons* and C35/C50), and STI. There are also a number of subsets of these latter techniques.

It should not be forgotten that it is not just sound reinforcement or public address systems where the resultant intelligibility may require assessment. Other related audio applications include telephone and intercom systems (telephone/headphone or loudspeaker based) as well as teleconferencing systems and other communication channels—e.g., radio. Hearing assistance systems for the hard of hearing can also be assessed and rated using a number of the techniques described below as can the effectiveness of noise masking systems where conversely a reduction in intelligibility is deliberately sought. Measurements may also need to be made in order to assess the natural intelligibility of a space perhaps so that the potential benefits or need for a speech reinforcement system can be evaluated and objectively rated (e.g., churches, classrooms and lecture rooms/auditoria, etc.).

Not all of the techniques are applicable to every application. The area of application is therefore noted at the end of each section. The practical limitations of each of the methods described are also briefly discussed.

40.14.1 Subject-Based Measures and Techniques

The fundamental measurement of intelligibility is of course speech itself. A number of techniques have been developed to rate speech intelligibility. The initial work was carried out in the 1920s and 1930s and was associated with telephone and radio communication systems. From this work the effects of noise, *SNR*, and bandwidth were established and subjective test methods formulated. (Much of this work was carried out at Bell Labs under the direction of Harvey Fletcher.) The sensitivity of the various test methods was also established and it was found that tests involving sentences and simple words were the least sensitive to corruption but often did not provide sufficiently detailed information to enable firm conclusions to be drawn regarding the effects and parameters under study to be definitely made.

The need to insure that all speech sounds were equally included led to the development of phonemically balanced (PB) word lists. Lists with 32, then 250, and finally 1000 words were developed.

Tests using syllables (logatoms) were also developed. These latter tests provide the most sensitive measure of speech information loss but are complex and very time consuming and costly in application.

The modified rhyme test (MRT) was developed as a simpler alternative to PB word lists and is suitable for use in the field with only a short training period. (The more sensitive methods can require several hours of training of the subjects before the actual tests can begin.) The various methods and their interrelationships are shown in Fig. 40-12 where the Articulation Index is used as the common reference.

40.14.2 Objective Measures and Techniques

40.14.2.1 Articulation Index

The Articulation Index (AI) was one of the first criteria and assessment methods developed to use acoustic measurements and relate these to potential intelligibility. AI is concerned with rating the effects of noise on intelligibility and was primarily developed for assessing telephone communication channels. Later corrections were added in attempt to take account of room reverberation but these methods are not considered sufficiently accurate for sound system use. AI is a very accurate and useful method of assessing and rating the effects of noise on speech. ANSI Standard S3.5 1969 (subsequently revised in 1988 and 1997) specifies the methods of calculation based on measurements of the spectrum of the interfering noise and desired speech signal. (Either in terms of 1/1-octave or 1/3-octave bands.)

The Index ranges from 0 to 1 with 0 representing no intelligibility and 1 representing 100% intelligibility. The Index is still very good for assessing the effects of noise on speech in range of applications where room reverberation effects are negligible—e.g., communications channels or aircraft cabins, etc.

Another important application relates to the assessment of speech privacy in offices and commercial environments. Here a very low AI score is required in order to insure that neighboring speech is not intelligible. This is extremely useful when setting up and adjusting sound masking systems and a speech privacy scale has been developed for this purpose. Unfortunately, few commercial analyzers incorporate the measurement, which would be an extremely simple matter to do if a 1/3-octave real-time spectrum display and data are available. Currently, most users of AI in this application either have to compute the result manually or by a simple spreadsheet procedure.

40.14.2.2 Articulation Loss of Consonants

This method was developed by Peutz during the 1970s and further refined during the 1980s. The original equation is simple to use and is in fact based on a calculation of the D/R ratio, although this is not immediately obvious from the equation. The long form of the equation takes into account both noise and reverberation—but unfortunately does not give exactly similar values to the simpler form—which is regarded by many to be overly optimistic. The original work was based on human talkers and not sound systems. (The original prediction equation was modified by Klein in 1971 to its now familiar form in order to do this.)

During 1986, a series of speech intelligibility tests were run that enabled a correlation to be found between MRT word scores carried out under reverberant conditions and a D/R measurement carried out on the TEF analyzer. For the first time this allowed the widely used predictive and design rating technique to be measured in the field. However the correlation does have a number of limitations which need to be considered when applying the method. The measurement bandwidth used at the time was equivalent to approximately 1/3-octave centered at 2kHz. Although three very different venues were employed, each with three significantly different loudspeakers and directivities, the correlation and hence method is only valid for a single source sound system. The measurement requires considerable skill on behalf of the operator in setting up the ETC measurement parameters and divisor cursors, so a range of apparently correct answers can be obtained. Nonetheless the measurement does provide a very useful method of assessment and analysis. In 1989 Mapp and Doany proposed a method for extending the technique to distributed and multiple source sound systems by extending the duration of the measurement window out to around 40 ms.

A major limitation of the method is that it only uses the 2kHz band. For natural speech where there is essentially uniform directivity between different talkers, single band measurements can be acceptably accurate. However, the directivity of patterns of many if not the majority of loudspeakers used in sound systems is far from constant and can vary significantly with frequency—even over relatively narrow frequency ranges. Equally, by only measuring over just one narrow frequency band, no knowledge is obtained regarding the overall response of the system. The accuracy of the measurement correlation can therefore become extremely questionable and any apparent *%Alcons* values extracted must be viewed with caution.

40.14.2.3 Direct-to-Reverberant and Early-to-Late Ratios

Direct-to-reverberant measurements or more accurately direct and early reflected sound energy-to-late reflected and reverberant energy ratios have been used as predictors of potential intelligibility in architectural and auditorium acoustics for many years. A number of split times have been employed as delineators for the direct or direct and early reflected sounds and the late energy. The most common measure is C50, which takes as its ratio the total energy occurring within the first 50ms to the total sound energy of the impulse response. Other measures include C35, whereby the split time is taken as 35ms and also sometimes C7 where this early split time effectively produces an almost pure D/R ratio.

A well-defined scale has not been developed, but it is generally recommended that for good intelligibility (in an auditorium or similar relatively large acoustic space) a positive value of C50 is essential and that a value of around +4dB C50 should be aimed for. (This is equivalent to about 0.60STI or *5%Alcons*.) Measurements are usually made at 1kHz or may be averaged over a range of frequencies. The method does not take account of background noise and is of limited application with respect to sound systems due to the lack of a defined scale and frequency limitations—although there is no reason why the values obtained at different frequencies could not be combined in some form of weighted basis.[1] Bradley has extended the C50 and C35 concept and introduced U50 and U80 etc. where U stands for useful energy. He also included signal-to-noise ratio effects. While the concept is a useful addition to the palette of speech intelligibility measures, it has not caught on to any extent—but it can be a very useful diagnostic tool and further extends our knowledge and understanding of speech intelligibility.

40.14.2.4 Speech Transmission Index STI, RaSTI, and STIPA

The STI technique was also developed in Holland at about the same time as Peutz was developing *%Alcons*. While the *%Alcons* method became popular in the United States, STI became popular and far more widely used in Europe and has been adopted by a number of International and European Standards and codes of practice relating to sound system speech intelligibility performance as well as International Standards relating to aircraft audio performance. (STI is also now cited in the more recent editions of NFPA 72 code for Fire Alarms and Signals.) It is interesting to note that while *%Alcons* was developed primarily as a predictive technique, STI was developed as a measurement method and is not straightforward to predict!

The technique considers the source/room (audio path)/listener as a transmission channel and measures the reduction in modulation depth of a special test signal as it traverses the channel, Figs. 40-32 and 40-33. A unique and very important feature of STI is that it automatically takes account of both reverberation and noise effects when assessing potential intelligibility.

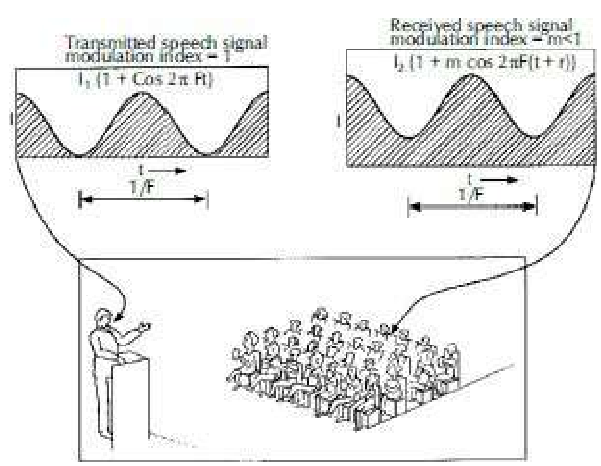


Figure 40-32. Principle of STI and modulation reduction of

speech by room reverberation.

Schroeder later showed that it is also possible to measure the modulation reduction and hence STI via a system's impulse response. Modern signal processing techniques now allow a variety of test signals to be used to obtain the impulse response and hence compute the STI—including speech or music. A number of instruments and software programs are currently available that enable STI to be directly measured. However, care needs to be taken when using some programs to insure that any background or interfering noise is properly accounted for.

The full STI technique is a very elegant analysis method and is based on the amplitude modulations occurring in natural speech, Figs. 40-33 and 40-34. Measurements are made using octave band carrier frequencies of 125Hz to 8kHz, thereby covering the majority of the normal speech frequency range. Fourteen individual low-frequency (speechlike) modulations are measured in each band over the range 0.63 to 12.5Hz.

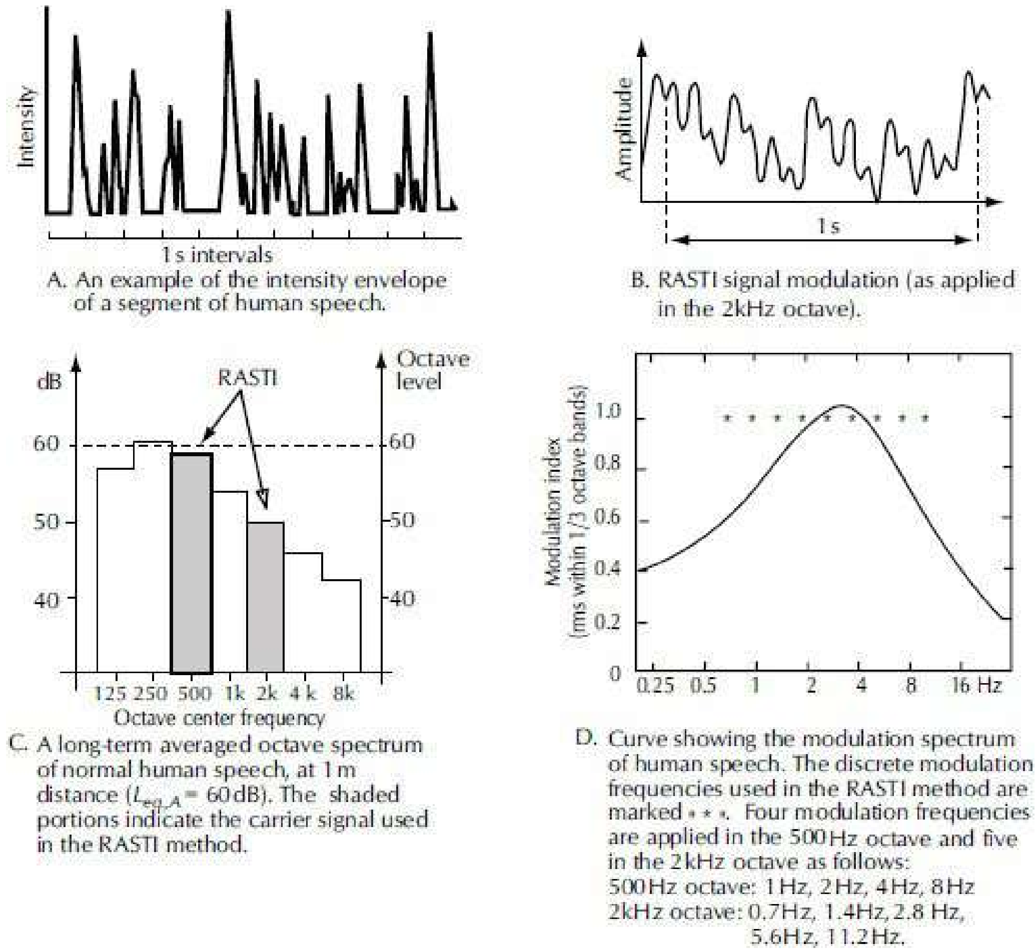


Figure 40-33. Principle of STI and RaSTI showing octave band spectrum and speech modulation frequencies.

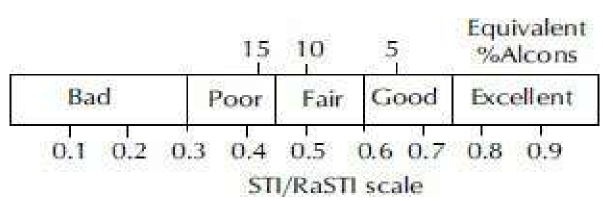


Figure 40-34. STI subjective scale and comparison with %Alcons.

A total of 98 data points are therefore measured for each STI value (7 octave band carriers each × 14 modulation frequencies). Because the STI method operates over almost the entire speech band it is well suited to assessing sound system performance. The complete STI data matrix is shown in Table 40-2. “X” represents a data value to be provided.

When STI was first developed, the processing power to carry out the above calculations was beyond economic processor technology and so a simpler derivative measure was conceived—RaSTI. RaSTI stands for Rapid Speech Transmission Index (later changed to Room Acoustic Speech Transmission Index when its shortfalls for measuring sound system performance were realized.[2,3] RaSTI uses just nine modulation frequencies spread over two octave band carriers thereby producing an order of magnitude reduction in the processing power required.

The octave band carriers are 500Hz and 2kHz, which, although well selected to cover both vowel and consonant ranges, does mean that the system under test has to be reasonably linear and exhibit a well-extended frequency response. Unfortunately many paging and voice alarm systems do not fulfill these criteria and so can give rise to readings of questionable accuracy. (However, this still takes account of a wider frequency range than the traditional D/R and *%Alcons* methods.) Fig. 40-35 shows a system response simulated by the author and evaluated via RaSTI. Although the majority of the speech spectrum is completely missing, the result was an almost perfect score of 0.99 STI! (Note that the most recent edition of IEC 60268-16 [Ed 4 2011] strongly recommends against using RaSTI.)

Table 40-2. STI Modulation Matrix

Carrier/Modulation Frequency (Hz)	125	250	500	1K	2K	4K	8K
0.63	X	X	X	X	X	X	X
0.80	X	X	X	X	X	X	X
1.0	X	X	X	X	X	X	X
1.25	X	X	X	X	X	X	X
1.6	X	X	X	X	X	X	X
2.0	X	X	X	X	X	X	X
2.5	X	X	X	X	X	X	X
3.15	X	X	X	X	X	X	X
4.0	X	X	X	X	X	X	X
5.0	X	X	X	X	X	X	X
6.3	X	X	X	X	X	X	X
8.0	X	X	X	X	X	X	X
10.0	X	X	X	X	X	X	X
12.5	X	X	X	X	X	X	X

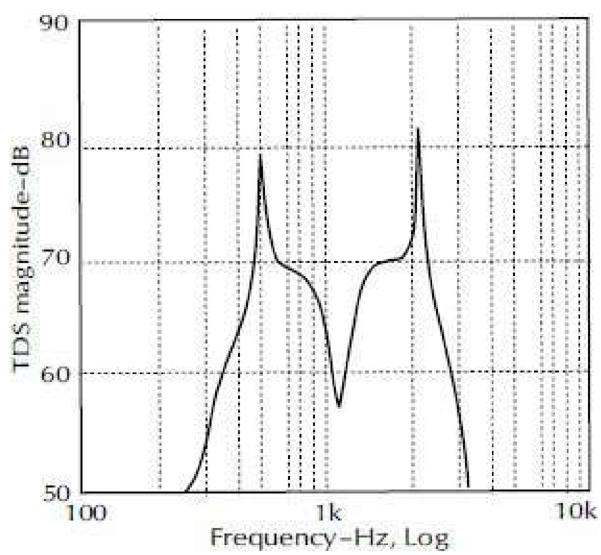


Figure 40-35. Simulated system frequency response curve favoring 500 and 2000Hz, giving an excellent RaSTI value but very poor sound quality.

The first commercially available instrument that could measure STI was the RaSTI meter introduced by Brüel and Kjaer in 1985. Modulated pink noise in the 500Hz and 2kHz octaves is generated and transmitted either acoustically or electronically into the system under test. A useful feature of the method is that the resultant signal is very speechlike in nature having a crest factor of around 12dB, which compares well with normal speech (at around 15–20dB). The relative levels of the 500Hz and 2kHz signals are also automatically transmitted in their correct ratios as compared to natural speech.

This makes setting up the test signal levels comparatively straightforward and enables measurements to be carried out by trained but nonexpert personnel. The introduction and adoption of theRaSTI test method has literally revolutionized the performance of many PA systems ranging from aircraft cabins and flight recording systems to trains, malls, and cathedrals as, for the first time, the intelligibility of such systems could be set and readily verified. As the limitations of RaSTI as a measure of sound system performance became more widely known and understood, it became clear that a replacement method would be required.[2,3] In 2001, STIPA (STI for PA systems) was introduced and became part of IEC 268-16 in 2003. Unlike RaSTI STIPA measures over virtually the complete speech bandwidth range from 125Hz to 8kHz. However, a sparse matrix is used to cut down the complexity of the stimulus and associated measurement processing time. Table 40-3 shows the modulation matrix for STIPA.

Table 40-3. STIPA Modulation Matrix

Carrier/Modulation Frequency (Hz)	125	250	500	1K	2K	4K	8K
0.63			X				
0.80						X	
1.0	X	X					
1.25					X		
1.6	?						
2.0				X			
2.5							X
3.15			X				
4.0						X	
5.0	X	X					
6.25					X		
8.0	?						
10.0				X			
12.5							X

The current version of IEC 60268-16 (Edition 4, 2011) slightly modified the above modulations, providing separate values for the 125Hz band.

Since its introduction, the STIPA technique has rapidly taken off with at least four manufacturers offering handheld portable measurement devices (though some are more accurate than others.[4*] In a similar manner to RaSTI, the STIPA signal is speech shaped and so automatically presents the correct signal for assessing a sound system. (STIPA is a protected name and refers to a measurement made with a modulated signal. Although STIPA can be derived from an impulse response, any such measurement must be clearly indicated as being an indirect STIPA). The new edition of the standard should also help insure that the various STIPA meters are fully compatible, so that any 2 meters using the same IEC 60268-16 Ed 4 test signal should give the same result.

The typical time required to carry out a single STIPA measurement is around 12–15s. However, as the test signal is based on a pseudorandom signal, there can be some natural variation between readings. For this reason it is recommended that at least three readings be taken to insure that a reliable measurement result is produced. STIPA correlates very closely with STI and overcomes most of the problems associated with RaSTI. However, just as with STI, it is a far from perfect measure and there are a number of limitations that need to be understood.

STIPA is vulnerable to some forms of digital signal processing and in particular CD player errors. STIPA test signals may be distributed on CDs, and some CD players can introduce significant errors. It is therefore essential to conduct a loop back measurement to insure that a valid signal is being generated. More recently, however, signals have been distributed as .wav files on solid state memory cards and can also can be directly downloaded, which helps overcome the problem.[4] Most hardware implementations include some form of error detection—particularly the detection of the occurrence of impulsive noise during a test. Not all STIPA meters have incorporated the level dependency relationship that exists between speech level and intelligibility—although it is clearly defined within the IEC 60268-16 standard. The same is also true of the masking function that the standard requires.

One of the major shortcomings of STI and STIPA is their inability to correctly assess the effect that equalization can have on a sound system. For example, STI measurements made on the system described earlier and whose frequency response is depicted in Fig. 40-27 were exactly the same pre and post equalization —although the word score intelligibility improved significantly. Adapting STI to correctly account for such occurrences is not a simple or straightforward matter and it will be some time yet before we have a measure that can accurately do this.

The STI/RaSTI scale ranges from 0 to 1. Zero represents total unintelligibility while 1 represents perfect sound transmission. Good correlation exists between the STI scale and subject-based word list tests. As with all current objective electroacoustic measurement techniques, STI does not actually measure the intelligibility of speech, but just certain parameters that correlate strongly with intelligibility. It also assumes that the transmission channel is completely linear. For this reason, an STI measurement can be fooled by certain system nonlinearities or time-variant processing. STI is also liable to corruption by the presence of late discrete arrivals (echoes). These, however, can be readily spotted by examination of the modulation reduction matrix.

The basic equation for the STI modulation reduction factor *m*(*f*) is

$$m(f) = \frac{1}{\sqrt{1 + \left[\frac{2FT}{13.8}\right]^2}} \times \frac{1}{1 + 10^{\frac{-L_{S/N}}{10}}} \qquad (40\text{-}5)$$

Unfortunately, this equation cannot be directly solved, making STI prediction a complex procedure requiring detailed computer modeling and analysis of the sound field. An approximate relationship exists between STI and *%Alcons*. Fig. 40-34 shows the two scales while Table 40-4 gives a numerical set of equivalent values.

The subjective scale adopted for STI (and RaSTI/STIPA) has led to considerable confusion when rating PA and sound systems. (For example a rating of 0.5 STI would normally be rated as good if heard in a highly reverberant or difficult acoustic environment rather than only fair. Also, in practice, there is usually a marked difference in perception between 0.45 and 0.50 STI (and more particularly 0.55 STI)—although they are all rated as fair. In an attempt to overcome the problem and also to add a degree of tolerance to the measure, the author has proposed that a new rating scale be employed for PA/sound systems (Mapp 2007).5 The proposed scale is shown in Fig. 40-36 and is based on a series of designated bands rather than absolute categories. While the bands will remain fixed, their application can vary so that, for example, an emergency voice announcement system may be required to meet category "G" or above whereas a high-quality system for a theater for a concert hall might be required to meet category "D," or an assistive hearing system might be required to meet category "B" or above, etc., Table 40-5. The new scale was adopted by IEC 60268-16 forming part of the fourth edition of the standard (2011).

Whereas STI does incorporate a degree of diagnostic ability—e.g., it can readily be determined if the speech intelligibility is primarily being reduced by noise or reverberation and the presence of late reflections—a visual display of the ETC or impulse response is invaluable when actually determining appropriate remedial measures and identifying the underlying cause or offending reflective surfaces. A combination of techniques is therefore employed by the author when tackling such problems including the use of directional and polar ETC measurements.

Table 40-4. STI and *%Alcons* Numerical Set of Equivalent Values

Quality	STI	%*Alcons*	Quality	STI	%*Alcons*
	0.20	57.7		0.62	6.0
	0.22	51.8		0.64	5.3
	0.24	46.5		0.66	4.8
	0.26	41.7		0.68	4.3
BAD	0.28	37.4		0.70	3.8
	0.30	33.6		0.72	3.4
	0.32	30.1		0.74	3.1
	0.34	27.0	GOOD	0.76	2.8
	0.36	24.2		0.78	2.5
	0.38	21.8		0.80	2.2
	0.40	19.5		0.82	2.0
	0.42	17.5		0.84	1.8
POOR	0.44	15.7		0.86	1.6
	0.46	14.1		0.88	1.4
	0.48	12.7		0.90	1.3
	0.50	11.4		0.92	1.2
	0.52	10.2	EXCELLENT	0.94	1.0
	0.54	9.1		0.96	0.9
	0.56	8.2		0.98	0.8
FAIR	0.58	7.4		1.0	0.0
	0.60	6.6			

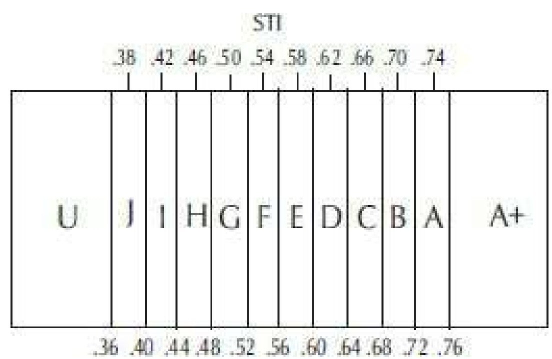


Figure 40-36. New STI scale proposal

Table 40-5. Possible Rating Scheme for Sound Systems

Category	Typical Use	Comment
A		
B	Theaters, speech auditoria, HOH systems	High speech intelligibility
C	Theaters, speech auditoria, teleconferencing	High speech intelligibility
D	Lecture theaters, classrooms, concert halls, modern churches	Good speech intelligibility
E	Concert halls, modern churches	High-quality PA systems
F	Shopping malls, public buildings offices, VA systems	Good quality PA systems
G	Shopping malls, public buildings offices, VA systems	Target requirement for VA/PA
H	VA & PA systems in difficult acoustic environments	Lower target for VA/ PA
I	VA & PA systems in difficult spaces	
J	Not suitable for PA systems	
U	Not suitable for PA systems	

40.14.2.5 SII Speech Intelligibility Index

This relatively new index (ANSI S3.5 *1997*) is closely related to the Articulation Index (AI) but also makes use of some STI concepts. SII calculates the effective signal-to-noise ratio for a number of frequency bands related to the speech communication bandwidth. Several procedures with different frequency domain resolutions are available. These include conventional 1/3-octave and 1/1-octave as well as a twenty one band critical bandwidth (ERB) analysis. An analysis based on seventeen equally contributing bands is also incorporated. The method would appear to be more suitable for direct communication channels rather than for sound reinforcement and speech announcement systems, but in situations where reverberation has little or no effect, the method would be applicable. It should also be useful for evaluating and quantifying the effectiveness of speech masking systems.

40.14.3 The Future for Speech Intelligibility Measurements

As can be seen from the foregoing discussions, we are still a long way from truly measuring speech intelligibility itself. Currently all we can do is to measure a number of physical parameters than correlate under certain conditions to intelligibility. An order of magnitude improvement is required for these to become less anomalous and fallible. The power of the modern PC should allow more perceptually based measurements to be made, as is already happening with telephone networks. However, it must not be forgotten that what is needed in the field is a simple to operate system that does not require highly trained staff to operate, as one thing is for certain: the need to measure and prove that intelligibility criteria have been met is going to rapidly expand over the next few years—indeed the introduction of STIPA is already hastening the process. The range of applications where such testing will need to be performed is also going to rapidly expand and will encompass almost all forms of public transport as well as all forms of voice-based life safety systems. The more traditional testing of churches, auditoria, classrooms, stadiums, and transportation terminals is also set to rapidly expand. As DSP technology continues to grow and as our understanding of psychoacoustics and speech continues to develop, the ability to manipulate speech signals to provide greater intelligibility will increase. Measuring such processes will be a further and interesting challenge.

Particular areas that are likely to see progress over the next few years are the development and use of binaural intelligibility measurements—probably using STI as their basis. The author has also tried using STI and STIPA to assess speech privacy and the effectiveness of speech masking systems. While potentially a promising technique, there are still many obstacles to overcome before it can become a viable technique—not least of which requires considerable research to be carried out between speech intelligibility and STI at the lower end of the STI scale.[5] The measurement and intelligibility assessment of assistive hearing systems is also currently under investigation[6] and is showing considerable promise. It is anticipated that a series of new criteria and measurement techniques will be developed specifically for this specialized but increasingly important field. The use of real speech and other conventional PA signals is also under research and should pave the way for less invasive measurement techniques.

Acknowledgments

The author gratefully acknowledges the assistance of JBL Professional for allowing a number of graphs and figures originally created for their Technical Note "Speech Intelligibility" (TN, Vol. 1, no. 26) to be reproduced in this chapter. The author would also like to thank Don and Carolyn Davis for holding the first Syn-Aud-Con Speech Intelligibility Workshop in 1986, which has acted as an inspiration to further work on this subject ever since.

References

1. J. P. A. Lochner and J. F. Burger, "The Influence of Reflections on Auditorium Acoustics," *Journal of Sound Vibration*, vol. 4, October 1964.

2. P. Mapp, "Limitations of Current Sound System Intelligibility Verification Techniques," *AES 113th Convention*, Los Angeles, 2002.

3. P. Mapp, "Systematic and Common Errors in Sound System STI and Intelligibility Measurements," *AES 117th Convention*, San Francisco, 2004.

4. P. Mapp, "Is STIPA a Robust Measure of Speech Intelligibility Performance," *AES 118th Convention*, Barcelona, 2005.

5. P. Mapp, "How Accurate Are Acoustic Models & Measurements with Regard to Speech Intelligibility?," *Proc. IOA*, vol. 29, Pt.7, 2007.

6. P. Mapp, "Assessing the Potential Intelligibility of Assistive Audio Systems for the Hard of Hearing and Other Users," *AES 124th Convention*, Amsterdam, 2008.

Bibliography

W. Ahnert and F. Steffens, *Sound Reinforcement Engineering*, London, Spon: 1999.

J. Borwick (ed.), *Loudspeaker and Headphone Handbook*, London, Butterworths: 1988, and 3rd Edition 2000.

J. Bradley "Relationships Among Measures of Speech Intelligibility in Rooms," *JAES*, vol. 48, no. 5, 1998.

Don and Carolyn Davis, *Sound System Engineering*, IN, Howard Sams: 1987.

J. Eargle, *Sound System Design Reference Manual*, JBL Professional: 1999.

J. Eargle, *Electroacoustical Reference Data*, Van Nostrand Reinhold, 1994.

B. J. More, *An Introduction to the Psychology of Hearing*, Academic Press, 1988.

B. S. Atal, M. R. Schroeder, and G. M. Sessler, "Subjective Reverberation Time and Its Relation to Sound Decay," *5th International Congress on Acoustics*, Liege, 1965.

M. Barron, "The Subjective Effects of First Reflections in Concert Halls—The Need for Lateral Reflections," *Journal of Sound Vibration*, Vol. 15, no. 4, 1971.

M. F. E. Barron, "The Effects of Early Reflections on Subjective Acoustical Quality in Concert Halls," Thesis, Southampton University, 1974.

F. A. Bilsen and I. Kievits, "The Minimum Integration Time of the Auditory System," *AES preprint #2746*, 1989 Convention.

Jens Blauert, *Spatial Hearing, the Psychoacoustics of Human Sound Localization*, Cambridge, MA: The MIT Press, 1983.

E. Corliss, "The Ear as a Mechanism of Communication," *J. Audio Eng. Soc.*, Vol 38, No. 9, 1990.

Don and Carolyn Davis, “Application of Speech Intelligibility to Sound Reinforcement,” *J. Audio Eng. Soc.*, Vol 37, no. 12, December 1989.

Don and Carolyn Davis, *Sound System Engineering*, 2nd ed. Carmel, IN: SAMS, a Division of Macmillan Computer Publishing, 1986.

Carolyn Davis, “Measurement of %Alcons,” *J. Audio Eng. Soc.*, Vol. 34, no. 11, November, 1986, pp. 905–909.

Don and Carolyn Davis, “Speech Intelligibility Workshop,” *Syn-Aud-Con Tech Topic*, Vol. 14, no. 1, Fall 1986, and vol. 14, no. 8, Summer 1987.

H. Haas, ‘The Influence of a Single Echo on the Audibility of Speech,” *Acustica*, Vol. 1, no. 49, 1951, reprinted *JAES*, 20, 1972.

H. G. Latham, “The Signal-to-Noise Ratio for Speech Intelligibility—An Auditorium Design Index,” *Applied Acoustics*, vol. 12, July 1979.

J. P. A. Lochner and J. F. Burger, “The Intelligibility of Reinforced Speech,” *Acustica* 9, 1959.

W. Klein, “Articulation Loss of Consonants as a Basis for Design and Judgment of Sound Reinforcement Systems,” *J. Audio Eng. Soc.*, Vol. 19, no. 11, pp. 920–922, December 1971.

F. Li and T. Cox, “Extraction of STI from Speech Signals Using Artifical Neutral Networks,” *AES110th Convention*, Amsterdam, 2001.

J. P. A. Lochner and J. F. Burger, (*Acustica*, 8,1,) 1958.

P. Mapp, “From Loudspeaker to Ear—Measurement and Reality,” *AES UK Conference*, 1977.

P. Mapp, “The loudspeaker-Room Interface,” *Hi-fi News & Record Review*, vol. 20, no. 8, 1981.

P. Mapp and P. Doany, “Speech Intelligibility Analysis and Measurement for a Distributed Sound Sysem in a Reverberant Environment,” *AES 87th Convention*, New York, 1989.

P. Mapp, “Investigations into Problems in Auditorium Acoustics,” *Proc. IOA*, vol. 17, pt. 1, 1995.

P. Mapp, “How Loud is Loud Enough?, *Sound & Video Contractor*, Vol. 14, no. 1, January 1996.

P. Mapp, “Practical Limitations of Objective Speech Intelligibility Measurements,” *AES 102nd Convention*, Munich, 1997.

P. Mapp, “Some Effects of Equalisation and Spectral Distortion on Speech Intelligibility,” *Proc. IOA*, Vol. 19, pt. 6, 1997.

P. Mapp, “Objective Speech Intelligibility Testing of Sound Systems,” *Proc. IOA*, Vol. 21, pt. 5, 1999.

P. Mapp, “Reaching the Audience,” *Sound & Video Contractor*, Vol. 17, no. 11, October 1999.

P. Mapp, “Sound Out of Doors,” *Sound & Video Contractor*, Vol. 18, no. 1, January 2000.

P. Mapp, “EQ article,” *Sound & Video Contractor*, Vol. 18, no. 14, Dec. 2000.

P. Mapp, “Sound Power: The Forgotten Loudspeaker and Sound System Parameter,” *AES 110th Convention*, Amsterdam, 2001.

P. Mapp, “Relationships between Speech Intelligibility Measures for Sound Systems,” *AES 112th Convention*, Munich, 2002.

E. Meyer and G. R. Schrodder, Gottinger Nachrichten Math. Phys. Kl., 11a, 1962.

V. M. A. Peutz, “Speech Information and Speech Intelligibility,” *AES 85th Convention*, 1988.

V. M. A. Peutz, “Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room,” *JAES*, Vol. 19, no. 11, 1971.

V. M. A. Peutz, “Speech Intelligibility,” *AES 75th Convention*, Paris, 1984.

H. J. Steeneken and T. Houtgast, “Some Applications of Speech Transmission Index (STI) in Auditoria,” *Acustica*, 51, pp. 229–234, 1982.

H. J. Steeneken and T. Houtgast, “RaSTI: A Tool for Evaluating Auditoria,” *Brüel & Kjaer Technical Review*, no. 3, 1985.

T. Houtgast, H. J. M. Steeneken, and R. Plomp, “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function,” *Acustica*, Vol. 46, September, 1980.

H. Wallach et al., *Am. J. Psychology*, Vol. 62, no. 315, 1949.

B. Yegnanarayana and B. S. Ramakrisina, “Intelligibility of Speech under Non-Exponential Decay Conditions,” *Journal of the Acoustical Society of America*, Vol. 58, 1975.

ANSI Standard S3.5- 1969 / 1988 / 1997 Methods for the Calculation of the Articulation Index.

IEC 289/BS 60286–16 Objective Rating of Speech Intelligibility by Speech Transmission Index, 1998.

IEC 849/BSEN 60849 (1998) Sound Systems for Emergency Purposes.

ISO 7240 Fire Detection and Alarm Systems, 2007

Chapter 41

Virtual Systems

by Ray Rayburn

41.1 The Design of Sound Systems

Sound systems are made of three primary components:

- Input transducers.
- Signal processing.
- Output transducers.

41.1.1 Analog Systems

Transducers are devices that convert energy from one form into another.

The primary type of input transducer used in sound systems is the microphone. It converts the form of acoustic energy we call sound into electrical energy carrying the same information. Other common audio input transducers include the magnetic tape head, the optical sensor, the radio receiver, and the phonograph pickup cartridge. Tape recorders, floppy and hard drives use magnetic heads to transform analog or digital magnetic patterns on the magnetic media into electrical signals. Optical free-space links, optical fiber receivers, and CD and DVD players all use optical sensors to turn optical energy into electrical energy. Radio receivers turn carefully selected portions of radio frequency energy into electrical energy. Phonograph cartridges turn the mechanical motion of the grooves in a record into electrical energy.

Similarly, the most common type of output transducer used in sound systems is the loudspeaker. It converts electrical energy back into the form of acoustical energy we call sound. Other common output transducers include headphones, magnetic tape heads, lasers, radio transmitters, and record cutting heads. Headphones are specialized electrical to acoustic transducers, which are intended to produce sound for one person only. Tape recorders, floppy and hard drives use magnetic heads to transform electrical signals into magnetic patterns on the magnetic media. Optical free-space links, optical fiber transmitters, CDR, CDRW, DVD±RW, and BD recorders all use lasers to turn electrical energy into optical energy. Radio transmitters turn electrical signals into radio frequency energy. Phonograph cutting heads turn electrical energy into the mechanical motion of the grooves in a record.

In general, we can't just connect a microphone to a loudspeaker and have a usable sound system. While there are exceptions such as "sound powered" telephones, in almost all cases there needs to be something that falls under the general heading of signal processing to connect the two.

In its most simplified form this processing might only consist of amplification. In general, microphones have low electrical power output levels, while loudspeakers require more electrical input power in order to produce the desired acoustic output level. Thus the processing required is amplification.

The next most common form of audio signal processing is the level control. This is used to adjust the amount of amplification to match the requirements of the system at this moment.

Multiple inputs to the signal processing are often each equipped with their own level control, and the outputs of the level controls combined. This forms the most basic audio mixer.

Much of the rest of what is done in signal processing can be classified as processing that is intended to compensate for the limitations of the input and output transducers, the environment of the input and output transducers, and/or the humans using and experiencing the sound system. Such processing includes, among other things, equalization, dynamics processing, and signal delay.

Equalization includes shelving, parametric, graphic, and the subcategories of filtering, and crossovers among others. Common filters include high pass, low pass, and all pass. Crossovers are made of filters used to separate the audio into frequency bands.

Dynamics processing is different in that the parameters of the processing vary in ways that are dependent on the current and past signal. Dynamics processors include compressors, limiters, gates, expanders, automatic gain controls (AGC), duckers, and ambient level controlled devices.

Signal delays produce an output some amount of time (usually fixed) after the signal enters the device.

The biggest early breakthrough in sound systems was the development of analog electrical signal processing. No longer was the designer limited to some sort of mechanical or acoustic system. This was taken a large step forward with the development of vacuum tube-based electronic circuitry.

Later, transistor circuitry allowed smaller product sizes and more complex processing to be done. The development of analog integrated circuits accelerated this trend.

Analog signal processing had its limitations, however. Certain types of processing such as signal delays and reverbs were very difficult to produce. Every time the signal was recorded or transmitted, quality was lost. Cascades of circuitry required to meet the ever more complex requirements of sound systems had reduced dynamic range when compared to the individual building blocks making them up.

41.1.2 The Introduction of Digital Devices in Analog Systems

These factors combined together spurred the application of digital signal processing (DSP) to audio.

DSP is the application of numerical processes to perform signal processing functions on signals that have been converted into sequences of numbers. The conversion of an analog signal into a sequence of numbers is done by an analog to digital converter or A/D. Similarly, the conversion of a sequence of numbers back into an analog signal is done by a digital to analog converter, or D/A.

Delays in the analog world almost always involved acoustic, mechanical, or magnetic systems. In other words, you had to use transducers to go from the electrical realm to some alternative form of energy and back, since it was very difficult to delay the signal enough to matter for audio systems while staying strictly in electrical form.

Early digital signal delays had very poor performance compared to today's digital products, but they were popular since the available analog delays were often of even worse audio quality, had very short maximum delay times, and often were not adjustable in delay time.

Digital signal delays offered much longer delay times, easy adjustability of the delay time, and often multiple outputs, Fig. 41-1.

Figure 41-1. Introduced in 1974, the Eventide Clockworks Model 1745M modular digital audio delay was the first to use random access memory (RAM) instead of shift registers for storage of sound. Options included pitch changing and flanging.

Analog reverbs always required some sort of mechanical, magnetic, or acoustic system.

The first analog reverbs were simply isolated rooms built with very reflective surfaces, and equipped with a loudspeaker for inserting the sound, and one or more microphones for picking up the reverberated sound. Obviously, there are major drawbacks to this approach. The size and cost limited the application of this technique, as did the difficulty in adjusting the characteristics of the reverberation.

Another analog technique involved vibrating a large thin steel plate with sound, and picking up the vibrations at multiple locations on the surface of the plate. This had the advantage of smaller size, and adjustable reverberation by moving acoustic damping materials closer to or farther away from the plate.

Smaller yet analog reverbs were made using a gold foil instead of the steel plate, or by using springs that were driven at one end and had vibration pickups at their other end. The gold foil technique resulted in quite acceptable sound, but the spring-based systems were often of low cost and barely usable sound quality.

The first digital reverbs were very expensive, and of a size comparable to that of the gold foil or better spring systems, but had the advantage of much greater control over the reverberation characteristics than could be achieved with the analog systems. As the cost of digital circuitry has come down over the years, so has the price and size of DSP-based reverbs.

Analog recording and transmission of sound have always involved significant reduction in the sound quality as compared to the original sound. Each time the sound was rerecorded or retransmitted, the quality was further reduced.

Digital recording and transmission of sound offered a dramatic difference. While the conversion of analog signals into digital always involves some loss, as long as the signal was kept in digital form and not turned back into an analog signal, making additional copies or transmitting the signal did not impose any additional losses. Therefore, the generational losses associated with the analog systems we had been using were eliminated.

41.2 Digitally Controlled Sound Systems

41.2.1 Digitally Controlled Analog Devices

The physical controls of an audio device have always constituted a significant portion of the cost of the product. This was fine for devices such as mixing consoles where the operator needed instant access to all the controls. There have always been controls, however, that while necessary to the initial setup of the sound system, were best hidden from easy adjustment by the user. These controls were often placed behind security covers to reduce the chance of their adjustment by unauthorized users. Crossovers, system equalizers, and delays are some of the more common examples of this class of device.

Until the introduction of the inexpensive graphically oriented personal computer, there was no practical way to avoid this cost, or to provide other than physical limitation of access to controls. Once graphical computers became commonplace and inexpensive, we saw the introduction of a new class of digitally controlled analog devices. These devices had greatly reduced numbers of physical controls, and in some cases had no physical controls. Instead the devices were adjusted by connecting them to a personal computer and running a control program on the computer. Once the "controls" were set using the program, the computer could be disconnected and the device would retain the settings indefinitely.

Since there were no or a limited subset of actual physical controls, there was no need for physical access restriction devices. Without both a computer and the appropriate control program, there was no way for the user to adjust the controls.

Presets, which could be recalled either from the front panel or remotely, now became possible. Different system configurations, which required multiple changes to controls, now became as simple as pressing a single button.

Computer control also allowed many more controls to be provided on a physically small product than would otherwise be practical. For example, a 1/3 octave equalizer might have 27 bands, plus a number of presets, and yet could be packaged in a small box.

Remote control of the device allows the control point to be distant from the audio device itself. This opened the possibility of reducing the amount of audio cabling in a system and replacing it with inexpensive data cabling to the operator's control point. The data cabling is much more resistant to outside interference than audio cabling.

In the initial versions of such control systems, a different control program and physical connection from the computer to the audio device was required for each device that was to be so controlled. This was fine in smaller systems, but in larger installations where there might be many such digitally controllable devices, it quickly became cumbersome.

To address this limitation, Crown developed what they called the IQ system. Crest developed a similar system they called NexSys. These systems used a single control program together with a control network that connected many digitally controllable devices. Thus they provided a single virtual control surface on a computer screen, which allowed the adjustment and monitoring of multiple individual audio devices.

41.2.2 Digitally Controlled Digital Audio Devices

Early digital audio devices had physical controls that mimicked the controls of analog devices. As with digitally controlled analog devices, the advantages of remote control programs quickly became apparent, particularly for those devices with many controls.

Digital audio devices already were internally digitally controlled, so providing for remote control was an easy and relatively inexpensive step. Some such devices provided physical controls that communicated with the signal processor. Others provided only control programs that would run on a personal computer and connect via a data connection back to the device controlled.

As with the digitally controlled analog devices, most such control schemes required an individual data line from the control computer to each device controlled. Several manufacturers including TOA and BSS developed techniques to allow many of their devices to be controlled by a single data line. These schemes were limited to products of a single manufacturer. Work went on for many years under the auspices of the Audio Engineering Society to try to develop a universally applicable common control scheme, but the requirements were so diverse that a universal standard has yet to be achieved.

This was one of the factors leading to the rise of universal control systems from companies such as Crestron and AMX that have the ability to control and automate remote controllable equipment using any control protocol. For the first time such control systems allow the user to have a single control surface that operates all these systems with their diverse control protocols. These control systems control audio, video, lighting, security, and mechanical systems, allowing a degree of total system integration never before achieved.

Despite the success of these universal control systems, often the user just needs to control all his or her audio system components from a single interface. This has been a driving force behind the continued efforts to develop a common control protocol, or some other easy way to bring all these controls together for the user. Besides the work that the AES has done toward developing such a common protocol, control and monitoring protocols developed for other industries have been adapted for use with audio systems. Among these protocols are Simple Network Management Protocol, or SNMP, and Echelon LonWorks.

The desire for unified systems with reduced control interfaces has also been one reason for the popularity of integrated devices that combine the functions of many formerly discrete devices in a single unified product.

41.2.3 Integration of Digital Products into Analog Systems

41.2.3.1 Dynamic Range

The dynamic range of analog systems is characterized by a noise floor, nominal operating level, and maximum output level before a rated distortion is exceeded. The noise floor is usually constant and does not change with the audio signal. Distortion generally increases with increasing level. The increase in distortion as the maximum output level is approached may be gradual or sudden.

Most professional analog audio equipment today has a maximum output level in the range from +18 to +28dB relative to 0.775V (dBu), and a nominal operating level in the 0 to +4dBu range. While optimum operation requires matching of the maximum output levels so that all devices in the signal chain reach maximum output level at the same time, in practice many engineers do not bother to match maximum output levels, relying instead on the easier matching of nominal levels. The best signal levels to run through typical analog equipment are moderate levels, far away from the noise floor, but not too close to the maximum output level.

Digital equipment, on the other hand, has a different set of characteristics. Distortion decreases with increasing level, and reaches the minimum distortion point just before the maximum output level. At that maximum output level distortion rises very abruptly. The noise floor of digital equipment is often not constant. In some cases the noise is very signal dependent, and sounds to the ear much more like distortion than noise. These characteristics come together to suggest that the optimal signal levels would be those close to but just a little below the maximum level.

41.2.3.2 Level Matching

As we combine analog and digital equipment in the same system, the different characteristics of the two technologies suggest that for maximum performance and widest dynamic range we must align the maximum output levels.

Each device has its own dynamic range, but those ranges will have different characteristics. In all cases we want the audio signal to stay as far as possible from the noise floor.

In any system some device will have the smallest dynamic range, and thus set the ultimate limitation on the performance of the system.

The system as a whole may not perform as well as the worst performing component, unless care has been taken to assure that all devices reach their own maximum output level at the same time.

41.2.3.3 Level Matching Procedure

To match the maximum output levels, apply a midfrequency tone to the input of the first device in the system. Increase the applied level and/or the gain of the device until the maximum output level is reached as determined by the increase in distortion. This point may be determined by using a distortion meter, watching the waveform using an oscilloscope for the onset of clipping, or listening with a piezo tweeter connected to the output of the device.

This latter technique was developed by Pat Brown of Syn-Aud-Con. If a frequency in the range of 400Hz is selected, the piezo tweeter can't reproduce it, and will remain silent. When the device under test exceeds its maximum output level, the resulting distortion will produce harmonics of the 400Hz tone that fall in the range the piezo tweeter can reproduce, and it will sound off in a very noticeable way. The level is then reduced until the tweeter just falls silent, and maximum level has been determined. Rane has produced a commercial tester based on this concept called the Level Buddy.

Once the first device is at its maximum output level, the gain of the second device is adjusted to achieve its maximum output level. In some cases the input of the second device will be overloaded by the maximum output level of the first device and no adjustment of the gain of the second device will eliminate the distortion. In such a case, an attenuator must be used between the devices to drop off the level so the second device's input is not overloaded. One place where such an attenuator is often needed is at the input of a power amplifier. Many times professional power amplifiers have input overload points far lower than the maximum output levels of any of the common devices used to drive them.

Once the second device is at maximum output level, the process is repeated in turn for each subsequent device in the system. Once all device interfaces have been optimized, the system is capable of maximum possible performance.

41.2.3.4 Minimization of Conversions

Up until now, digital components have been treated just like the analog components they have replaced in the system. This is not the optimal way, however, to integrate digital processing into a system.

All devices, analog or digital, impose quality limitations on the performance of the system. In a properly designed digital device, the major performance limitations are due to the conversions from analog into digital and back. Properly done DSP will not introduce significant distortions or other artifacts into the signal. Early analog to digital converters had significantly worse distortions than early digital to analog converters. With modern converters the pendulum has swung back the other way with analog to digital converters often having less distortion than digital to analog converters. In any case, the majority of the distortions and other quality degradations in a properly designed digital based audio signal processor are due to the converters.

This suggests that we should consider carefully how many A/D and D/A converters are used in our systems, with an eye to minimize the number of converters in any given signal path.

This requires a change in how we treat digital audio devices in our systems. No longer can we consider them to be interchangeable with traditional analog components. Instead we must use system design practices that will allow a reduction in the number of converters our audio must go through.

One powerful technique is to group all the digital devices together in only one part of the signal flow of our systems, and use digital interconnects between the devices instead of analog. While not all of our digital processors are available with digital interconnections, many of them are.

The most popular two channel consumer digital interconnect standard is known as SPDIF, while the most popular two channel professional interconnect standard is AES3. The European Broadcasting Union (EBU) adopted the AES3 standard with only one significant change. The EBU required transformer coupling, which was optional under AES3. As a result, this interconnect standard is often called AES/EBU. Many products are made with these interconnects, and converters are available to go from SPDIF to AES3 and from AES3 to SPDIF.

There are many interfaces that carry more than two channels. One that is popular in the home studio market is the ADAT interface, which carries eight channels. Most of the interfaces that originated in the home studio market are limited in the distance they can be run.

To address the need for greater distances and larger numbers of channels in professional applications, CobraNet was developed. It also differs from the other digital interfaces in that it allows point to multipoint connections instead of only point to point. This is due to it running on Ethernet, which is an industry standard computer networking protocol. Today there are several different digital interface systems sold that use some or all of the Ethernet Standard to transmit digital audio.

More information on digital audio interconnects can be found in Chapter 42, *Digital Audio Interfacing and Networking*.

By grouping as many of our digital devices in one portion of the system as possible, and making all the interconnections between them in the digital domain, we have minimized the number of conversions our signal has gone through, and maximized the potential performance.

41.2.3.5 Synchronization

Digital audio consists of a series of consecutive numeric samples of the audio, each of which must be received in sequence. If the samples are not received in proper sequence, or if samples are lost or repeated, then the audio will be distorted.

In order for digital audio devices to interconnect digitally, both ends of each connection must run at the same sampling rate. If the source is running at even a very slightly faster rate than the receiver, sooner or later the source will output a sample that the receiver is not ready to receive yet. This will result in the sample being lost. Similarly, if the source is running at even a very slightly slower rate than the receiver, eventually the receiver will be looking for a sample before the source is ready to send it. This will result in a new false sample being inserted into the data stream.

In a simple chain of interconnected digital audio devices, it is possible for each device to look at the sampling rate of the incoming digital audio, and lock itself to that incoming rate. One problem with this system is that the sampling rate as recovered from the incoming digital audio is less than a perfect steady rate. It will have slight variations in its rate known as jitter. While there are techniques available to reduce this jitter, they add cost, and are never perfect. Each consecutive device in the chain will tend to increase this jitter. As a result, it is not recommended to cascade very many digital audio devices in this manner.

If a single digital audio device such as a mixer will be receiving digital audio from more than one source, then this simple scheme for synchronizing to the incoming digital audio signal breaks down, since there is more than one source. There are two ways to solve this problem.

One way is to use a sample rate converter (SRC) on each input to convert the incoming sample rate to the internal sample rate of the processor. Such a SRC will add cost to the input, and will in some subtle ways degrade the quality of the audio. Of course, there are different degrees of perfection available in SRCs at correspondingly different levels of complexity and cost. Some SRCs will only handle incoming digital audio that is at a precise and simple numeric ratio to the internal sample rate. Others will accept any incoming sample rate over a very wide range and convert it to the internal sampling rate.

This second sort of SRC is very useful when you must accept digital audio from multiple sources that have no common reference, and convert them all to a common internal sampling rate.

As implied above, the other way to handle inputs from multiple digital audio sources is to lock all the devices in the digital audio system to a common reference sample rate. In large systems this is the preferred solution, and the Audio Engineering Society has developed the AES11 Standard, which explains in detail how to properly implement such a system. Such a system can have excellent jitter performance since each device directly receives its sampling rate reference from a common source. Interconnections between the digital audio devices can be rearranged freely since we do not have to be concerned about synchronization and jitter changes as the signal flow is changed.

The only flaw in this scheme, is that some digital audio devices may not have a provision for accepting an external sampling rate reference. As a result, in many complex systems while there may be a master sample rate clock that most equipment is locked to, there often is still a need for sample rate converters to accept the output of those devices which can't lock to the master clock, or that operate at a different sample rate.

41.2.3.6 Multifunction Devices

Once we grouped most or all of our digital devices in a single subsection of our system, the next natural question is why not combine these multiple separate devices into a single product. Obviously, such a combined device greatly reduces or eliminates the need for the system designer to be concerned with synchronization issues, since the equipment designer has taken care of all the internal issues. Only if the system contains more than one digital audio device with digital interconnections does the issue of synchronization arise.

Some of the first examples of such combination products were digital mixers and loudspeaker processors.

Digital mixers were developed that combined not only the traditional mixing and equalization functions, but also often added reverb and dynamics processors inside the same device. Depending on the intended application, such digital mixers might also integrate automation systems, and remote control surfaces. Remote control surfaces allow the separation of the signal processing from the human operated controls. This might allow all the audio inputs and outputs to remain on stage, for example, while only the control surface is placed at the operator's position, Fig. 41-2.

Figure 41-2. The CL3 Digital Mixing Console from Yamaha is part of a family of audio mixers. The size of the mixer has no direct relationship to the number of audio inputs and outputs controlled. The audio input and output connections are in stage boxes that connect to the console via a Dante network. If desired a separate mixer can be used for monitor mixing yet share the same input connections.

Loudspeaker processors are another common example of integrated digital subsystems. Such devices might include input-level adjustment, compression, signal delay, equalization, and crossover functions. Each crossover output might include further signal delay, equalization, level adjustment, and limiting. Often manufacturers provide standard settings for their loudspeakers using these processors, thus optimizing the performance of the loudspeaker to a degree not otherwise possible, while allowing one universal processor to be used for many different products in their line.

The limitation of such products is that their internal configuration is fixed, and therefore the possible applications are limited to those the manufacturer anticipated.

One solution is the one pioneered by Dave Harrison in his analog console designs. In an age when most recording consoles were custom built, and provided just the features and signal flow capabilities requested by the studio owner, David designed a console with so many features, and such flexible signal routing, that it could meet the needs of a very wide range of users. Any one user may not need any but a small subset of the available features. A few users might have requested additional features if they were having a custom console manufactured. Through innovative engineering, David was able to design this console in such a way that it could be mass produced for significantly less cost than the more limited custom consoles it replaced.

Applying this same concept to integrated digital devices led to devices designed with signal processing and routing capabilities well beyond the average user's requirements. This, of course, made such a device capable of application to more situations than a more limited device would have been.

41.2.3.7 Configurable Devices

The next significant advance in integrated digital signal processing was the user configurable device. In such a device, the basic configuration of the signal flow and routing remains constant, or the user can select from one of several different possible configurations. Next, the user can select the specific signal processing that takes place in each of the processing blocks in the selected configuration, within certain constraints.

This sort of device is fine for situations where the basic functions needed are limited, but some degree of customization to suit the job is required. The TOA Dacsys II was an early example of this sort of system, and was available in two in by two out and two in by four out versions, Fig. 41-3.

Figure 41-3. TOA Dacsys II digital audio processors (center and right). TOA's second generation digital audio processor (the SAORI was the first), it had a Windows-based control program, which allowed a limited amount of internal reconfiguration of the signal flow. On the lower left is a digitally controlled analog matrix mixer, which could be controlled with the same program.

For example, a complex processor for a loudspeaker might have multiple inputs optimized for different types of audio inputs. There might be a speech input that is bandlimited to just the speech frequency range, equalized for speech intelligibility, and has moderate compression. There might be a background music input that has a wider frequency range, music-oriented equalization, and heavy compression. There might be a full range music input which has music equalization, and no compression.

The input processing chain for each will have a level control and a high pass filter for subsonic reduction or speech bandwidth reduction. The speech input chain might next have a low pass filter to reduce the high-frequency range. All three inputs will then have multiband parametric equalizers to tailor their frequency response. The speech and background music inputs would then have compressors for dynamic range control. The three input processing chains would end in a mixer that would combine them into a single mixed signal to drive the output processing chains.

Such a system might have three outputs, one for the low frequencies, one for the midfrequencies, and one for the high frequencies. The low-frequency processing chain will have a high-pass filter set to eliminate frequencies below the reproduction range of the woofer. Next, it would have a low-pass filter to set the crossover frequency to the midrange speaker. This may often be followed by a multiband parametric equalizer used to smooth the response of the woofer. Lastly will come a limiter to keep the power amplifier from clipping and/or provide some degree of protection for the woofer. The midfrequency processing chain will have a high-pass filter to set the crossover frequency from the woofer, and a low-pass filter to set the crossover frequency to the high-frequency speaker. It might also have a multiband parametric equalizer and a limiter. The high-frequency processing chain will have a high-pass filter to set the crossover frequency from the midrange, and might have a low-pass filter to set the high-frequency limit. It may have a shelving equalizer to compensate for the high-frequency response of the driver. It will also have a multiband parametric equalizer and a limiter.

Some of these fixed configuration audio processors can allow quite complex systems to be built. For example the Ashley Protea™ ne24.24M provides up to 24 inputs and 24 outputs. Each input has a mic preamp with up to 60dB gain, 48V phantom power and "Push to Talk" switching, input level with polarity, passive potentiometer or active RD-8C remote level control, signal delay, fifteen band fully parametric EQ, noise gate, autoleveler, and ducker. This is followed by a cross point mixer. Each output has high and low pass filters, signal delay, fifteen EQ filters, gain, remote level control, and limiter. The combination of all these facilities allows quite complex systems to be built, Fig. 41-4.

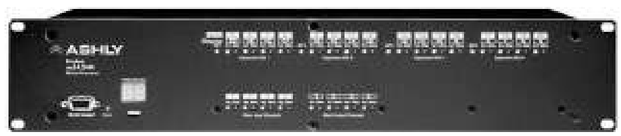
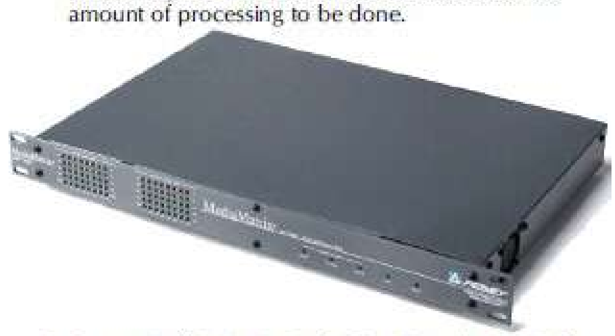


Figure 41-4. Ashley Protea™ ne24.24M matrix DSP processor.

41.3 Virtual Sound Processors

At times, however, even the most complex fixed configuration processor will not meet the needs of a project. In 1993, the Peavey MediaMatrix system introduced the concept of the virtual sound processor. It allowed the designers to choose from a wide variety of virtual audio processing devices, and wire them in any configuration they desired. Integrated digital sound systems could now be designed with the same flexibility of configuration formerly enjoyed in the analog world, and with much greater ease of wiring. Changes to the configuration could be rapidly made on screen and loaded into the processor at the click of a button. Complex systems, which would not have been possible using analog technology due to circuitry drift or cost, now became routine. Systems with as many as 256 inputs and 256 outputs, and 70,000 or more internal controls became practical, Fig. 41-5. Several years later, BSS came out with the Soundweb digital audio processor with similar capabilities but a different physical architecture, Fig. 41-6A. MediaMatrix was based on a PC that supports digital audio processing cards inserted in it. Soundweb consists of a family of digital audio processor boxes that interconnect using category 5 UTP cables, hence the web part of the name. Both products provide similar functions in different ways.

A. Peavey MediaMatrix Frame 980nt. The first Virtual Sound Processor. It is a modular system based on a Windows/Intel PC with added Digital Processing Units (DPUs). From one to eight DPUs may be inserted as needed based on the amount of processing to be done.



B. Peavy MediaMatrix Break Out Box (BoB). Each BoB provides eight audio inputs and eight audio outputs for a MediaMatrix processor.

Figure 41-5. Peavey MediaMatrix mainframe system, the first virtual sound processor.

Biamp, BSS Audio, Electro-Voice, Innovative Electronic Designs, Meyer Sound, Peavey, QSC Audio Products, Symetrix, Yamaha, and others have come up with a variety of products that also give the user the ability to wire virtual devices together. These range from large systems similar to Q-Sys or MediaMatrix, to a small module from QSC that provides processing for a single power amplifier, Fig. 41-6A–E.

Many of these signal processors provide multiple options for audio input and output. For example, Media-Matrix provides options for analog I/O, AES3 I/O, and a CobraNet or Dante interface.

In order for a virtual sound processor to replace all the analog processing used in a sound system, a wide variety of virtual devices must be available. MediaMatrix now provides nearly 700 standard audio processing and control logic virtual devices on its menu. It is also possible to build your own complex devices from simpler devices appearing on the menu or existing inside menu devices. Almost any audio processing device desired may be either found on the menu or built from components available.

Systems are designed in a manner very similar to drawing a schematic on a CAD system. Virtual devices are taken from the menu and placed on a work surface. They have audio input nodes on the left, and audio output nodes on the right side of the device. Some systems have control input nodes on the top, and control output nodes on the bottom of the devices. Wires are drawn interconnecting the I/O nodes and the virtual devices, Fig. 41-7.

A. The Biamp Audia family of virtual signal processors.

B. One of the BSS Audio London series of virtual signal processors.

C. Peavey's MediaMatrix NION N6 virtual signal pro

D. QSC's BASIS series of virtual signal processors.

E. Yamaha's DME Satellite Series virtual signal proc

Figure 41-6. Virtual sound processors by Biamp, BSS, Peavey, QSC, and Yamaha.

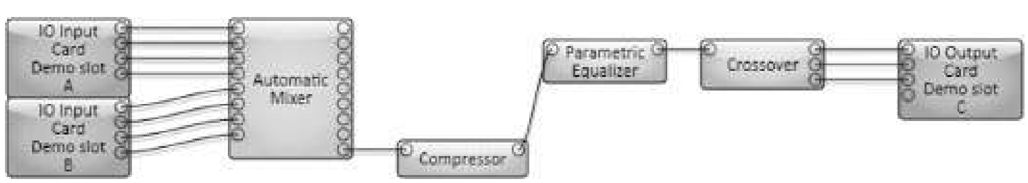
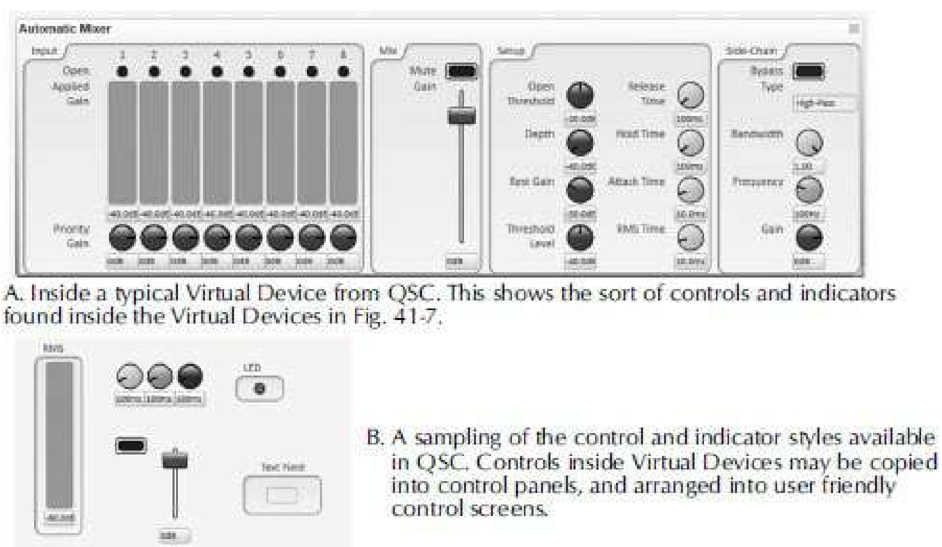


Figure 41-7. A simple example of a virtual sound processor schematic from QSC Q-Sys.

Any number of virtual devices may be used until the available DSP processing power is exhausted. All of the systems provide some means for displaying the amount of DSP used. Devices may be added to the schematic until 100% utilization is reached. Expandable systems such as MediaMatrix and Soundweb allow the addition of more cards or boxes to add additional processing power as needed. MediaMatrix also allows the selection of sampling rate. Slower sampling rates trade off reduced bandwidth for increased processing capability.

Since the schematic may be edited at any time, one major advantage of these systems is that changes may easily be made in the field to accommodate changed requirements or field conditions. Since it is rare that a system is 100% utilized, often the needed additional virtual devices, or wiring changes, may just be added. If the change exceeds the available DSP resources, often some other change may be made in a less critical area to reduce the required DSP resources. By contrast, in an analog system physical rewiring or the purchase of additional components would be required. Both of these add significant cost. Thus often the use of virtual sound processors results in significant savings in the total project cost, over and above the cost savings of the initial equipment purchase, and a more optimized finished system.



A. Inside a typical Virtual Device from QSC. This shows the sort of controls and indicators found inside the Virtual Devices in Fig. 41-7.

B. A sampling of the control and indicator styles available in QSC. Controls inside Virtual Devices may be copied into control panels, and arranged into user friendly control screens.

Figure 41-8. Inside a typical virtual device from QSC Q-Sys. This shows the sort of controls and indicators found inside the virtual devices in Fig. 41-7.

Generally, double-clicking on a virtual device will open it, allowing the internal controls to be seen, Fig. 41-8. Inside each device is a control panel with the controls and indicators needed by that device. Sometimes seldom used controls will be placed in sub-windows.

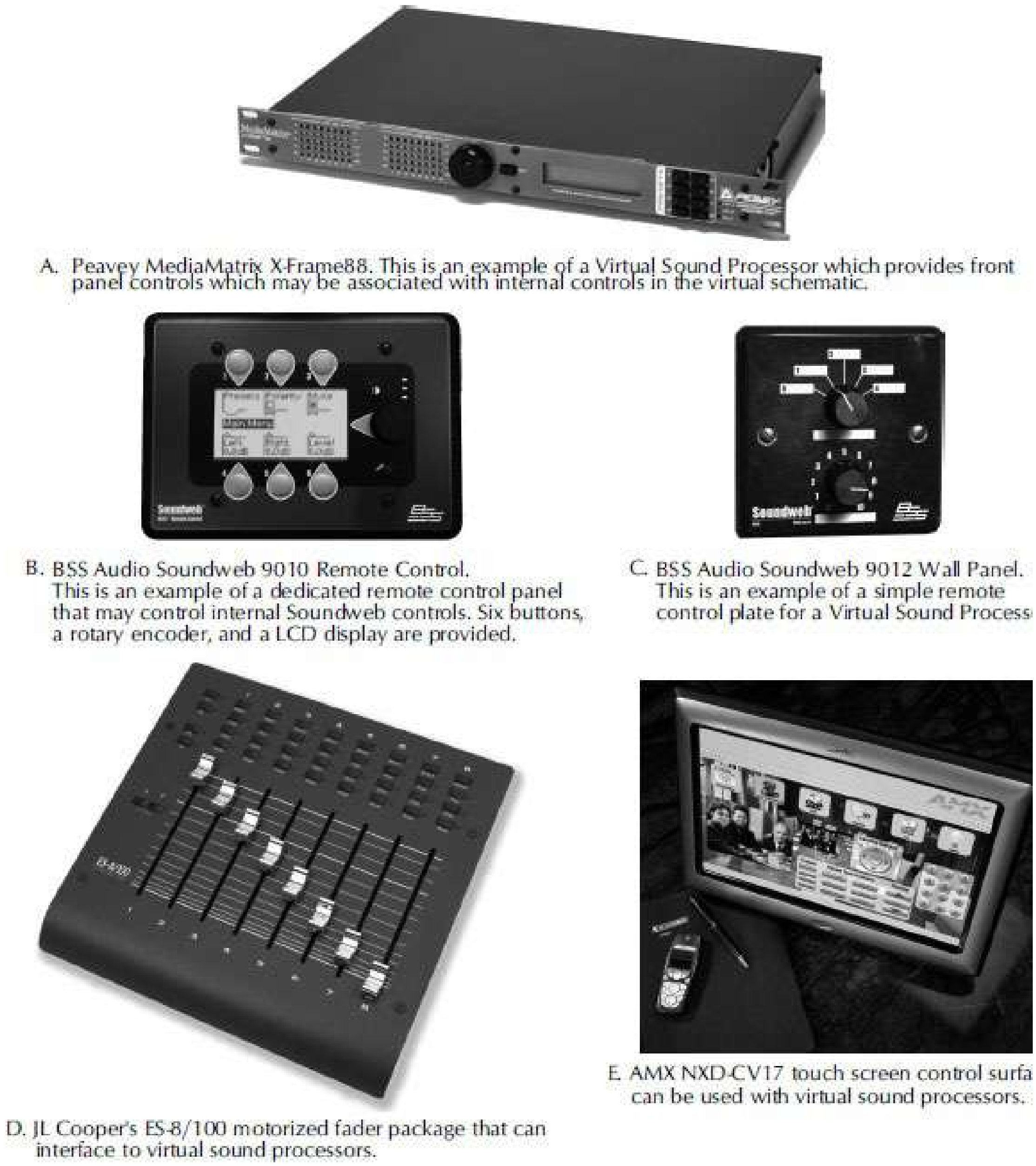
Selected controls from the devices may be copied and placed in control panels. This is done using the standard Windows copy and paste commands. The schematic may then be hidden, and the user only allowed access to the controls that the designer wishes, placed on the control panels. Multiple controls may be ganged, and the settings of many controls may be recalled using presets or subpresets. Presets recall the settings of all the controls in the system, while subpresets recall the settings of just a selected subset of the controls. Controls may also be edited to change their style, size, color, and orientation. This capability allows the designer to develop a very user-friendly interface. Often bitmaps may be inserted to serve as backgrounds.

Besides the virtual interface, some systems require physical interfaces. To support this requirement, most virtual sound processors provide remote control capability in addition to their virtual control surfaces. Some have a few front panel controls available, Fig. 41-9A. Many virtual sound processors provide control inputs to which external switches or level controls may be connected. Control outputs allow lamps and relays to be driven. Serial control interfaces using RS 232, RS 485, or MIDI are often available. Some processors also provide Ethernet interfaces. Others have dedicated programmable remote control panels. When remote control needs are extensive, but the user interface must be simple, touch screen operated control systems such as by AMX or Crestron are often used. These usually control the virtual audio processor by means of serial RS232, RS485, or Ethernet control lines, Fig. 41-9A–E.

Designing and using a virtual sound processor is similar to designing an analog system, except that you have the ability to more precisely optimize the system. The cost of each individual virtual device is very low, and you have the ability to wire precisely the configuration you need. Thus designs may be more efficient, and may also more exactly meet the system requirements.

Let’s take our previous example of a loudspeaker processor with three inputs, each optimized for a different type of program, and three-way outputs, and design a virtual sound processor for this task using QSC Q-Sys. Other virtual sound processors could also be used for the same purpose although some details would be different.

The first step is to place the virtual devices for the audio inputs and outputs. Since we wish analog inputs and outputs, we will select analog I/O cards for inputs and outputs 1 through 4 from the Device menu, Fig. 41-10.



A. Peavey MediaMatrix X-Frame88. This is an example of a Virtual Sound Processor which provides front panel controls which may be associated with internal controls in the virtual schematic.

B. BSS Audio Soundweb 9010 Remote Control. This is an example of a dedicated remote control panel that may control internal Soundweb controls. Six buttons, a rotary encoder, and a LCD display are provided.

C. BSS Audio Soundweb 9012 Wall Panel. This is an example of a simple remote control plate for a Virtual Sound Process

D. JL Cooper's ES-8/100 motorized fader package that can interface to virtual sound processors.

E. AMX NXD-CV17 touch screen control surfa can be used with virtual sound processors.

Figure 41-9. Examples of physical controls for virtual sound processors.

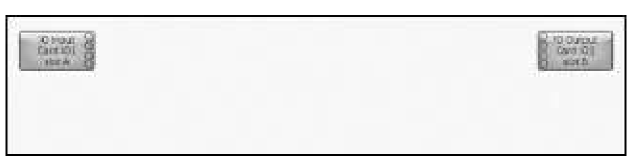


Figure 41-10. QSC software showing virtual devices for audio input and output.

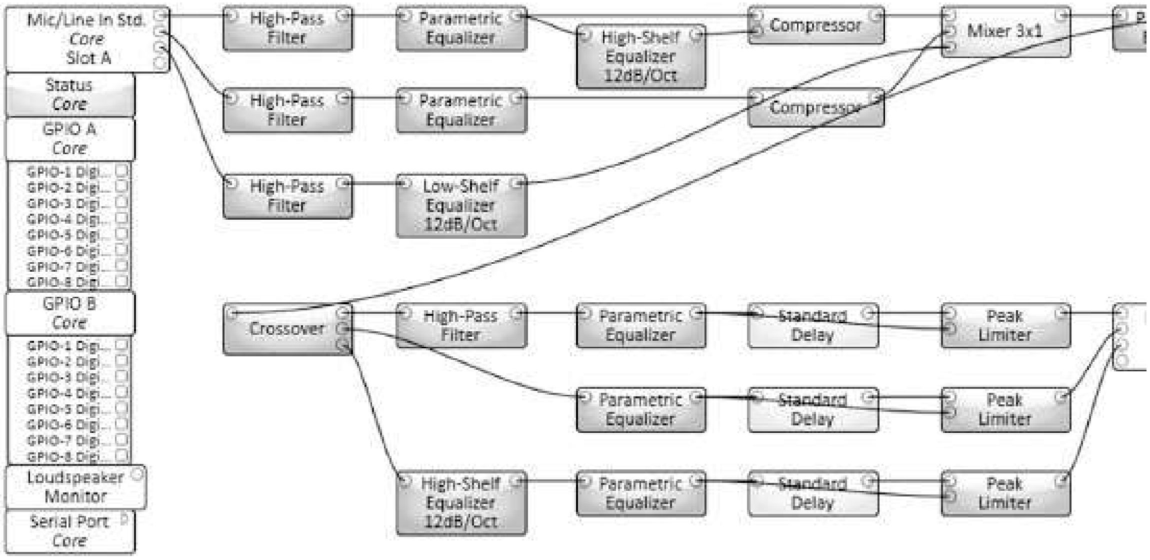


Figure 41-11. Completed crossover example schematic.

Input 1 will be our speech input. We will use a high pass filter, a 1 band parametric equalizer, a high-frequency shelving equalizer, a compressor, and a three input mixer. Input 1 will wire to the input of the high pass filter. The output of the high pass filter feeds the parametric equalizer, which feeds both the main input of the compressor, and the shelving equalizer. The output of the compressor wires to the first input of the mixer. The high pass filter will be adjusted to be a 125Hz Butterworth 24dB/octave filter. This band limits the input to the speech range, and prevents the entry of low-frequency noise. The parametric equalizer will be set for a bandwidth of two octaves, and 3dB boost at 3kHz. This provides a gentle emphasis of the speech intelligibility range. The compressor is left at its default settings of Soft Knee, 0dB threshold, and a ratio of 2:1. The high-frequency shelving equalizer will be set to a frequency of 8kHz, and 8dB of boost. In combination with the compressor, this serves as a de-esser. By boosting the sibilance range at the input to the side chain of the compressor, those frequencies will be compressed more easily, and excessive high-frequency sibilance will be controlled, Fig. 41-11.

Input 2 will be for the background music. We will use a high pass filter, two bands of parametric equalization, and a compressor. The high pass filter will be set to 80Hz with a Q of 2. This produces an underdamped response with a bass boost just above the low-frequency roll-off. One band of the parametric equalizer is set to 1.5kHz with a bandwidth of two octaves, and a cut of 5dB, while the other is set to 8kHz with a bandwidth of one octave, and a boost of 5 dB. The combination of the high pass filter and the parametric equalizer produces the desired background music response. The compressor is set to Soft Knee, −10dB threshold, and a ratio of 4:1. This provides a more aggressive compression. The output of the compressor is wired to the second mixer input.

Input 3 is for full range music. It has a high pass filter, and a low-frequency shelving equalizer. The high pass filter is set to 30Hz at 12dB/octave, and the low-frequency EQ is set to +10dB at 100Hz. The output of the EQ is wired to the third mixer input.

The output of the mixer will drive a 6 band parametric equalizer for overall system EQ. Next comes a three-way 24dB/octave crossover. The low-frequency output of the crossover is wired to a high pass filter with the *Q* adjusted so it optimally tunes and protects the woofer. Next comes a three band parametric EQ, five millisecond delay, and limiter. The side chain input of the limiter is wired directly from the output of the EQ bypassing the delay. This combination of a delay and limiter wired so that the main input of the limiter sees a delayed signal, while there is no delay on the side chain input, is known as a *look-ahead limiter*. By setting the delay to about three times the attack time of the limiter, the limiter has time to react to an audio signal before that signal reaches the limiter. Since the limiter has an attack time of 1ms, we will set the delay to 3ms. A look-ahead limiter is able to accurately limit audio transients without the distortions inherent in ultrafast limiters.

The mid- and high-frequency outputs of the crossover are processed similarly. The midfrequency output just has the parametric EQ, delay, and limiter, while the high-frequency output has a shelving equalizer to compensate for the CD horn, parametric EQ, delay, and limiter, Fig. 41-11.

As you can see, while this circuit is relatively simple, using a virtual audio processor allowed us to optimize it in ways not possible using either commonly available analog components or a fixed configuration digital processor. This schematic utilizes about 3% of the resources in the smallest version of the QSC Q-Sys virtual processor. By way of comparison a similar schematic took 19% of the available DSP resources on a single MediaMatrix board. This shows the great improvement in DSP processing speed in the latest generation of virtual processors.

The larger and more complex the system, the greater the advantages of the virtual audio processor over previous technologies. Legislative chambers, stadiums, ballrooms, theme parks, and churches are among the facilities utilizing virtual audio processors.

One technique commonly used in legislative sound systems is called *mix-minus*. Often such systems will have a microphone and loudspeaker for each legislator. In order to prevent feedback, each loudspeaker receives a mix that does not contain its associated microphone signal. Signals from other nearby microphones are at a reduced level in the mix. The U.S. Senate sound system utilizes this technique. Since there are 100 senators each of whom has his or her own microphone and loudspeaker, plus leadership microphones and loudspeakers, there were over 100 microphones with over 100 associated loudspeakers, which would have required over 100 mixers each with over 100 inputs if it had been implemented with a straightforward matrix mixer. To reduce this complexity, the mix-minus technique was developed. It works on the concept that only a small number of microphone inputs need to be muted or reduced in level on any given output. A single large mixer is used to produce a mix of all the inputs called the sum. Each output mixer receives the sum and just those inputs that must be muted or reduced in level. The polarities of the direct inputs of the mixer are reversed, so that as their level is increased, they cancel out part or all of their audio from the sum at the output of the mixer. If a direct input is set to unity gain, it will perfectly subtract from the sum signal, thus eliminating that input from the mixer output. While this technique has been used in analog system designs, circuit stability restricted its practical use in larger systems. Digital systems add another potential complexity. As signals are processed and transferred between DSP processing chips, delays may be introduced. If the sum and direct input signals do not arrive at the mixer at exactly the same time, the direct signal will not properly cancel. Small amounts of signal delay on selected inputs may be required to assure that all the signals reach the input of any given mixer at the same time. Some virtual signal processors automatically provide such compensation, or provide it as an option. It is always possible to manually insert very small delays as required.

Today, virtually all larger sound reinforcement systems utilize some form of virtual sound processor. The advantages of more optimized system design, the ability to make easy changes, and reduced cost, have made this sort of processor the overwhelming favorite of consultants and contractors worldwide.

Systems installed today still use analog microphones and microphone preamps, and in some cases analog mixing consoles, the outputs of which are fed to the virtual sound processor. Likewise, the outputs of the virtual sound processors are usually connected in the analog domain to conventional power amplifiers, which are then wired to loudspeakers. Thus considerable portions of the total sound system remain outside the scope of the virtual sound processor. There is a better way, however, that was first used by the U.S. Senate sound system installed in 1994, and continued in the new system installed in 2006.

Each senator has his or her own small microphone equipped with a tiny Kevlar reinforced cord. Suitable microphones with direct digital outputs were not available. The cord is managed by a Servoreeler Systems servo controlled reeler located in the senator's desk, under the control of the virtual system. No slip rings are used, and the far end of the cord is directly connected to the preamp, also located in the desk. The analog gain of the preamp is also under the control of the virtual system. The output of the preamp drives an A/D converter, which is connected to a DSP processor, and also located in the desk. The initial audio and control processing is done in the desk. Audio, control signals, and power are carried on a single cable between the desk and the central processor portion of the virtual sound system. The central processor performs all the mix-minus processing, and many of the auxiliary functions. Outputs from the central processor go back over the same cable to the desk, where further processing is done, still under the control of the virtual system, and the power amplifiers and speakers are driven.

What is special about this system is that not only is the central processing done in a virtual sound processor, but the processing associated with the microphones and loudspeakers is also part of the virtual system. The entire system consisting of redundant central processors, over 100 desk units, custom operator's console, and several displays, is all part of a single integrated virtual sound system. All of the DSP processing, including both central processors, and the over 100 remote processors in the desks, is loaded with their operating code and controlled, from a single common virtual sound system program. Even the microphone reelers, which are a servo controlled electromechanical subsystem, and the analog microphone preamps, are under the control of the virtual sound system. There are no unnecessary A/D and D/A conversions, and the longest analog interconnection is the microphone cable. Sound is converted from analog into digital at the end of the microphone cable, and remains in the digital domain until it is in the loudspeaker enclosure. The U.S. Senate sound systems, both the original of 1994 and the updated 2006 system, can be considered to be prototypes for the virtual sound systems of the future.

41.4 Virtual Sound Systems

41.4.1 Microphones

The virtual sound system of the future will be programmed and controlled through a single unified user interface program. It will have no analog interconnections.

Microphones will have a direct digital output, and receive power and control signals through the microphone cable. The Audio

Engineering Society Standards Committee has issued the AES-42-2010 Standard defining a digital interface for microphones. The digital audio transmission scheme used is based on the AES3 Standard, but adds digital phantom power, microphone control, and synchronization features. The microphones can phase-lock their internal sampling clocks to that of the equipment to which they are connected. The first microphones meeting this standard contain conventional analog microphone elements with conversion into the digital domain done inside of the microphone body. In the future, we may see microphones that produce digital signals directly out of the microphone element. In either case, these new smart digital microphones can be controlled by the virtual system to which they are connected. Some of these microphones will even allow their directional patterns to be changed, and in some cases to be steered toward the sound sources under the control of the virtual sound system. By dynamically adjusting the pickup pattern and direction of each of the microphones, the sound system may adaptively optimize its performance.

Microphone arrays will enhance the control of the directional patterns and aiming of microphones. Microphone arrays consist of from three to hundreds of microphone elements whose outputs are processed to produce one or more virtual microphones with controllable directional patterns and orientation. TOA has developed a Real-time Steering Array Microphone which is the first commercially available steerable array microphone suitable for use in sound reinforcement systems. It triangulates the location of the talker and steers a pickup beam that follows the talker's position. As the talker moves the pickup beam dynamically adapts. The range of talker locations the microphone will track is adjustable during setup. Gain before feedback is excellent even with the array microphone several feet from the talker, and similar to that of a gooseneck microphone located within a few inches of the talker's mouth. Because of this capability, dynamically self steered array microphones will be capable of picking up intelligible sound from a greater distance than traditional microphones. This allows sound systems to be designed without visible microphones, even in difficult acoustic environments.

Since the outputs of the individual microphone elements in an array are processed into the single virtual microphone output by DSP processing, adding more DSP will allow additional virtual microphones to be produced from the same array. Each virtual microphone can be aimed in a different direction, with its own directional pattern. We already see the beginnings of this in some of the microphone systems that provide 5.1 surround outputs for recording from a single compact microphone array. In a speech reinforcement system, each virtual microphone could track an individual talker. Talkers would be identified by their unique individual voiceprint. When there are multiple microphone arrays in the room, each talker's individual virtual microphone can automatically be formed using the optimum nearby array. As each talker moves around the room, his personal virtual microphone will always be formed using a nearby array, and will move from array to array as he moves around the room.

Since each virtual microphone will stay with its assigned talker, the output of each microphone may be individually optimized for the person to which it is assigned. When logging of the activities in the room is required, if desired each person could be recorded on her own individual track. Where speech to text conversion is utilized, having a separate virtual microphone for each talker is a significant advantage. Speech to text conversion is much easier when the system can learn the voice of a single individual. By providing outputs to the speech to text system that only contains the voice of a single individual, accuracy is greatly improved.

Microphone arrays will also have the ability to selectively reject sounds coming from certain sources. During system setup, the virtual microphone processors will be taught the location of the system loudspeakers, and of any significant noise sources in the room. This will allow them to keep a null in the directional pattern always aimed in those directions. As a result, the chances of feedback and the pickup of noise will be significantly reduced.

It will also be possible to define regions in 3D space from which speech will not be amplified. In legislative systems, for example, it is extremely important to make sure side conversations are never amplified. By defining an area slightly back from the desks as a privacy zone, the legislators will be able to lean back and have a private conversation with their aides even if they forget to turn their microphones off.

Voice tracking microphone arrays have been limited in their bandwidth, added significant signal latency, and were costly. These factors made them unattractive for sound reinforcement applications. However, improvements in processing algorithms, coupled with the dramatic reductions in the cost of DSP processing power we have seen each year, will soon bring this technology to a host of new applications including sound reinforcement.

41.4.2 Loudspeakers

Many loudspeakers today are powered with integrated power amplifiers and crossovers. Some loudspeakers have expanded on this concept by directly accepting digital audio and control signals. They contain DSP processing, which, integrated with the loudspeaker system design, allow much improved loudspeaker performance and protection. Modern DSP-based line array loudspeakers have steerable directional patterns, and in some cases can produce multiple acoustic output beams from the same loudspeaker. They may even send back an audio sample of their acoustic output for confidence monitoring.

As with microphone arrays, DS-based loudspeaker arrays allow sound to be steered to where it is needed, and kept from where it is not wanted. Dynamically controlled loudspeaker arrays will allow the loudspeaker coverage to change as room and system conditions change. Loudspeaker arrays may be produced as lines or flat panels, which mount flush with the walls, ceilings, and other architectural room elements. No longer is it necessary for loudspeakers to be aimed in the direction we wish the sound to go. For example, it is quite feasible to mount a flat panel loudspeaker array in a convenient location on the sidewall of the room, and direct the sound downwards and back into the audience area. Loudspeaker coverage patterns and directions may be changed under the control of the virtual system for different uses of the facility. This is a tremendous advantage over the older technology, which required either multiple sets of speakers, or physically changing the loudspeaker aiming for different applications.

A single loudspeaker array may be used to simultaneously produce multiple sound coverage patterns, each of which may be driven by its own independent sound source if so desired. One application of this technique would allow greatly enlarging the area in a room where accurate multichannel reproduction could be heard. Those located towards the edges of the room could now receive properly balanced sound from all loudspeakers in the room, even though they were much closer to some of the loudspeakers than to others, thus preserving the spatial reproduction. In another application, the same loudspeaker could aim direct sound at the audience, while simultaneously aiming ambient effects toward other portions of the room.

Control feedback from the virtual sound system will allow automatic modification of the loudspeaker coverage pattern as environmental conditions change. Such changes might include audience size and location, ambient noise, temperature, and wind speed and direction. Integration of DSP processing will also allow other useful functions to be moved into the loudspeaker cabinet. These will include source signal selection and mixing, delay, equalization, compression, limiting and driver protection, and ambient level compensation. The programming and control of the DSP processing will be over the same connection that brings the digital audio to the loudspeaker. This will allow the integration of all loudspeaker functions as part of the common virtual sound system.

41.4.3 Processing System

Those portions of the audio processing that are not contained either in the microphones or the loudspeakers will be contained in the central processing system. This may be either a single processor, or a networked array of processors. In either case there will be a single user interface for programming and controlling the entire system.

Control and monitoring of the virtual sound system may occur from many locations concurrently. The system will be controllable from PCs running either dedicated control software, or even standard Web browsers. For situations where control via a mouse is not acceptable, touch screen controllers will be available. Where physical controls are desired, a variety of standard, modular, control panel elements will be available. These will allow implementation of physical controls as simple as a wall mounted volume control, or as complex as a large mixing console.

Virtual sound processors have evolved substantially since the first products of this type were introduced in the early 90s. As the processing power available in these products has grown so have the capabilities.

Sound systems exist in a real-world environment, which also contains many other elements with which the sound system operation must be integrated. The most advanced of today's virtual sound processors contain powerful control logic subsystems to ease this integration. High-speed control connections allow exchange of data with external room and building control systems.

QSC Audio's Q-Sys audio processing product suite has advanced the reliability, sophistication, and capabilities of virtual audio processing products. The Q-Sys offering incorporates many functions that previously were available only in distinctly separate products. These include advanced virtual devices such as FIR filters, feedback suppressions, and ambient level sensing. It also greatly reduces the amount of time needed to compile, as well as incorporates a very low, and fixed, latency between all inputs and outputs. The Q-Sys allows the designer to easily create a fully redundant system, answering much of the concern that was initially generated by the use of digital systems for all of a facility's audio signal processing and control.

One very significant advantage of the most advanced virtual sound processing systems is the ease with which it is possible to make the various processing subsections interact with each other. For example, an automatic microphone mixer can be thought of as multiple-level meters and gain blocks, where the signal level at the various inputs is used to adjust the instantaneous gain of the various gain blocks. Such automatic microphone mixers exist in analog, digital, and virtual form. However, that sort of interaction can be expanded greatly to the system level in a virtual sound processor. For example, each microphone input processing chain might contain an AGC. The maximum possible gain an individual AGC can insert while still keeping the entire sound system stable will depend in part on the amount of gain or loss the AGCs for every other microphone are applying. In a virtual system it is possible to let each AGC know what the other AGCs are doing, and based on that information modify its behavior.

There are devices on the market that dynamically insert notch filters to keep a sound system from going into feedback. They do this by monitoring the onset of feedback and very quickly applying the corrective filters. This means the system must slightly start to ring before correction can be applied. A virtual sound system, by contrast, can monitor all the factors that impact system stability and insert corrective notch filters selectively in only the signal path required, and do so before the system starts ringing.

A virtual sound system can be programmed to know which are the most critical microphones and loudspeaker zones, and if trade-offs must be made to get optimum performance, can optimize the most important inputs and outputs. For example, if there is a person who must be heard, and that person is speaking in a very soft tone of voice and as a result the gain can't be gotten high enough, the virtual sound system can bring the gain up higher in the most critical loudspeaker zones while not increasing the gain everywhere, thus keeping the overall system stable.

A virtual sound processor may have many thousands of controls that need to be adjusted during a systems initial setup. Today's advanced virtual processors contain control tools that allow the system commissioning engineer a much simplified interface for adjusting those controls. This greatly reduces the time required and the chances for error in setup.

In short a well-designed virtual sound system can apply all the little tweaks to the system's controls that a very skilled operator would have applied if he or she could respond to conditions in a split second and adjust hundreds of controls at once.

41.4.4 Active Acoustics

The virtual sound system may also be used to modify the acoustic environment.

The reverberation time and reflection patterns of the space may be dynamically varied at any time to meet the needs of the program material. This requires that the physical acoustics of the space be at the low end of the desired reverberation range. The virtual sound system will add the initial reflections from the proper spatial directions, and the enveloping reverberant tail, to produce the desired acoustic environment. The ability to change the acoustics on an almost instantaneous basis allows each portion of a program to be performed in its optimum acoustics. For example, spoken portions of the program may only utilize a few supportive reflections. At the other extreme, choral or organ music may have a very long reverberation time. This technology may also be used to simulate the acoustic environment of the room in outdoor performance venues. LARES Associates and Meyer Sound are two of the leading firms designing digitally-variable acoustics system.

Environmental noise, particularly that of a lowfrequency and/or repetitive nature, may be actively canceled by the virtual sound system. As the cost of DSP processing comes down, and the power handling of transducers goes up, this technology will become more attractive in comparison to traditional noise control and isolation methods. Vibration and low-frequency sounds are the most difficult and costly to isolate using traditional passive methods. High displacement isolation mounts together with large amounts of mass are often required for good low-frequency performance. At higher frequencies often far less expensive techniques and materials are effective. By comparison, active noise and vibration control techniques are most effective at the lowest frequencies, but find it increasingly difficult to obtain satisfactory performance over large areas at higher frequencies. Therefore, including active noise control techniques in a virtual sound system to control low-frequency noises may prove beneficial in reducing the total project cost.

41.4.5 Diagnostics

The virtual sound system will monitor its own operation, and the environment in which it operates. The entire signal path will be monitored for failures. Depending on the level of system design, the operator may just be notified, or redundant equipment may be automatically utilized to assure uninterrupted operation. Most systems will utilize multiple microphones and loudspeakers. In itself, this provides a significant degree of redundancy. If the coverage pattern of the microphones or loudspeakers is controllable, then the virtual system can compensate for any given failure of a microphone or loudspeaker. Redundancy may also be designed into the interconnections and processing subsystems of the virtual sound system. With careful design, systems with few or no single points of failure can be built.

Environmental conditions that will impact the long term health of the system, such as temperature and airflow, will be monitored and trends logged. The performance of the microphones and loudspeakers in the system will be monitored and recorded to spot degradation of performance before it becomes audible. The acoustic environment will also be monitored to spot changes that might impact on the subjective performance of the sound system. System health reports will be automatically generated and sent to the system operator, installer, and designer when any of the parameters monitored are outside of expected tolerances. This capability will result in much more consistent performance over the life of the system, and will extend that life for years.

41.4.6 The Sound System of the Future

When all these techniques are combined, the virtual sound system of the future will have better performance, be more invisible to the user, be easier to operate, and have a longer life than any current system.

Chapter 42

Digital Audio Interfacing and Networking

by Ray Rayburn

42.1 Background

In most cases it is preferred to interface digital audio devices in the digital domain, instead of using analog interconnections. This is because every time audio is transformed from analog to digital, or digital to analog, there are inevitable quality losses. While analog interfacing is simple and well understood, there are few cases in which it would be desirable to interface two digital audio devices in the analog domain. If the digital audio devices are not provided with digital audio interfaces, for example, analog interfacing will be required. Such an analog interface, however, will result in subtle changes in the digital audio from one side of the interface to the other. The exact sequence of numbers that make up the digital audio will not be reproduced at the far side of an analog interface.

The numbering system commonly used in digital audio is called binary. Each of the digits (called bits) in the binary numbering system can be either a 1 or a 0. If two binary numbers are identical, then all their bits will match.

Digital audio interfaces have the potential to allow bit accurate transfer of the digital audio from one digital audio device to another, thus insuring no changes in the sequence of numbers that makes up the digital audio, and therefore potentially perfect accuracy. In order for this potential to be realized both digital audio devices must be synchronized.

Digital audio consists of a series of consecutive numeric samples of the audio, each of which must be received in sequence. If the samples are not received in proper sequence, or if samples are lost or repeated, then the audio will be distorted.

In order for digital audio devices to interconnect digitally, both ends of each connection must run at the same sampling rate. If the source is running at even a (very slightly faster) rate than the receiver, sooner or later the source will output a sample that the receiver is not ready to receive yet. This will result in the sample being lost. Similarly, if the source is running at even a (very slightly slower rate) than the receiver, eventually the receiver will be looking for a sample before the source is ready to send it. This will result in a new false sample being inserted into the data stream.

42.1.1 Synchronous Connections

The most straightforward way to carry digital audio is over a synchronous connection. In such a scheme, the data is transmitted at the exact same rate it is created, in other words, at the sample rate. When additional data is sent along with the audio in a synchronous system, it is added to the audio data and the whole package of information is transmitted at the audio sampling rate. Such systems send information in fixed-size groups of data, and introduce very little signal delay or latency. In a synchronous system the audio data words are sent and received at the audio sampling rate and both ends of the system must be locked to the same master sampling rate clock, Fig. 42-1. AES3 and IEC 90958 are examples of synchronous digital audio interconnection schemes.

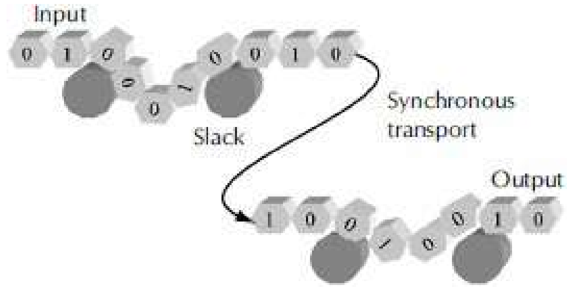


Figure 42-1. Synchronous connection between input and output.

42.1.2 Asynchronous Connections

An asynchronous system is in many ways the exact opposite of a synchronous system. Information is not sent at any particular time. The size of a given packet of information may vary. The time that it takes to get a given piece of information across an asynchronous connection may well be indeterminate. There is no common master clock that both ends of the connection refer to.

Examples of asynchronous transmission abound. When you mail a letter, it may contain just a short note on a single page, or it might contain the manuscript for a book. You put your letter in the mailbox (the outbound buffer) in the expectation that it will be picked up sometime later that day. The letter will pass through many different stages of transmission and storage along the way to its destination. You might know that the average delivery time is 3 days, however, in some cases the delivery might happen in 2 days, and in others it might be 6 days. You can be (almost) certain the letter will reach its destination eventually, but the exact delivery time can’t be known.

Other examples of asynchronous transmission include the Internet, and most common computer interfaces and networks including RS-232 serial, and Ethernet networking.

RealAudio and Windows Media Audio (WMA) are two common schemes for providing a synchronous audio connection over the asynchronous Internet. While there can be a large receive buffer of as much as 5 to 10 seconds, you will still experience dropouts in the audio due to network conditions on the Internet inserting delay on some audio packets that exceed the receive buffer delay. Generally the audio gets through OK, but every once in a while it drops out.

42.1.3 Isochronous Connections

Isochronous connections share properties of both the synchronous and asynchronous systems, and bridge the gap between the two. Information is not sent at a constant rate locked to a master clock. It provides a maximum delivery time for information that is only rarely if ever exceeded. By using buffers at each end, it can carry information such as audio where the delivery of audio words in proper sequence, and with a known and constant delay, is essential. In a properly designed isochronous system latency can be very low providing near real time operation, and reliability can be very high, Fig. 42-2.

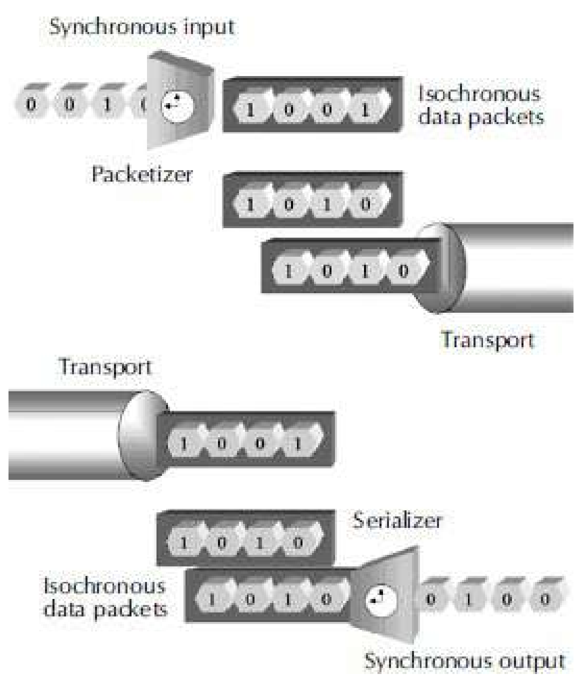


Figure 42-2. An isochronous system.

Examples of isochronous systems include ATM that is commonly used for transmitting telephone calls and computer data, and

CobraNet® audio networking. FireWire (IEEE-1394) is a networking scheme that combines both isochronous and asynchronous elements.

42.1.4 AES5

AES5 standardizes on a primary sampling frequency for professional audio use of 48kHz. It also allows 44.1kHz to be used when compatibility with consumer equipment is required. For broadcast and transmission-related applications where a 15kHz bandwidth is acceptable it allows a sampling frequency of 32kHz to be used. For applications where a wider than 20kHz bandwidth is desired, or a relaxed slope of the antialiasing filter is preferred, a sampling rate of 96kHz may be used.

Higher and in some cases much higher sampling rates are in use internally in digital audio equipment. When such a higher rate appears on an external digital audio interface, the AES recommends the rate be a multiple of a factor of two of one of the approved sampling rates above.

AES5 discourages the use of other sampling rates, although others are in use.

The above information is based on AES5-2008 (reaffirmed 2013). It is always advisable to obtain the latest revision of the standard.

42.1.5 Digital Audio Interconnections

In a simple chain of interconnected digital audio devices, it is possible for each device to look at the sampling rate of the incoming digital audio, and lock itself to that incoming rate. One problem with this system is that the sampling rate as recovered from the incoming digital audio is less than a perfect steady rate. It will have slight variations in its rate known as jitter. While there are techniques available to reduce this jitter, they add cost, and are never perfect. Each consecutive device in the chain will tend to increase this jitter. If the jitter gets too high, the receiving device may not correctly interpret the digital audio signals, and bit accuracy will be lost. Worse, the performance of analog to digital and digital to analog convertors is very dependent on a precise and steady clock. Even very small amounts of jitter can significantly degrade the performance of convertors. As a result, it is not recommended to cascade very many digital audio devices in this manner.

If a single digital audio device such as a mixer will be receiving digital audio from more than one source, then this simple scheme for synchronizing to the incoming digital audio signal breaks down, since it is only possible to synchronize to a single source at a time. There are two ways to solve this problem.

One way is to use a sample rate converter” (SRC) on each input to convert the incoming sample rate to the internal sample rate of the digital audio device. Such a SRC will add cost to the input, and will in some subtle ways degrade the quality of the audio. The accuracy will be better than with an analog interfacing, but the digital audio will not be transferred with bit accuracy. Of course, there are different degrees of perfection available in SRCs at correspondingly different levels of complexity and cost. Some SRCs will only handle incoming digital audio that is at a precise and simple numeric ratio to the internal sample rate. Others will accept any incoming sample rate over a very wide range and convert it to the internal sampling rate.

This second sort of SRC is very useful when you must accept digital audio from multiple sources that have no common reference, and convert them all to a common internal sampling rate.

As implied above, the other way to handle inputs from multiple digital audio sources is to lock all the devices in the digital audio system to a single common reference clock rate. In large systems this is the preferred solution, and the Audio Engineering Society has developed the AES11 Standard that explains in detail how to properly implement such a system. Such a system can have excellent jitter performance since each device directly receives its sampling rate reference from a common source. Interconnections between the digital audio devices can be rearranged freely since we do not have to be concerned about synchronization and jitter changes as the signal flow is changed.

42.1.6 AES11

AES11 defines a digital audio reference signal (DARS) that is merely an accurate AES3 signal used as the common reference clock for a facility. The DARS may contain audio signals, but is not required to do so.

There are four basic modes of operation defined in AES11: use of a DARS, use of the embedded clock in the AES3 signal, use of a common master video reference clock from which a DARS is derived, or use of a GPS receiver to reference a DARS and provide a time of day sample address code in bytes 18 to 21 of channel status. Use of a DARS is considered normal studio practice. As mentioned above cascading AES3 signals through devices without a DARS can lead to increased jitter.

The only flaw in this scheme is that some digital audio devices may not have provisions for accepting an external sampling rate reference. As a result, in many complex systems, while there may be a master sample rate clock that most equipment is locked to, there often is still a need for sample rate convertors to accept the output of those devices that can't lock to the master clock. AES11 acknowledges this limitation.

AES11 specifies two grades of DARS, grade 1 and grade 2. A DARS that has as its primary purpose studio synchronization should be identified in byte 4 bits 0–1 of the AES3 channel status. More details are given on this below.

A grade 1 DARS is the highest quality and is intended for use in synchronizing either a multiroom studio complex or a single room. It requires a long term stability of ±1ppm. Devices producing a grade 1 DARS are only expected to themselves lock to signals of grade 1 quality. Devices that are only expected to lock to grade 1 signals are required to lock to signals over a range of ±2ppm.

A grade 2 DARS is intended for use in synchronizing only within a single room where the added expense of a grade 1 solution can't be justified. It requires a long term stability of ±10ppm. Devices expected to lock to grade 2 signals are required to lock to signals over a range of ±50ppm.

The above information is based on AES11-2009. It is always advisable to obtain the latest revision of the standard.

42.2 AES3

The Audio Engineering Society titled their AES3 Standard "Serial transmission format for two-channel linearly-represented digital audio data." Let's break that title apart as our first step in examining the AES3 Standard.

This standard sends the data in serial form. In other words it sends the information to be transmitted as a sequence of bits down a single transmission path, as opposed to sending each bit down a separate transmission path. Each bit of data making up a single sample of the audio is sent in sequence starting with the least significant bit on up to the most significant bit. The least significant bit is the bit that defines the smallest change in the audio level, while the most significant bit is the one that defines the largest change in the audio level.

AES3 normally is used to transmit two channels of audio data down a single transmission path. The data for channel one of a given audio sample period is sent first, followed by the data for channel two of the same sample. This sequence is then repeated for the next sample period.

Most professional digital audio today is linearly represented digital audio data. This is also sometimes called pulse code modulation, or PCM. In such a scheme for numerically representing audio, each time sample of audio is represented by a number indicating its place in a range of equal-sized amplitude steps. If eight bits were being used to represent the audio level, there would be 2^8 or 256 equal-sized amplitude steps between the smallest level that could be represented and the largest. The smallest amplitude change that could be represented is exactly the same at the low-level portion of the range as at the highest-level portion. This is important to understand since not all digital audio uses such a linear representation. For example, often telephone calls are encoded using nonlinear techniques to maximize the speech quality transmitted using a limited number of bits. In professional audio we generally use larger numbers of bits, usually in the range of 16 to 24, that allow excellent performance with linear representation. Linear representation makes it easier to build high-quality converters and signal processing algorithms.

The bits that make up an audio sample word are represented in two's complement form, and range from the least significant bit (LSB) that represents the smallest possible amplitude change, to the most significant bit (MSB) that represents the polarity of the signal.

AES3 adds a considerable amount of structure around the basic sequence of bits described above in order to allow clock recovery from the received signal, provide a robust signal that is easily transmitted through paths of limited bandwidth, and provide for additional signaling and data transmission down the same path.

Each of the two audio channels that can be carried by an AES3 signal is formatted into a sequence of two subframes, numbered 1 and 2, each of which follows the following format.

The following information is based on AES3-1-2009, AES3-2-2009, AES3-3-2009, and AES3-4-2009 which together make up the AES3 Standard. It is always advisable to obtain the latest revision of the standard.

42.2.1 Subframe Format

First, additional bits are added before and after the digital audio to make a subframe of exactly 32 bits in length. The bits are transmitted in the sequence shown from left to right, Fig. 42-3.

If the audio data contains 20 or fewer bits, subframe format B is used. If the audio data contains 21 to 24 bits, subframe format A is used. In either case if data containing less than 20 or 24 bits is used, extra zeros are added to the LSB to bring the total number of bits to

20 or 24. Since the data is in two's complement form, it is important that the MSB, representing the signal polarity, always be located in bit 27.

The preamble is used to indicate if the audio to follow is channel one or two, and to indicate the start of a block of 192 frames.

If 20 or less audio bits are carried, then AES3 allows 4 bits of other data to be carried by the AUX bits.

The validity bit is zero if it is permissible to convert the audio bits into analog, and one if the conversion should not be done. Neither state should be considered a default state.

The user bit may be used in any way for any purpose. A few possible formats for using this bit are specified by the standard. Use of one of these formats is indicated by the data in byte one, bits 4 through 7 of the channel status information. If the user bit is not used it defaults to zero.

The channel status bit carries information about the audio signal in the same subframe, in accordance with a scheme that will be described later.

The parity bit is added to the end of each subframe and is selected so the subframe contains an even number of ones and an even number of zeros. This is called even parity. It allows a simple form of error checking on the received signal.

42.2.2 Frame Format

A subframe from channel two follows a subframe from channel one. The pair of subframes in sequence is called a frame. In normal use frames are transmitted at exactly the sampling rate. Again the data is transmitted in the sequence shown from left to right.

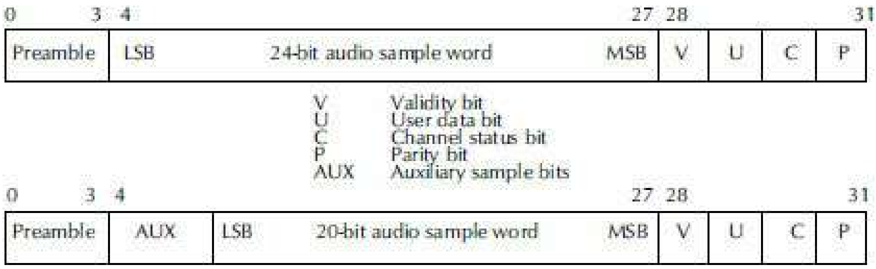


Figure 42-3. AES3 subframe format. Note that the first bit is called bit 0.

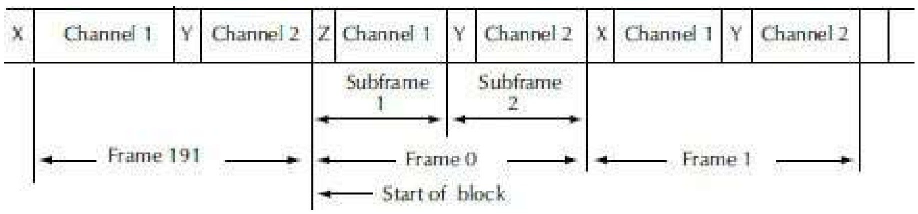


Figure 42-4. AES3 frame format. Note that the subframes are numbered 1 and 2, but frames are numbered starting with frame 0.

The parts shown as X, Y, and Z in Fig. 42-4 represent the three versions of the preamble portion of each subframe. When version Z is used, it indicates the start of a block of 192 frames. When version X or Z is used, it indicates that the channel data to follow is from channel one. When version Y is used, it indicates that the channel data to follow is from channel two.

Blocks are used to organize the transmission of channel status data, Fig. 42-4.

42.2.3 Channel Coding

AES3 needs to be able to be transmitted through transformers. Transformers can’t pass direct current (dc). Ordinary binary data can stay at 1 bit level for any arbitrary length of time, and thus by its nature can contain a dc component. Therefore a coding scheme is needed that eliminates this possibility.

We must also be able to recover the sampling rate clock from the AES3 signal itself. It was desired not to have to rely on a separate connection to carry the sampling rate clock. Since ordinary binary can stay at a given bit level for any arbitrary length of time, it is not possible to extract the clock from such a signal.

It was also desired to make AES3 insensitive to polarity reversals in the transmission media.

To meet these three requirements, all of the data except the preambles is coded using a technique called biphase-mark.

The binary data shown in the source coding portion of the diagram above has the sequence 100110.

The clock marks shown are at twice the bit rate of the binary source coding, and specify a time called the unit interval (UI), Fig. 42-5.

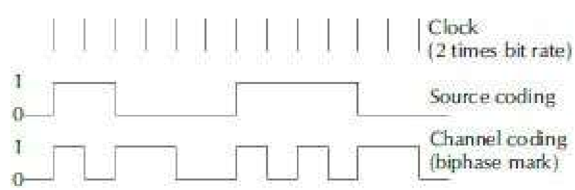


Figure 42-5. AES3 channel coding. The time between clock pulses is called the unit interval (UI).

The channel coded data sequence has a transition at every boundary between bits of the original source coding, whether or not the source coding has such a transition. This allows extraction of the original clock rate from the received signal since there always is a transition at every source bit boundary.

If the source coding data is a one, the channel coding will insert a transition in the middle of the source coding bit time. If the source coding data is a zero, the channel coding will not insert any additional transition.

The combination of these channel coding characteristics provides the desired features. There is no dc component, so the signal may be transmitted through transformers. The sampling rate clock may be extracted from the signal. The signal is insensitive to polarity reversals since the source data state is carried by the presence or absence of an additional signal transition rather than the coded data state itself.

42.2.4 Preambles

The single portion of the subframe that is not encoded using biphase-mark coding is the preamble. In fact the preambles are deliberately designed to violate the biphase-mark rules. This is done to allow easy identification of the preamble and to avoid any possibility that some data pattern could by chance duplicate a preamble.

This also allows the receiver to identify the preamble and synchronize itself to the incoming audio within one sample period. This makes for a robust reliable transmission scheme.

As mentioned in the Frame Format section above, there are three different possible preambles. Each preamble is sent at a clock rate equal to twice the bit rate of the source coding. Thus the eight states of each preamble are sent in 4 bit time slots at the beginning of each subframe.

The state of the beginning of the preamble must always be opposite that of the second state of the parity bit that ends the subframe before it.

Preceding State	0	1	
	Channel Coding		
"X"	11100010	00011101	Subframe 1
"Y"	11100100	00011011	Subframe 2
"Z"	11101000	00010111	Subframe 1 and block start

You will note that the two versions of each preamble are simply polarity reversed versions of each other.

In practice, due to the nature of the positive parity used for the bit before the preamble, and the biphase coding, only one version of each preamble will ever be transmitted. However, to preserve the insensitivity to polarity inversions, AES3 receivers must be able to accept either version of each preamble, Fig. 42-6.

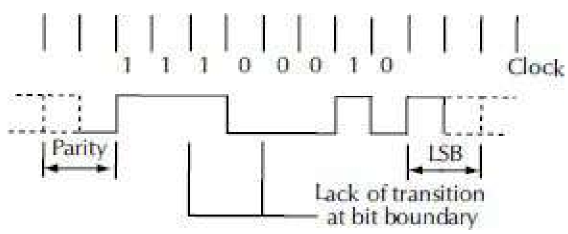


Figure 42-6. AES3 Preamble X (11100010). The time between clock pulses is called the unit interval (UI).

Like biphase-mark coding, the preambles are dc free and allow for clock recovery while differing from valid biphase-mark coding at least twice.

The clock rate shown above is at twice the source bit rate. Note that the second state of the parity bit is always zero, and therefore the preamble will always start with a transition from zero to one. Also note that in this preamble, as in all possible preambles, there are at least two places where there is no transition at a bit boundary thus violating the rules for biphase-mark coding and providing positive identification of the preamble.

42.2.5 Channel Status Format

Each audio channel has its own channel status bit. The data carried by that bit is associated with its own audio channel. There is no requirement that the data for each channel be identical, although it could be, Fig. 42-7.

The sequence of 192 channel status bits in a given block is treated as 24 bytes of data, as shown in Table 42-1.

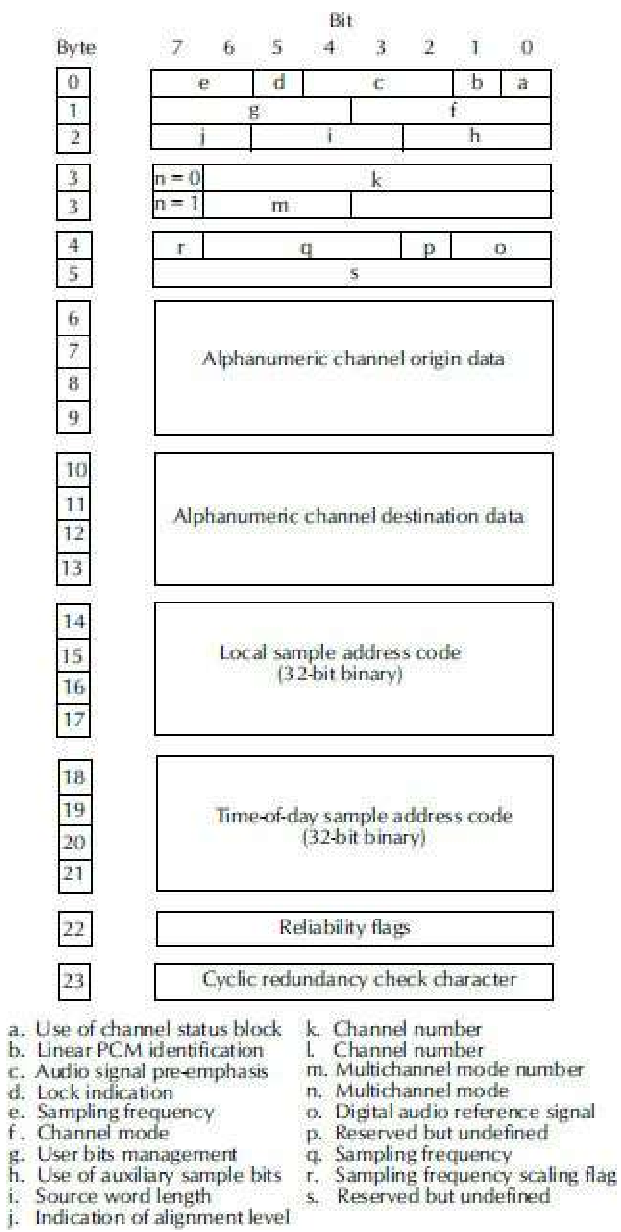


Figure 42-7. AES3 Channel Status Data Format. Note both the bits and bytes are numbered starting with 0.

Table 42-1. Channel Status Data Format Details

Byte 0		
Bit 0	0	Contents of the channel status block conform to IEC 60958-3 "consumer use" Standard. (The remainder of this table only applies to "professional use") (See Note 1.)
	1	Contents of the channel status block as to the AES3 "professional use" Standard.
Bit 1	0	Audio words consist of linear PCM samples.
	1	Audio words consist of something other than linear PCM samples.
Bits 2–4	Encoded Audio Signal Emphasis	
Bit	2 3 4	
State	0 0 0	No emphasis indicated. Receiver defaults to no emphasis but may be manually overridden.
	1 0 0	No emphasis used. Receiver may not be manually overridden.
	1 1 0	50/15μs emphasis used per ITU-R BS.450. Receiver may not be manually overridden.
	1 1 1	ITU-T J.17 emphasis (with 6.5dB insertion loss at 800Hz). Receiver may not be manually overridden.
	All other possible states of bits 2–4 are reserved and are not to be used unless defined by the AES in the future.	
Bit 5	0	Lock not indicated. This is the default.
	1	Source sampling frequency unlocked.
Bits 6–7	Encoded Sampling Frequency (See Notes 2, 3 and 4.)	
Bit	6 7	
State	0 0	No sampling frequency indicated. This is the default. Receiver defaults to interface frame rate, and automatic sampling rate determination or manual override is enabled.
	0 1	48kHz sampling rate. Automatic sampling rate determination or manual overrides are disabled.
	1 0	44.1kHz sampling rate. Automatic sampling rate determination or manual overrides are disabled.
	1 1	32kHz sampling rate. Automatic sampling rate determination or manual overrides are disabled.

Note 1. Other than the use of the Channel Status block of information, the rest of the data format is identical between the AES3 "professional use" Standard and the IEC 60958-3 "consumer use" Standard. The electrical format is different, however. For these reasons it should never be assumed that a "consumer use" receiver would function correctly with a "professional use" transmitter, or vice versa.

Note 2. If Byte 0 Bit 1 indicates other than linear PCM, the validity bit must be set for that channel and certain other bits are reserved.

Note 3. It is not a requirement that the sampling frequency used be indicated by these bits, nor is the use of one of the sampling frequencies that can be indicated by these bits. If the transmitter does not support sampling frequency indication, the sampling frequency is unknown, or the sampling frequency is not one that can be indicated by these bits, then the bits should be set to 0 0. Bits 3–6 of Byte 4 of the Channel Status may indicate other possible sampling rates.

Note 4. If Bits 0–3 of Byte 1 indicate single channel double sampling frequency mode, then the sampling frequency indicated by Bits 6 and 7 of Byte 0 is doubled.

Byte 1

Bits 0–3	Encoded Channel Mode	
Bit	0 1 2 3	
State	0 0 0 0	No mode indicated. Receiver defaults to two channel mode but may be manually overridden.

0 0 0 1 Two channel mode used. Receiver may not be manually overridden.

0 0 1 0 Single channel (monophonic) mode used. Receiver may not be manually overridden.

0 0 1 1 Primary/Secondary (subframe 1 is primary) mode used. Receiver may not be manually overridden.

0 1 0 0 Stereophonic (subframe 1 is left) mode used. Receiver may not be manually overridden.

0 1 0 1 Reserved for user defined applications.

0 1 1 0 Reserved for user defined applications.

0 1 1 1 Single channel double sampling frequency mode. Subframes 1 and 2 contain successive samples of the same signal. Sampling frequency is double the frame rate and double the rate indicated in Byte 0 (if a rate is indicated there), but not double the rate indicated in Byte 4 (if a rate is indicated there). Receiver may not be manually overridden. Byte 3 may indicate channel number.

1 0 0 0 Single channel double sampling frequency mode—left stereo channel. Subframes 1 and 2 contain successive samples of the same signal. Sampling frequency is double the frame rate and double the rate indicated in Byte 0 (if a rate is indicated there), but not double the rate indicated in Byte 4 (if a rate is indicated there). Receiver may not be manually overridden.

1 0 0 1 Single channel double sampling frequency mode—right stereo channel. Subframes 1 and 2 contain successive samples of the same signal. Sampling frequency is double the frame rate and double the rate indicated in Byte 0 (if a rate is indicated there), but not double the rate indicated in Byte 4 (if a rate is indicated there). Receiver may not be manually overridden.

1 1 1 1 Multichannel mode. Byte 3 indicates the channel numbers.

All other possible states of bits 0–3 are reserved and are not to be used unless defined by the AES in the future.

Bits 4–7 Encoded User Bits Management

Bit	4 5 6 7	
State	0 0 0 0	No user information indicated—default.
	0 0 0 1	192 bit block of user data, starting with the Preamble "Z."
	0 0 1 0	Data to the AES18 Standard.
	0 0 1 1	User defined.
	0 1 0 0	Data to the IEC 60958-3 Standard.
	0 1 0 1	192 bit block as specified in AES52, starting with the Preamble "Z".
	0 1 1 0	Data to the IEC 62537 Standard
	All other possible states of bits 4–7 are reserved and are not to be used unless defined by the AES in the future.	
Byte 2		
Bits 0–2	Encoded use of auxiliary sample bits.	

Bit	0 1 2		
State	0 0 0	Maximum 20 bit audio words, auxiliary sample bits usage not defined, default.	
	0 0 1	Maximum 24 bit audio words, auxiliary sample bits used for audio.	
	0 1 0	Maximum 20 bit audio words, auxiliary sample bits used for a coordination signal per Annex B of AES3-2-2009.	
	0 1 1	User defined applications.	
	All other possible states of bits 0–2 are reserved and are not to be used unless defined by the AES in the future.		
Bits 3–5	Encoded audio word length (see notes 1, 2, 3, and 4).		
Bit	3 4 5	Audio sample word length if bits 0–2 indicate maximum 20 bit length.	Audio sample word length if bits 0–2 indicate maximum 24 bit length.
State	0 0 0	Length not indicated, default	Length not indicated, default
	0 0 1	19 bits	23 bits
	0 1 0	18 bits	22 bits
	0 1 1	17 bits	21 bits
	1 0 0	16 bits	20 bits
	1 0 1	20 bits	24 bits
	All other possible states of bits 3 – 5 are reserved and are not to be used unless defined by the AES in the future.		
Bits 6–7	Alignment level indication.		
Bit	6 7		
State	0 0	Not indicated.	
	0 1	SMPTE RP155 (alignment level is 20 dB below maximum level).	
	1 0	EBU R68 (alignment level is 18.06 dB below maximum level).	
	1 1	Reserved for future use.	

Note 1. If the default state or bits 3–5 is indicated, the receiver should default to 20 or 24 bits as specified by bits 0–2, but allow manual override or auto set.

Note 2. If other than the default state of bits 3–5 is indicated, the receiver should not allow manual override or auto set.

Note 3. No matter which audio word length is indicated, the MSB representing the signal polarity is always bit 27 of the subframe.

Note 4. Knowledge of the actual encoded audio word length can be used to allow the receiving device to properly re-dither the audio to a different word length if so required.

Byte 3

Bit 7	Defines the meaning of bits 0–6.	
State	0	Undefined multichannel mode, default.
	1	Defined multichannel modes.
Bits 0–6	Channel number if bit 7 is 0. Channel number is value of bits 0–6 (bit 0 is LSB) plus 1.	
Bits 4–6	Multichannel mode if bit 7 is 1.	
Bit	4 5 6	
State	0 0 0	Multichannel mode 0. Bits 0–3 specify the channel.
	1 0 0	Multichannel mode 1. Bits 0–3 specify the channel.
	0 1 0	Multichannel mode 2. Bits 0–3 specify the channel.
	1 1 0	Multichannel mode 3. Bits 0–3 specify the channel.
	1 1 1	User defined multichannel mode. Bits 0–3 specify the channel.
Bits 0–3	Channel number if bit 7 is 1. Channel number is value of bits 0–3 (bit 0 is LSB) plus 1.	

Note 1. It is intended that the defined multichannel modes will identify mappings between channel numbers and function. Standardized mappings have yet to be defined.

Note 2. Some equipment may only consider the channel status data carried in one of the two subframes. Therefore if both subframes specify the same channel number, subframe 2 has a channel number one above channel 1 unless single channel double sampling frequency mode is in use.

Note 3. If bit 7 is 1, bits 0–3 correspond to the consumer mode channel status specified in IEC 60958-3. Consumer mode channel A is equivalent to channel 2, and Consumer mode channel B to channel 3 and so on.

Byte 4

Bits 0–1	Digital audio reference signal to the AES11 Standard.	
Bit	0 1	
State	0 0	This is not a reference signal, default.
	0 1	Grade 1 reference signal.
	1 0	Grade 2 reference signal.
	1 1	Reserved for future use.
Bit 2	PMC audio contains hidden information.	
	0	No information (default)
	1	PMC audio contains information in LSBs per AES55.
Bits 3–6	Sampling frequency.	
Bit	3 4 5 6	
State	0 0 0 0	No frequency indicated, default.
	1 0 0 0	24kHz.
	0 1 0 0	96kHz.
	1 1 0 0	192kHz.
	0 0 1 0	384kHz.
	1 0 1 0	Reserved.
	0 1 1 0	Reserved.
	1 1 1 0	Reserved.
	0 0 0 1	Reserved for vectoring.
	1 0 0 1	22.05kHz.
	0 1 0 1	88.2kHz.
	1 1 0 1	176.4kHz.
	0 0 1 1	352.8kHz.
	1 0 1 1	Reserved.
	0 1 1 1	Reserved.
	1 1 1 1	User defined.
Bit 7	Sampling frequency scaling flag.	
	0	No scaling, default.

1	Multiply sampling frequency indicated in byte 0 bits 6–7, or byte 4 bits 3–6, by 1/1.001.

Note 1. Bit 2 refers to the LSBs of the audio word, not the aux bits.

Note 2. When bit 2 is set to 1, audio processing including dithering, sample rate conversion, and level changes should not be done. The receiver seeing bit 2 set may look for information in the LSBs such as MPEG surround sound per ISO/IEC 23003-1.

Note 3. The sampling frequency as indicated in byte 4 is independent of the channel mode as indicated in byte 1.

Note 4. There is no requirement to use a particular sampling frequency, nor to use a sampling frequency that can be indicated in bytes 0 or 4. If the transmitter does not support indication of sampling frequency, the frequency is unknown, or the sampling frequency is not one that can be indicated in this byte, then bits 3–6 should be set to "0 0 0 0."

Note 5. It is intended to assign sampling frequencies in the future to the currently reserved states of byte 4 bits 3– 6 (except 0 0 0 1) such that if the rates are related to 44.1kHz bit 6 will be set, and if they are related to 48kHz bit 6 will be cleared. Do not use these reserved states unless defined in the future by the AES.

Byte 5

Bits 0–7	Reserved. Set to 0 unless defined in the future by the AES.

Bytes 6–9

	Alphanumeric channel origin data. Byte 6 contains the first character.
Bits 0–7 (each byte)	7 bit no parity data International Organization for Standardization (ISO) 646, Bit 7 is always 0. Transmit LSBs first. Nonprintable characters (codes 01 to 1F hex and 7F hex) must not be used. Default is all 0s (code 00 hex or ASCII null).

Bytes 10–13

Alphanumeric channel destination data. Byte 10 contains the first character.

(each byte)	7 bit no parity datano parity data ISO 646. IRV Bit 7 is always 0. Transmit LSBs first. Nonprintable characters (codes 01 to 1F hex and 7F hex) must not be used. Default is all 0s (code 00 hex or ASCII null).

Bytes 14–17

Local sample address code sent as 32 bit binary with LSBs first. Value is of the first sample in this block.

Bits 0–7

(each byte)	Transmit LSBs first. Default is all 0s.

Note 1. This serves the same function as an index counter on a recorder.

Bytes 18–21

Time of day sample address code sent as 32 bit binary with LSBs first. Value is of the first sample in this block.

Bits 0–7

(each byte)	Transmit LSBs first. Default is all 0s.

Note 1. This time of day is the time of the original analog to digital conversion, and should not be changed thereafter. Midnight is represented by all 0s. In order to convert this sample code into correct time, the sampling frequency must be known accurately.

Byte 22

Reserved until redefined by the AES in the future.

Byte 23

Channel status data Cyclic Redundancy Check Character (CRCC).

Bits 0–7 The CRCC allows the receiver to check for correct reception of the bytes 0 through 22 of the channel status block. It is generated by $G(x) = x^8 + x^4 + x^3 + x^2 + 1$. CRCC data is always required. The AES3 standard provides further information on how to calculate this.

Non-PCM channel status reservations. When byte 0 bits 0 & 1 are both set to 1 the following status bits are reserved as indicated

Byte	Bit	Function
0	5	Lock Indication
0	6–7	Sampling Frequency
1	4–7	User Bits Management
2	0–2	Use of Aux Bits
3	0–7	Multichannel Mode Indications
4	3–7	Sampling Frequency Multipliers & Scaling Flag
23	0–7	Channel Status Data CRCC

The sequence of channel status bits for each channel starts in the frame with Preamble Z.

42.3 AES3 Implementation

42.3.1 AES3 Transmitters

In a standard implementation, an AES3 transmitter must encode and transmit the audio words, validity bit, user bit, parity bit, the three preambles, and bytes 0, 1, and 23 (CRCC) of the channel status.

An enhanced implementation provides additional capabilities beyond the standard implementation.

All transmitters must be documented as to which of the channel status capabilities they support.

42.3.2 AES3 Receivers

All receivers must decode the channel status as required, and reject channel status blocks that have CRCC errors. Audio content must not be muted or rejected because of channel status errors.

All receivers must be documented as to which of the channel status capabilities they support.

42.3.3 Electrical Interface

AES3 uses a balanced 110Ω electrical interface based on the International Telegraph and Telephone Consultative Committee (CCITT) Recommendation V. 11. When used with shielded twisted pair 110Ω cable distances of up to 100 meters are supported without equalization and at frame rates of up to 50kHz. See Fig. 42-8.

The use of transformers and blocking capacitors in the driving and termination and isolation networks is strongly recommended. The AES42 (AES3-MIC) Digital Interface for Microphones Standard calls for digital microphones to be powered by 10Vdc digital phantom power that is a variation on the phantom power scheme used for analog microphones. This provides another excellent reason to provide dc blocking on all AES3 inputs and outputs, transformer based or not, to prevent damage if such a phantom power scheme were to be applied. Structured wiring using Cat 5 or high rated cable and RJ45 connectors is now a permitted alternative interconnect scheme for AES3 signal. These interconnects are also used for Ethernet, which may have power over Ethernet (PoE) applied. They are also used for plain old telephone service (POTS), which will have 48V battery and 90V ring signals. If structured cabling is used for AES3, consideration must be given to the survivability of the line driver and receiver circuits if accidentally interconnected to PoE or POTS lines.

Transformers will make possible higher rejection of common mode interfering signals, electromagnetic interference (EMI), and grounding problems than common active circuits. The European Broadcasting Union (EBU) in its version of this standard (EBU Tech. 3250-E) requires the use of transformers. This is the major difference between the standards. It is common to see the AES3 Standard referred to as the AES/EBU Standard even though that is not strictly correct since AES3 makes the transformers optional, while the EBU requires them.

Cabling to be used for AES3 may be 110Ω balanced twisted pair with shield or shielded (STP), or unshielded (UTP) Category 5e (Cat 5e) or better wiring. The impedance must be held over a frequency range from 100kHz to 128 times the maximum frame rate to be carried. The line driver and line receiver circuits must have an impedance of 110Ω±20% over the same frequency range. While the acceptable tolerance of the cable impedance is not specified, it is noted that tighter impedance tolerances for the cable, driver, and receiver will result in increased distance for reliable transmission, allow higher data rates, reduce susceptibility to electromagnetic interference, and reduce electromagnetic emissions.

If a 32kHz sampling rate mono signal were carried in single channel double sampling frequency mode, the interface frequency range would only extend to 2.048MHz. If a 48kHz sampling frequency two-channel signal were to be carried, the interface frequency range would extend to 6.144MHz, or about the 6MHz bandwidth commonly quoted for AES3. However, if a 192kHz sampling frequency two-channel signal were to be carried, the interface frequency range would extend to 24.576MHz. As you can see, some uses of AES3 can extend the frequency range far beyond 6MHz. If you are using a mode that has extended interface frequency, make sure that the transmitter, interconnect system, and receiver are all designed to meet specifications over the entire frequency range in use.

When AES3 was originally introduced it was thought that ordinary analog audio shielded twisted pair cable would be acceptable for carrying AES3 digital audio, and indeed that is often the case for shorter distances. However, the impedance and balance of common audio cable vary widely, and it was quickly determined that purpose built AES3 110Ω cable performed significantly better for AES3 than ordinary analog audio cable. It was later determined that such AES3 rated cable often also performed significantly better as analog audio cable than ordinary cable, so today we commonly see AES3 rated cable used in both analog and digital applications.

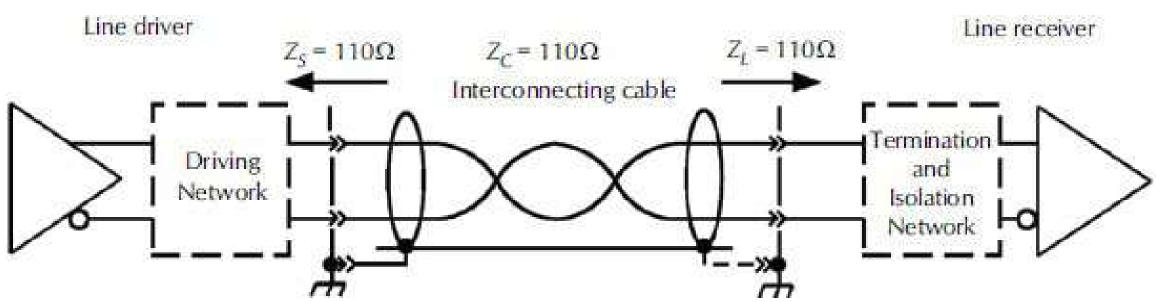


Figure 42-8. AES general circuit configuration.

While the AES3 Standard makes mention of interconnect lengths of “a few hundred meters,” in practice distances beyond about 100m often require the use of equalization to compensate for losses in the cabling. If such equalization is used, it must never be applied to the transmitter, but only to the receiver.

As an alternative to purpose built AES3 rated digital audio cable, structured wiring meeting Cat 5e or greater is acceptable. Such cabling can be of either shielded twisted pair (STP) or unshielded twisted pair (UTP) construction. To deliver satisfactory performance, only one cable type (Cat 5e or higher STP, Cat 5e or higher UTP, or AES3 digital audio rated) may be used for the entire path from driver to receiver. If Cat 5e or greater UTP is used, distances of 400 meters unequalized or 800 meters with equalization are possible. If structured cabling is used, the AES3 signal should be on pins 4 and 5 of the "RJ45" connector.

42.3.4 Line Drivers

Just like AES3 cabling, line drivers are specified as having a balanced output with an impedance of 110Ω±20% over the entire frequency range from 100kHz to 128 times the maximum frame rate. The driver must be capable of delivering an output level between 2–7V (measured peak to peak) (typical 4V±10%) into a 110Ω resistive termination directly across its output terminals with no cable present. The balance must be good enough so that any common mode output components are at least 30dB lower in level than the balanced output signal.

The rise and fall times of the output, as measured between the 10–90% amplitude points, must be no faster than 0.03UI, and no slower than 0.18UI into a 110Ω resistive termination directly across its output terminals with no cable present. If the frame rate is 48kHz, this is equivalent of no faster than 5ns, and no slower than 30ns. A faster rise and fall time often improves the eye pattern at the receiver, but a slower rise and fall time often results in lower electromagnetic interference (EMI) radiated. Equipment must meet the EMI limits of the country in which it is used.

Equalization must not be applied at the driven end of the line.

42.3.5 Jitter

All digital equipment has the potential for introducing jitter, or small timing variations in the output signal. Extreme amounts of jitter can actually cause data errors. More moderate amounts of jitter may not change the actual data transmitted, but can lead to other ill effects. An ideal D/A would ignore the jitter on the incoming signal and perfectly produce the analog output based solely on the data carried. Unfortunately many real-world A/D and D/A converters are far from ideal, and allow jitter to change or modulate the output. Therefore keeping jitter low can have significant audible benefits.

AES3 divides the jitter at the output of a line driver into two parts intrinsic and pass through. The pass through portion of the jitter is due to jitter in the timing reference used. If such an external timing reference is used AES3 requires that there never be more than 2dB of jitter gain at any frequency. The external timing reference may be derived from an AES3 input signal, or from a digital audio reference signal (DARS), which is an AES3 signal used as a clock reference as specified in AES11. If cascades of digital devices are built where each device uses as its clock reference the AES3 signal received from the previous device in the chain, it is possible for the pass through jitter to eventually increase the output jitter to an unacceptable level.

Many of today's better A/D and D/A converters provide jitter attenuation from the timing reference, Fig. 42-9.

Intrinsic jitter is measured with the equipment's own internal clock and with the equipment locked to an effectively jitter free external reference clock. Intrinsic jitter is measured through a minimum-phase one-pole high-pass filter whose –3dB down point is 700Hz, and which accurately provides that characteristic down to at least 70Hz. The pass band of the filter has unity gain. Measuring at the transition zero crossings and through the filter the jitter must be less than 0.025 unit intervals (UI), Fig. 42-5.

42.3.6 Line Receivers

Just like AES3 cabling and line drivers, line receivers are specified as having a balanced output with an impedance of 110 Ω±20% over the entire frequency range from 100kHz to 128 times the maximum frame rate. The receiver must be capable of accepting an input level of 2–7V (measured peak to peak). Early versions of AES3 required the receiver be able to accept 10V. Only one receiver may be connected to an AES3 line. Early versions of AES3 permitted multiple receivers, but it became clear this was not good practice and the standard was modified.

An AES3 receiver must correctly interpret data when a random data signal that is not less than V_{min} = 200mV and T_{min} = 0.5UI, as shown in Fig. 42-10, is applied to the receiver.

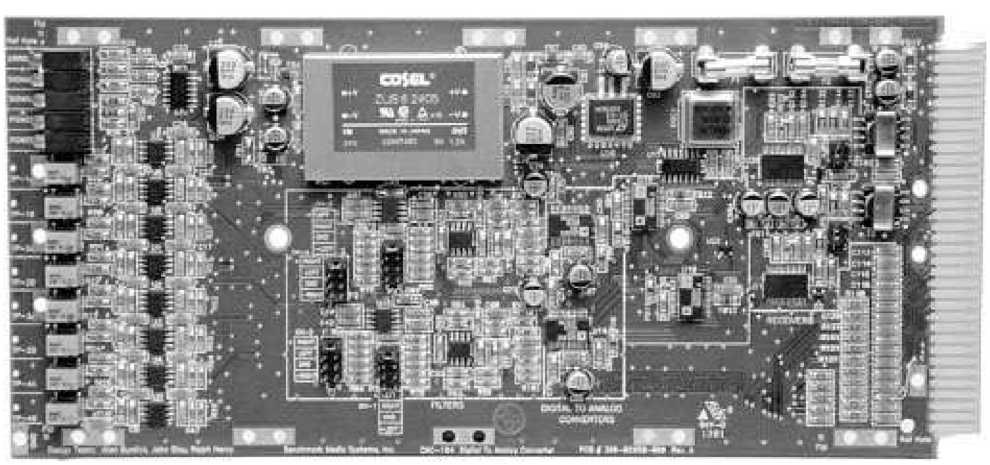


Figure 42-9. Benchmark Media Systems DAC-104 is a four channel 24 bit 96kHz sampling rate D/A converter. An example of a high performance D/A, it provides jitter reduction of 100dB at 1kHz and 160dB at 10kHz. Total harmonic distortion plus noise (*THD* + N) is less than 0.00079% at -3dB FS at any sampling rate and test frequency.

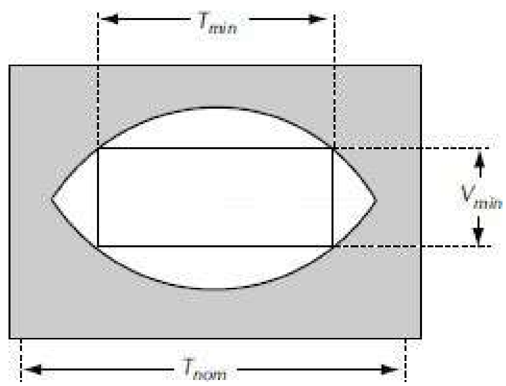


Figure 42-10. AES3 eye diagram. T_{nom} = 1 unit interval (UI) (see Fig. 42-5); T_{min} = 0.5 Tno_m; $V_{mi}n$ = 200mV. The eye diagram is one of the most powerful tools used to examine the quality of received data. The larger the open area of the eye the better. The limits shown are the most closed an eye should ever be for correct reception of the AES3 data.

If cable lengths of over 100m are to be used, optional receiver equalization may be applied. The amount of equalization needed depends on the cable characteristics, length, and the frame rate of the AES3 signal. The AES3 Standard suggests that at a 48kHz frame rate an equalizer with a boost that rises to a maximum of 12dB at 10MHz would be appropriate.

The receiver must introduce no errors due to the presence of common mode signals as large as 7Vp at any frequency from dc to 20kHz. This is not enough range to protect an AES3 receiver from the application of 10Vdc digital phantom power as specified in the AES42 (AES3-MIC) Digital Interface for Microphones Standard.

The receiver must introduce no data errors from jitter that does not exceed 10 unit intervals (UI) at frequencies below 200Hz decreasing to not exceeding 0.25UI at frequencies over 8kHz. Of course the recovered clock from such a high jitter signal may cause other problems, but at least the data must be decoded correctly.

42.3.7 AES3 Connectors

The connector for AES3 signals is what is commonly called the XLR, and is standardized in IEC 60268-12 as the circular latching 3 pin connector. Outputs are on male connectors and inputs are on female connectors just as in common analog usage of this same connector. The shield or ground connection is on pin 1, and the signal connections are on pins 2 and 3. With AES3 digital signals, the relative polarity of pins 2 and 3 is unimportant.

To avoid confusion with analog audio signal connectors, AES3 suggests that manufacturers label AES3 outputs “digital audio output,” or “DO;” and AES3 inputs “digital audio input,” or “DI.”

An alternative modified XLR connector has been proposed to help make clear that the signal on the connector is digital and not analog, and via a keying scheme reducing the chances of accidental interfacing of inputs and outputs that are incompatible. There has been much discussion in the AES about changing to this connector, either for all AES3 signals, or at least for AES42 (AES3-MIC) digital microphone signals, but no consensus has been reached. It should also be noted that since the analog audio bandwidth usually does not significantly exceed 20kHz, and the AES3 spectrum does not go below 100kHz, it is possible for a single cable to carry both an analog audio signal and an AES3 signal at the same time. The proposed modified XLR connector could also allow such a dual use condition.

If Cat 5e or greater UTP or STP is used, the 8-way modular connector specified in IEC 60603-7 (commonly but incorrectly called “RJ45”) must be used. Pins 4 and 5 of the “RJ45” are the preferred pair, with pins 3 and 6 the suggested alternative pair. It is suggested that if adaptors from XLR to “RJ45” are used, pin 2 of the XLR should connect to pin 5 (or other odd numbered pin) of the “RJ45”, and pin 3 of the XLR should connect to pin 4 (or other even numbered pin) of the “RJ45” connector.

If coaxial transmission of AES3 is used, the connector must mechanically be type BNC as specified in IEC 60169-8, but with a 75Ω impedance.

42.4 Coaxial Transmission of AES3

AES3 can be carried over unbalanced 75Ω coaxial cable (formerly called AES3-id) instead of over 110Ω balanced cable. It can allow the transmission of AES3 information over distances of 1000m or greater. Analog video distribution equipment and cable may often be suitable for coaxial transmission of AES3 data. This of course is a great convenience in video facilities.

At distances of up to 1000m, receiver equalization may not be needed. Equalization must never be applied at the line driver end.

AES-3-4-2009 and AES2id-2012 provides extensive tables and circuit diagrams showing active and passive circuits for AES-3 coaxial transmission. Canare, among others, sells passive adapters between 110Ω balanced AES3 and coaxial 75Ω unbalanced AES-3.

The following information is based on AES-3-4-2009 and AES2id-2012. It is always advisable to obtain the latest revision of the information document.

42.4.1 Line Driver

AES-3 coaxial line drivers must have an impedance of 75Ω and exhibit a return loss in excess of 15dB from 100kHz to 128 times the frame rate (6MHz for a 48kHz frame rate). Much but not all modern video gear will have the bandwidth to correctly handle higher sampling rates.

The peak to peak output voltage into a 75Ω 1% tolerance resistor must be between 0.8V and 1.2, with a dc offset not to exceed 50mV. The rise and fall times should be between 0.185 and 0.27UI (30ns and 44ns for a frame rate of 48kHz). These output voltage, dc offset, and rise and fall times have been chosen for compatibility with analog video distribution equipment. Lower dc offset values are desirable for longer transmission distances.

42.4.2 Interconnect System

AES-3 coaxial cable must be 75±3Ω over the range from 100kHz to 128 times the frame rate (6MHz for a 48kHz frame rate). It is to be equipped with BNC connectors as described in IEC 60169-8 but with an impedance of 75Ω instead of 50Ω.

42.4.3 Line Receiver

AES-3 coaxial line receivers must have a resistive impedance of 75Ω and exhibit a return loss in excess of 15dB from 100kHz to 128 times the frame rate (6MHz for a frame rate of 48kHz). The receiver must be capable of correctly decoding signals with input levels of 0.8V and 1.2V (measured peak to peak).

An AES-3 coaxial receiver must correctly interpret data when a random data signal that is not less than V_{min} = 320mV and T_{min} = 0.5UI as shown in Fig. 42-11 is applied to the receiver. For reliable operation at distances in excess of 1000m, a receiver that operates correctly with a V_{min} = 30mV may be required.

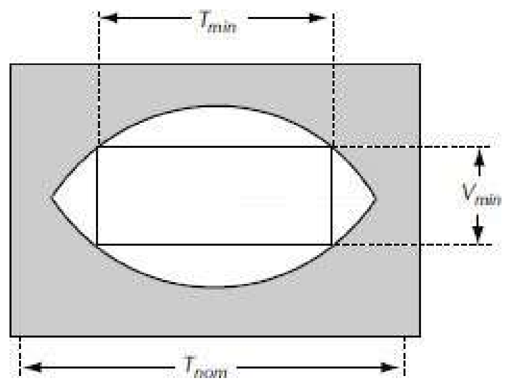


Figure 42-11. AES3 coaxial Eye diagram. T_{nom} = 1 unit interval (UI) (see Fig. 42-5); T_{min} = 0.5UI; V_{min} = 320mV. The eye diagram is one of the most powerful tools used to examine the quality of received data. The larger the open area of the eye the better. The limits shown are the most closed an eye should ever be for correct reception of the AES-3id data.

42.5 AES42

AES42 is a variant on AES3 designed to meet the needs of interfacing microphones that have direct digital outputs. The first most significant difference is that the transmitter and receiver use center tapped (on the cable side) transformer, which allow a digital phantom power (DPP) of +10 (+0.5, −0.5)Vdc at up to 250mA to be supplied to the microphone. No more than 50mVp-p ripple is allowed on the DPP. The microphone may draw no more than 250mA from the DPP, and may not present a load in excess of 120nF to the DPP. The microphone must not be damaged by the application of any of the analog microphone phantom powers specified by IEC 61938 including common 48V phantom (P48). The techniques described by this standard may be applied to portable AES3 output devices other than microphones, however, AES42 only covers microphones.

Optionally a modulation from +10 to +12V (resulting in a peak current of 300mA) may be applied to the DPP for remote control purposes. This modulated signal thus travels in common mode from the AES42 input back to the AES42 microphone over the same cable that is carrying the AES3 audio data from the microphone to the AES42 input. Because it is sent in common mode, the data rate must be far slower than that of AES3 to avoid interference. If the AES3 frame rate (FR) is 44.1kHz or 48kHz, the bit rate of the remote control signal is FR/64 bits per second (bit/s). For a FR of 88.2kHz or 96kHz the remote control bit rate is FR/128bit/s. For a FR of 176.4kHz or 192kHz the remote control bit rate is FR/256 bit/s. As a result, the remote control bit rate is 750bit/s if the AES3 FR is 48kHz, 96kHz, or 192kHz, and 689.06bit/s if the FR is 44.1kHz, 88.2kHz, or 176.4kHz.

The remote control signals are sent as required, except if used for synchronization, in which case they will be sent on a regular basis of not less than six times per second.

The following information is based on AES42-2010. It is always advisable to obtain the latest revision of the standard.

42.5.1 Synchronization

There are two primary possible modes of operation for a microphone meeting the AES42 Standard, Fig. 42-12.

Mode 1 allows the microphone to free run at a rate determined by its own internal clock. No attempt is made to lock the microphone's clock rate to an external clock, and if such a lock is desired, sample rate conversion must be performed external to the microphone. This technique is the simplest way for an AES42 microphone to operate, and does not require the use of the optional remote control signal.

Mode 2 uses the remote control signal to send data back to the microphone that allows its sampling rate to be varied, and phase locked to an external reference. The mode 2 microphone (or other AES42 device) contains a voltage controlled crystal oscillator (VCXO), which has its frequency controlled by a digital to analog converter (DAC). The DAC receives control information via the remote control signal from the AES42 receiving device. The receiving device compares the current sample rate of the microphone to the external reference and uses a phase locked loop (PLL) to generate a correction signal, which is sent back to the microphone. This results in the sampling rate of the microphone becoming frequency and phase matched to the reference signal. If multiple microphones or other AES42 mode 2 sources are locked to the same reference, this has the additional advantage of providing a consistent and near zero phase relationship between the sampling times of the various sources. When multiple microphones sample correlated signals, for example, in stereo or multichannel recording techniques, this results in stable imaging.

If the receiver does not support mode 2 operation, the mode 2 microphone automatically reverts to mode 1 operation.

42.5.2 Microphone ID and Status Flags

AES42 defines the use of the user data channel in AES3 to optionally allow the microphone to identify itself and send back status information. Imagine the benefits in a complex setup of not having to worry which input a given mic is plugged into. The receiving device could use the microphone ID information to automatically route the microphone to the correct system input, no matter to which physical input it was connected.

42.5.3 Remote Control

AES42 defines three possible sets of remote control instructions, simple, extended, and manufacturer specific. If a device supports the extended instruction set, it must also support the simple instruction set. If a device supports at least the simple instruction set, it must have predetermined default settings it enters if no instructions are received on power up. If a device has switches on it, those will have priority over received instructions.

42.5.4 Simple Instruction Set

The simple instruction set is sent as a 2 byte signal with a minimum of 1 byte break between commands sent. Each byte is sent MSB first.

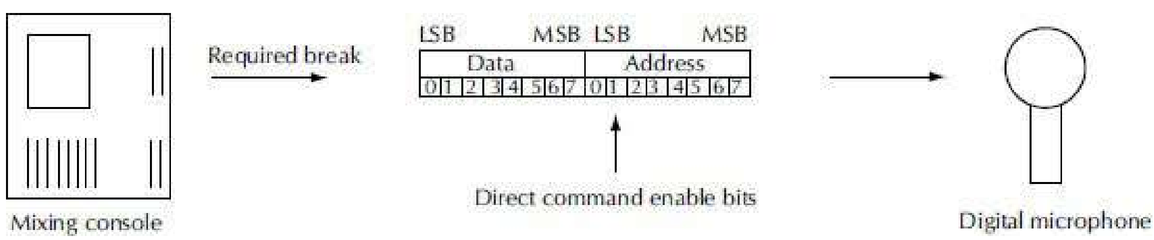


Figure 42-12. AES42 data format for the simple instruction set.

Time flows from right to left in this diagram. Note that the MSB of each byte is sent first. After each 2 byte instruction there is a required break of 1 byte's worth of time. Addressing for the simple instruction set is contained in bits 0–2 of the address byte.

42.5.5 *Direct Commands*

Table 42-2 shows the direct commands.

Table 42-2. Direct Commands

Direct Command Address Byte

Bits 0–2	Direct Command Enable bits	
Bit	0 1 2	
State	0 0 0	Identifies this command using the Extended Command set. See Extended Instruction Commands below.
	1 0 0	Direct Command 1 (low-cut filter, directivity control, preattenuation).
	0 1 0	Direct Command 2 (mute, limiter, signal gain).
	0 0 1	Direct Command 3 (synchronization).
	All other possible states of bits 0–2 are reserved and are not to be used unless defined by the AES in the future.	
Bits 3–7	Optional synchronization control word extension.	
Bit	3 4 5 6 7	
State	0 0 0 0 0	Default if synchronization control word extension not used.
	x x x x x	If the optional synchronization control word extension is used, bit 7 will be the MSB and bit 3 the LSB of the extension.

Direct Command 1 Data Byte

Bits 0–1	Low-Cut Filter	
Bit	0 1	
State	0 0	No filter, default.
	1 0	40 Hz (−3 dB) Low-cut filter.
	0 1	80 Hz (−3 dB) Low-cut filter.
	1 1	160 Hz (−3 dB) Low-cut filter.
Bits 2–5	Directivity	
Bit	2 3 4 5	
State	0 0 0 0	Manufacturer defined directivity, default.
	1 0 0 0	Omnidirectional.
	0 1 0 0	
	To	Increasing directivities.
	0 0 1 0	

	1 0 1 0	Subcardioid.
	0 1 1 0	Increasing directivities
	1 1 1 0	Through this state.
	0 0 0 1	Cardioid.
	1 0 0 1	Increasing directivity.
	0 1 0 1	Supercardioid.
	1 1 0 1	Hypercardioid.
	0 0 1 1	
	To	Increasing directivities.
	0 1 1 1	
	1 1 1 1	Figure of eight.
Bits 6–7	Preattenuation	
Bit	6 7	
State	0 0	No attenuation, default.
	1 0	6 dB(minimum).
	0 1	12 dB.
	1 1	18 dB (maximum).

Direct Command 2 Data Byte

Bit 0	Mute	
	0	Mute off, default.
	1	Mute on.
Bit 1	Peak Limiter	
	0	Peak Limiter disabled, default.
	1	Peak Limiter enabled.
Bits 2–7	Gain	
Bit	2 3 4 5 6 7	
State	0 0 0 0 0 0	0 dB gain, default.
	10 0 0 0 0	+1 dB gain.
	x x x x x x	Increasing 1 dB per count.
	1 1 1 1 1 1	+63 dB gain.

Direct Command 3 Data Byte

Bits 0–7	Synchronization	
Bit	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	Maximum negative tuning of VCXO.
	0 0 0 0 0 0 0 1	Center frequency of VCXO.
	1 1 1 1 1 1 1 1	Maximum positive tuning of VCXO.

Synchronization control word can be expanded from 8 to 13 bits by using the 5 extended address bits as LSB extended data. See AES42-2010 Annex C for details.

Extended Instruction Commands

Note that not all possible commands are defined.

Extended Instruction Address Byte

Bits 0–2 Direct Command Enable Bits.

Bit	0 1 2	
State	0 0 0	Identifies this command uses the Extended Command set.
	1 0 0	Direct Command 1. See Direct Commands above.
	0 1 0	Direct Command 2. See Direct Commands above.
	0 0 1	Direct Command 3. See Direct Commands above.

All other possible states of bits 0–2 are reserved and are not to be used unless defined by the AES in the future.

Bits 3–7 Extended address bits.

Bit	3 4 5 6 7	
State	0 0 0 0 0	Command 0, default.
	1 0 0 0 0	Command 4.
	0 1 0 0 0	Command 5.
	1 1 0 0 0	Command 6.
	0 0 1 0 0	Command 7.
	x x x x x	Commands follow in sequence.
	1 1 1 1 1	Command 34.

Extended Command 0 data Byte

Bits 0 – 7 System Commands

Bit 01234567

State	00000000	Reset to AES42 default settings (Note 1).
	1 0 0 0 0 0 0 0	Reset to user defined settings (Note 1 & 2).
	0 1 0 0 0 0 0 0	Store user defined settings.
	1 1 0 0 0 0 0 0	
	To	Reserved.
	0 1 1 1 1 1 1 1	
	1 1 1 1 1 1 1 1	No operation (default).

NOTE 1 The AES42 interface must send its supported control commands immediately after powering the microphone or sending a reset command. This ensures both interface and microphone have equivalent parameter settings.

NOTE 2 If supported, user setting are used after powering the microphone instead of AES42 default settings.

Extended Command 4 data Byte

Bits 0–1	Light Control.	
Bit	0 1	
State	0 0	No light, default.
	1 0	Light 1 on/light 2 off.
	0 1	Light 2 on/light 1 off.
	1 1	Both lights.

Bits 2–3 Test Signal.

Bit	2 3	
State	0 0	No test signal, default.
	1 0	Test signal 1 (under consideration).
	0 1	Test signal 2 (under consideration).
	1 1	Test signal 3 (under consideration).
Bit 4	ADC Calibration.	
	0	No calibration, default.
	1	Calibrate ADC
Bit 5	Reset	
	0	No reset, default.
	1	Reset.
Bits 6–7	Microphone Status Data Page Request.	
Bit	6 7	
State	0 0	Page 0, default.
	1 0	Page 1.
	0 1	Page 2.
	1 1	Page 3.

Extended Command 5 Data Byte

Bits 0–3	Dither and noise shaping.	
Bit	0 1 2 3	
State	0 0 0 0	No dither or noise shaping, default.
	x x x x	All other states are under consideration.
Bits 4–6	Sampling frequencies.	
Bit	4 5 6	
State	0 0 0	44.1 kHz, default.
	1 0 0	48 kHz.
	0 1 0	88.2 kHz multiple = 2.
	1 1 0	96 kHz multiple = 2.
	0 0 1	176.4 kHz multiple = 4.
	1 0 1	192 kHz multiple = 4.
	011	352.8 kHz multiple = 8.
	111	384 kHz multiple = 8.
Bit 7	x	Reserved.

Extended Command 6 Data Byte

Bits 0–6	XY balance, (Notes 1, 2).	
Bit	0 1 2 3 4 5 6	
State	0 0 0 0 0 0 0	Left 0.5, Right 0.5, Center, default.
	1 1 1 1 1 1 0	Left 1.0, Right 0.0, Left (A) channel only.
	0 0 0 0 0 0 1	Left 0.0, Right 1.0, Right (B) channel only.
Bits 0–v6	MS width (Notes 1, 2).	
Bit	0 1 2 3 4 5 6	
State	0 0 0 0 0 0 0	Mid 0.5, Side 0.5, Stereo, default.
	1 1 1 1 1 1 0	Mid 1.0, Side 0.0, Mono.
	0 0 0 0 0 0 1	Mid 0.0, Side 1.0, Difference only.
Bit 7	XY or MS Select (Note 2).	
	0	XY stereo, default.
	1	MS stereo.

Note 1. Signed two's complement notation is used to encode the channel weights. Bit 6 is the sign extension = -2^0, bit 5 = 2^{-1} . . . Bit 0 = 2^{-6}.

Note 2. If XY is selected, then AES3 Channel 1 carries the Left audio channel, AES3 Channel 2 carries the Right audio channel, and bits 0–6 control the XY balance. If MS is selected, then AES3 Channel 1 carries the Mid or sum signal, AES3 Channel 2 carries the Side or difference signal, and bits 0–6 control the MS width.

Extended Command 7 Data Byte

Bits 0–7	Equalization Curve Select.	
Bit	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	No equalization, default.
	x x x x x x x x	All other states, manufacturer specific equalization.

Extended Command 8 Data Byte

Bit 0	Polarity Select	
Bit	0	
State	0	No polarity reverse (default).
	1	Polarity reverse.
Bit 1	Stereo/Mono	
Bit	1	
State	0	Stereo (default for stereo microphones).
	1	Mono (both channels transmit the same signal).

NOTE Default setting of bit is stereo. Mono mics ignore this bit.

Bits 2 - 7	x x x x x x	Reserved

Extended Command 9

Bites 0 – 7 Periodic transmission control

Bit	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	No change allowed (default).
	1 x x x x x x x	Annex D, Page 0, Byte 20 to 22.
	x 1 x x x x x x	Annex D, Page 1, Byte 13 to 20.
	x x 1 x x x x x	
	To	Reserved
	x x x x x x x 1	

NOTE Some microphone status information is transmitted periodically using bytes with alternate meanings. If the receiver supports decoding of these special bytes, command 9 shall be sent providing a handshake mechanism for safe receiving and decoding. Every data bit of command 9 corresponds to a defined data byte block described in Annex D, pages 0 to 3. A logical 1 allows the microphone to change the corresponding data bytes.

Extended Commands 10 through 27 Reserved.

Extended Command 28

Bits 0 - 3 Peak Limiter Threshold Setting.

Bit	0 1 2 3	
State	0 0 0 0	0 dB FS (default).
	1 0 0 0	−1 dB FS.
	0 1 0 0	−2 dB FS.
	1 1 0 0	−3 dB FS.
	0 0 1 0	−4 dB FS.
	1 0 1 0	−5 dB FS.

	0 1 1 0	−6 dB FS.
	1 1 1 0	−7 dB FS.
	0 0 0 1	−8 dB FS.
	1 0 0 1	−9 dB FS.
	0 1 0 1	−10 dB FS.
	1 1 0 1	−11 dB FS.
	0 0 1 1	−12 dB FS.
	1 0 1 1	−13 dB FS.
	0 1 1 1	−14 dB FS.
	1 1 1 1	−15 dB FS.
Bits 4–7		Reserved.

Extended Command 29

Bits 0–3 Compressor/limiter attack time.

Bit	0 1 2 3	
State	0 0 0 0	0 ms (default).
	1 0 0 0	01 ms.
	0 1 0 0	0.3 ms.
	1 1 0 0	1 ms.
	0 0 1 0	3 ms.
	1 0 1 0	10 ms.
	0 1 1 0	30 ms.
	1 1 1 0	100 ms.
	0 0 0 1	Auto attack mode 1.
	1 0 0 1	Auto attack mode 2.
	0 1 0 1	Auto attach mode 3.
	1 1 0 1	Auto attack mode 4.
	0 0 1 1	Auto attack mode 5.
	1 0 1 1	Auto attack mode 6.
	0 1 1 1	Auto attack mode 7.
	1 1 1 1	Auto attack mode 8.

Bits 4 - 7 Compressor/limiter recovery time

Bit	4 5 6 7	
State	0 0 0 0	Manufacturer specific recovery time (default).
	1 0 0 0	50ms.
	0 1 0 0	0.1s.
	1 1 0 0	0.2s.
	0 0 1 0	0.5s.
	1 0 1 0	1s.
	0 1 1 0	2s.
	1 1 1 0	5s.
	0 0 0 1	Auto recovery mode 1.
	1 0 0 1	Auto recovery mode 2.
	0 1 0 1	Auto recovery mode 3.
	1 1 0 1	Auto recovery mode 4.
	0 0 1 1	Auto recovery mode 5.
	1 0 1 1	Auto recovery mode 6.
	0 1 1 1	Auto recovery mode 7.

	1 1 1 1	Auto recovery mode 8.

NOTE The terminology for attack and recovery times are defined in IEC 268-8. If the definitions for the attack and recovery times are not compliant with this standard they shall be defined in the manufacturer's specifications.

Extended Command 30

Bits 0–2	Compressor/limiter ratio	
Bit	0 1 2	
State	0 0 0	1.2:1 (default).
	1 0 0	1.5:1.
	0 1 0	2:1.
	1 1 0	3:1.
	0 0 1	4:1.
	1 0 1	6:1.
	0 1 1	8:1.
	1 1 1	∞:1 (limiter).
Bit 3	Reserved	
Bits 4–5	Side chain frequency response	
Bit	4 5	
State	0 0	Flat (default).
	1 0	1 kHz high-pass filter.
	0 1	2 kHz high-pass filter.
	1 1	4 kHz high-pass filter.
Bits 6–7	Reserved	

Extended Command 31

Bits 0–5	Compressor/limiter threshold	
Bit	0 1 2 3 4 5	
State	0 0 0 0 0 0	0 dB FS (default).
	1 0 0 0 0 0	−1 dB FS.
	0 1 0 0 0 0	
	To	Increasing at 1 dB per count.
	0 1 1 1 1 1	
	1 1 1 1 1 1	−63 dB FS
Bit 6	Reserved	
Bit 7	Compressor/limiter enabled	

Bit	7	
State	0	Compressor/limiter disabled (default).
	1	Compressor/limiter enabled.

Extended Command 32

Bits 0–3		Brightness light 1.
Bit	0 1 2 3	
State	0 0 0 0	Brightness 1 (dark).
	1 0 0 0	Brightness 2.
	0 1 0 0	
	To	Continuing increments.
	1 1 1 0	
	0 0 0 1	default.
	1 0 0 1	
	To	Continuing increments.
	0 1 1 1	
	1 1 1 1	Brightness 16.
Bits 4–7	Brightness light 2.	
Bit	4 5 6 7	
State	0 0 0 0	Brightness 1 (dark).
	1 0 0 0	Brightness 2.
	0 1 0 0	
	To	Continuing increments.
	1 1 1 0	
	0 0 0 1	default.
	1 0 0 1	
	To	Continuing increments.
	0 1 1 1	
	1 1 1 1	Brightness 16.

Extended Command 33 Manufacturer Specific Instruction begin.

Extended Command 34 Manufacturer Specific Instruction end.

Manufacturer Specific Instructions Are Under Consideration.

42.5.6 Remote Control Pulse Structure

There are two modes of transmitting remote control pulses: standard and optional fast mode. A feature bit (Fast DPP Remote Data see Annex D) is set if the fast mode is supported by the microphone. In the following, fast mode values are in parenthesis.

The remote control pulses are added to the DPP voltage and have a peak to peak amplitude of 2±0.2V. They carry information in the form of pulse width modulation.

For AES3 frame rates (FR) of 48kHz or multiples thereof the remote control data rate is 750 (9600) bit/s, while for FR of 44.1kHz or multiples thereof the remote control data rate is 689 (8820)bit/s. The transmission time T per bit corresponds to the inverted bit rate –1.33ms (104ps) to 1.45ms (113ps).

A logical 1 is represented by a pulse width of 7/8 (2/3) × T, and must follow the preceding pulse at an interval of 1/8 (1/3) × T A logical 0 is represented by a pulse width of 1/8 (1/3) × T, and must follow the preceding pulse at an interval of 7/8 (2/3) × T. Thus in both cases the total time used by a bit is T, a byte is 8 × T, and the combination of the command and data bytes is 16 × T.

It is possible that in the future an extended command byte may be added preceding the existing command and data bytes. In any case the entire sequence of extended command byte (if defined in the future), command byte, and data byte is sent with no interruptions in the flow of pulses.

The minimum interval between the end of one command and data bytes block and the beginning of the next is 8 × T or a 1 byte interval. The end of a command-data block is defined by a minimum break time of 4 × T which allows detection of the end of the command and data bytes and for the data to be latched into the microphone.

The command byte is transmitted first, immediately followed by the data byte. Within each byte the MSB is transmitted first and the LSB last.

The rise and fall times of the pulses (measured from the 10% and

90% amplitude points) is to be 10µs ±5µs, independent of the permitted capacitive load conditions (C_{load} =0–170nF including the cable capacitance), Fig. 42-13.

42.5.7 Synchronization

A mode 2 AES42 transmitter contains a VCXO and a DAC that set its operating frequency. The corresponding PLL resides in the AES42 receiver. The receiver sends a regular stream of control voltage commands to the microphone using Direct Command 3 of the simple instruction set. The commands are repeated not less than once every ⅙ s, and can have 8 to 13 bits of resolution. The ADC and DAC must have an accuracy of ±½LSB, and be monotonic.

If on power up a mode 2 AES42 transmitter does not see synchronization commands sent to it by the receiver, it will run in mode 1 at its default sampling rate or at the rate specified by the extended instruction set if supported. If while running a mode 2 AES42 transmitter stops receiving synchronization commands, it should hold the last value of control voltage sent to it until synchronization commands are restored.

Mode 2 transmitters identify themselves to mode 2 receivers by means of a command that is part of the user data bits in the AES3 data stream. When a mode 2 capable receiver sees this signal it switches to mode 2 operation.

AES42 specifies the mode 2 AES42 receiver characteristics for 48kHz or 44.1kHz operation. Since there is a linear relationship between comparison frequency and loop gain, operation at higher frequencies will require either frequency division down to 48kHz or 44.1kHz, or a corresponding reduction of the loop gain.

The phase comparator is a frequency-phase (zero-degree) type, and the PLL has a proportional, integrating, differentiating (PID) characteristic. The proportional constant *Kp* is 1LSB at 163ns time error (2.8° at 48kHz). The integration time constant *Ki* is 1LSB/s at 163ns time error. The differential constant Kd is 1LSB at 163 ns/s change of time error. The differential signal maximum gain is 8LSB at 163ns time error (fast change). The master reference clock must have an accuracy of ±50ppm or better.

The AES42 mode 2 transmitter must have a VCXO basic accuracy of ±50ppm, a minimum tuning range of ±60ppm + basic accuracy, a maximum tuning range of ±200ppm, and a tuning slope that is positive with f_{max} for control data = 0xFF. The control voltage low pass filter has a dc gain of unity, a stage 1 filter that is first order with a corner frequency of 68 mHz (0.068Hz) and maximum attenuation for frequencies greater than 10Hz of 24dB constant, and a stage 2 filter that is first order with a corner frequency of 12Hz. Means may be used to raise the corner frequencies when the rate of tuning change is great. This will allow faster lockup on power on.

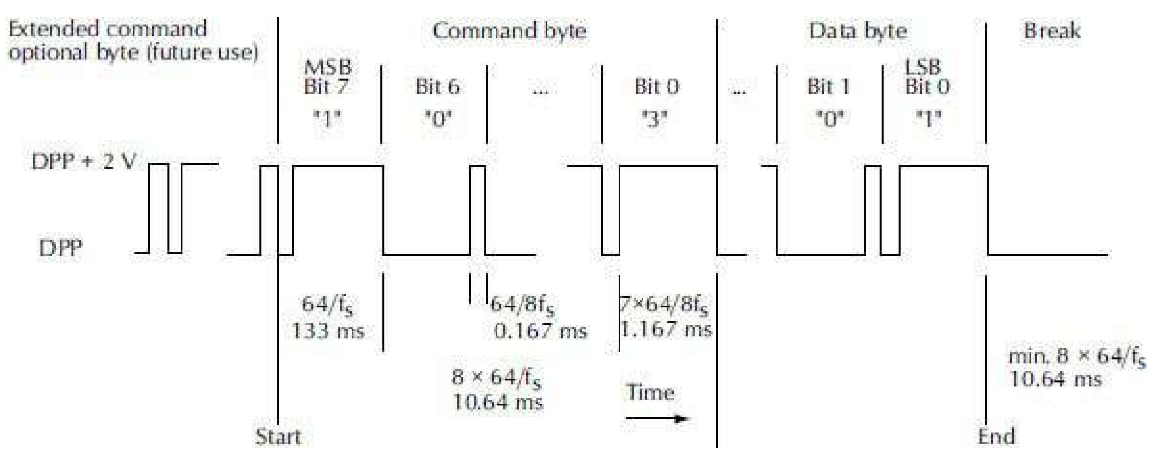


Figure 42-13. AES42 command and data byte bit structure at a 48kHz frame rate (FR).

AES42 provides schematics showing how such a mode 2 control system might be implemented.

42.5.8 Microphone Identification and Status Flags

AES42 compliant transmitters may send status information to the receiver using the user data bit as defined in AES3. The channel status block start preamble is used to identify the start of blocks of 192 bits of user data. Each subframe contains user data. This allows different information to be sent in each subframe that is associated with that subframe.

In monophonic microphones where the audio data is repeated in both subframes, the user data must also be repeated.

Microphone status data is sent LSB first in pages of 192 bits each. The pages are organized into 24 bytes. Byte 0 of all pages always contains the same data, including a page identifier, and time critical bits. This assures the delivery of the time critical bits no matter which page is being sent. Page 0 is sent continuously with each additional page being sent at least once per second. The receiver may request additional pages using the page request command in Extended Command data byte 4.

In order for the receiver to properly interpret the user data, the transmitter must set byte 1 bits 4–7 of the AES3 channel status data to 0 0 0 1. This indicates the user data bit is in use, and it is organized into 192 bit blocks starting with the AES3 subframe Z preamble.

Organization

All status bytes are sent LSB first, Table 42-3.

Table 42-3. Status Data Page

Status Data Page 0		
Status Byte 0—Starts All Status Data Pages		
Bits 0–2	Reserved.	
Bit	0 1 2	
State	0 0 0	Reserved. Must always be set to 0 0 0.
Bit 3	Mute	
State	0	Not muted, default.
	1	Muted.
Bit 4	Overload.	
State	0	No overload, default.
	1	Overload.
Bit 5	Limiter.	
State	0	Limiter not active, default.
	1	Limiter active.
Bits 6–7	Page Identifier.	
Bit	6 7	
State	0 0	Status Page 0 Mic Status.
	1 0	Status Page 1 Mic identification.

	0 1	Status Page 2 Mic revision.
	1 1	Status Page 3, reserved.

Status Data Page 0 Byte 1—Microphone Configuration Echo

Bits 0–1 Low-Cut Filter Status Echo.

Bit	0 1	
State	0 0	No filter, default.
	1 0	40 Hz (−3 dB) Low-cut filter.
	0 1	80 Hz (−3 dB) Low-cut filter.
	1 1	160 Hz Low-cut filter.

Bits 2–5 Directivity Status Echo.

Bit	2 3 4 5	
State	0 0 0 0	Manufacturer defined directivity, default.
	1 0 0 0	Omnidirectional.
	0 1 0 0	
	To	Increasing directivities.
	0 0 1 0	
	1 0 1 0	Subcardioid.
	0 1 1 0	
	To	Increasing directivities.
	1 1 1 0	
	0 0 0 1	Cardioid.
	1 0 0 1	Increasing directivity.
	0 1 0 1	Supercardioid.
	1 1 0 1	Hypercardioid.
	0 0 1 1	
	To	Increasing directivities.
	0 1 1 1	
	1 1 1 1	Figure of eight.

Bits 6–7 Preattenuation Status Echo.

Bit	6 7	
State	0 0	No attenuation, default.
	1 0	6 dB (minimum).
	0 1	12 dB.
	1 1	18 dB (maximum).

Status Data Page 0 Byte 2—Microphone Configuration Status

Bits 0–4 Mute		
Bit	0	
State	0	Mute off.
Bit	1	Mute on.
Bit 1		Peak Limiter.
Bit	1	
State	0	Peak Limiter On.
	1	Peak Limiter Off.
Bits 2 - 7		Signal Gain.
Bit	2 3 4 5 6 7	
State	0 0 0 0 0 0	0 dB (default).
	1 0 0 0 0 0	+1 dB

	0 1 0 0 0 0	
	To	Increasing at 1 dB per count.
	0 1 1 1 1 1	
	1 1 1 1 1 1	+63 dB.

Status Data Page 0 Byte 3—Microphone Remote Control Feature Indicator 1 (sound)

Bit 0	EQ Curve Selection	
State	0	Equalization curve not available, default.
	1	Equalization curve available, set using Extended Command Data Byte 7.
Bit 1	Balance-Width.	
State	0	MS width or XY balance not available, default.
	1	MS width or XY balance available, set using Extended Command Data Byte 6, bits 0–6.
Bit 2	MS-XY Switch.	
State	0	MS or XY selection not available, default.
	1	MS or XY selection available, set using Extended Command Data Byte 6, bit 7.
Bit 3	Limiter.	
State	0	Limiter not available, default.
	1	Limiter available, set using Direct Command Data Byte 2, bit 1.
Bit 4	Gain Control.	
State	0	Signal gain settings not available, default.
	1	Signal gain settings available, set using Direct Command Data Byte 2, bits 2–7.

Bit 5	Low Cut Filter.	
State	0	Low cut filter settings not available, default.
	1	Low cut filter settings available, set using Direct Command Data Byte 1, bits 0–1.
Bit 6	Pattern Control.	
State	0	Directivity pattern settings not available, default.
	1	Directivity pattern settings available, set using Direct Command Data Byte 1, bits 2–5.
Bit 7	Attenuation.	
State	0	Attenuation settings not available, default.
	1	Attenuation settings available, set using Direct Command Data Byte 1, bits 6–7.

Status Data Page 0 Byte 4—Microphone Remote Control Feature Indicator 2 (control)

Bit 0	Mode 2 Synchronization.	
State	0	External synchronization not available, default.
	1	External synchronization available.
Bit 1	Dither-Noise Shaping.	
State	0	Dither and noise shaping not available, default.

	1	Dither and noise shaping available, set using Extended Command Data Byte 5, bits 0–3.
Bit 2	Multiple Sampling Frequency.	
State	0	Multiple sampling frequencies not available, default.
	1	Multiple sampling frequencies available, set using Extended Command Data Byte 5, bits 4–6.
Bit 3	Light Control.	
State	0	Light control selection not available, default.
	1	Light control selection available, set using Extended Command Data Byte 4, bits 0–1.
Bit 4	Test Signal.	
State	0	Test signal selection not available, default.
	1	Test signal selection available, set using Extended Command Data Byte 4, bits 2–3.
Bit 5	ADC Calibrate.	
State	0	ADC calibration function not available, default.
	1	ADC calibration function available, set using Extended Command Data Byte 4, bit 4.
Bit 6	Reset.	
State	0	Reset function not available, default.
	1	Reset function available, set using Extended Command Data Byte 4, bit 5.

Bit 7	Mute.	
State	0	Mute selection not available, default.
	1	Mute selection available, set using Direct Command Data Byte 2, bit 0.

Status Data Page 0 Byte 5—Microphone Remote Control Feature Indicator

Polarity		
State	0	Polarity selection not available.
	1	Polarity selection available.
Bit 1	Compressor/limiter.	
State	0	Compressor/limiter not available.
	1	Compressor/limiter available.
Bit 2	Stereo.	
State	0	Stereo setting not available.
	1	Stereo setting available.
Bit 3	Peak Limiter Threshold	
State	0	Peak Limiter Threshold setting not available.
	1	Peak Limiter Threshold setting available.
Bit 4	Reserved, bit set to 0	
Bit 5	Microphone Configuration Status 3	
State	0	Microphone Configuration Status 3 feed-back not supported.

	1	Microphone Configuration Status 3 feedback supported.
Bit 6	Reserved, bit set to 0.	
Bit 7	Fast DPP Remote Data	
State	0	Fast remote data not supported.
	1	Fast remote data supported.

Status Data Page 0 Byte 6—Wireless Microphone Status Flags

Bits 0–4	Reserved.	
Bit	0 1 2 3 4	
State	0 0 0 0 0	Reserved, must always be set to 0 0 0 0 0.
Bit 5	Squelch.	
	0	Receiver squelch inactive, default.
	1	Receiver squelch active.
Bit 6	Link Loss.	
	0	RF link is operating, default.
	1	RF link not operating.
Bit 7	Low Battery.	
	0	No low battery condition, default.
	1	Low battery condition.

Note 1. This byte is only used for wireless microphones, wired microphones should set all bits to 0.

Status Data Page 0 Byte 7—Wireless Microphone Battery Status

Bits 0–1 Reserved.

Bit	0 1	
State	0 0	Reserved, must always be set to 0 0.

Bits 2–5 Battery Charge Proportion.

Bit	2 3 4 5	
State	0 0 0 0	100%, default.
	1 0 0 0	90%.
	0 1 0 0	80%.
	1 1 0 0	70%.
	0 0 1 0	60%.
	1 0 1 0	50%.
	0 1 1 0	40%.
	1 1 1 0	30%.
	0 0 0 1	20%.
	1 0 0 1	10%.
	0 1 0 1	0%.
	x x x x	All other states reserved.

Bits 6–7 Battery Type.

Bit	0 1	
State	0 0	Not indicated, default.
	1 0	Battery type is primary cell.
	0 1	Battery type is rechargeable.
	1 1	Reserved.

Note 1. Microphones supporting low battery indication use byte 6 bit 7. Microphones supporting battery charge indication must use both byte 7 bits 2 – 5, and byte 6 bit 7.

Status Data Page 0 Byte 8—Wireless Microphone Error Handling Flags

Bits 0–2 Reserved.

Bit	0 1 2	
State	0 0 0	Reserved, must always be set to 0 0 0.

Bits 3–4 Error Concealment

Bit	3 4	
State	0 0	Error concealment not in use, default.
	1 0	Error concealment in use.
	0 1	Reserved.
	1 1	Reserved.

Bits 5–7 FEC Capacity.

Bit	5 6 7	
State	0 0 0	FEC capacity used 0%.
	1 0 0	FEC capacity used 20%.
	0 1 0	FEC capacity used 40%.
	1 1 0	FEC capacity used 60%.
	0 0 1	FEC capacity used 80%.
	1 0 1	FEC capacity used 100%.
	0 1 1	FEC capacity overloaded.
	1 1 1	Reserved.

Status Data Page 0 Byte 9 – Supported Sampling Rates

Bits 0–7 Sampling Rates (multiple flags can be set: x = don't care).

Bits	0 1 2 3 4 5 6 7	
State	1 x x x x x x x	44.1kHz supported.
	x 1 x x x x x x	48kHz supported.
	x x 1 x x x x x	88.2kHz supported.
	x x x 1 x x x x	96kHz supported.
	x x x x 1 x x x	176.4kHz supported.
	x x x x x 1 x x	192kHz supported.
	x x x x x x 1 x	352.8kHz supported.
	x x x x x x x 1	384kHz supported.

Status Data Page 0 Byte 10 – Microphone Switch Monitoring

Bits 0–4 Reserved, bits set to 0.

Bits 5–6 Call Button.

Bits	5 6	
State	0 0	No button Pressed.
	1 0	Button 1 pressed.
	0 1	Button 2 pressed.
	1 1	Buttons 1 & 2 pressed.

Bit 7 Remote off.

Bit	7	
State	0	Remote parameter setting enabled.
	1	Remote parameter setting disabled.

Status Data Page 0 Byte 11 – Microphone error status

Bits 0–7

Bits	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	No error.

	1 x x x x x x x	Microphone capsule error.
	x 1 x x x x x x	
	To	Reserved.
	x x x x x x x 1	

Status Data Page 0 Byte 12

Status Data Page 0 Bytes 13–19—Reserved

All bits of Page 0 Bytes 13-19 are reserved and should be set to 0.

Status Data Page 0 Byte 20-Microphone Configuration status 3-periodic transmission

Bits 0–7 Extended command counter.

Bits	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	
	To	Meaning of corresponding data in byte 22 – See Annex A.2.
	1 1 1 1 1 1 1 1	

Status Data Page 0 Byte 21-Microphone Configuration status 3-periodic transmission

Bit 0 Extended Command.

Bits	0	
State	0	Extended command byte not used.
	1	Extended command byte valid.

Bits 1–2 Reserved.

Bits 3–7 Command Counter.

Bits	3 4 5 6 7	
State	0 0 0 0 0	
	To	Meaning of corresponding data in byte 22 – See Annex A.2.
	1 1 1 1 1	

Status Data Page 0 Byte 22-Microphone Configuration status 3-periodic transmission

Bits 0–7	Status Data.	
Bits	0 1 2 3 4 5 6 7	
State	x x x x x x x x	Microphone configuration status—See Annex A.2.

Status Data Page 0 Byte 23-Compressor/limiter gain reduction

NOTE Details not specified by AES42-2010.

Status Data Page 1 Bytes 1–12—Manufacturer Identification

Manufacturer identification information should be sent in 7 bit ASCII form using only printable characters in the range of 00, and 20 to 7E Hex, and starting in byte 1. The use of nonprintable characters in the range of 01 to 1F Hex is not allowed. Each byte has a 0 of reserved usage in bit 7, followed by the MSB of the ASCII code in bit 6, through the LSB in bit 0. This allows 12 characters for manufacturer identification. Fill any unused bytes with zeros.

Status Data Page 1 Bytes 13–20—Microphone Model Identification

Manufacturer identification information should be sent in 7 bit ASCII form using only printable characters in the range of 00, and 20 to 7E Hex, and starting in byte 1. The use of nonprintable characters in the range of 01 to 1F Hex is not allowed. Each byte starts with the LSB of the ASCII code in bit 0, through the MSB in bit 6, and has a 0 of reserved usage in bit 7. This allows 12 characters for manufacturer identification. Fill any unused bytes with zeros.

Status Data Page 1 Bytes 13–20—Device Information-periodic transmission

Device information should be sent in 7 bit ASCII form using only printable characters in the range of 00, and 20 to 7E Hex, and starting in byte 13. The use of nonprintable characters in the range of 01 to 1F Hex is not allowed. Each byte starts with the LSB of the ASCII code in bit 0, through the MSB in bit 6, and has a 0 of reserved usage in bit 7. This allows 8 characters for device information. Fill any unused bytes with zeros.

Status Data page 1 Byte 21-Device Identification (meaning of bytes 13 – 20)-periodic transmission

Bits 0–7

Bits	0 1 2 3 4 5 6 7	
State	0 0 0 0 0 0 0 0	Microphone model (default).
	1 0 0 0 0 0 0 0	Microphone head.
	0 1 0 0 0 0 0 0	Wireless transmitter.
	1 1 0 0 0 0 0 0	Wireless receiver.
	0 0 1 0 0 0 0 0	Wireless channel number.
	1 0 1 0 0 0 0 0	Wireless channel name.
	0 1 1 0 0 0 0 0	
	To	Reserved.
	1 1 1 1 1 1 1 1	

Status Data Page 1 Bytes 22—Reserved

All bits of Page 1 Byte 22 are reserved and should be set to 0.

Status Data Page 1 Byte 23-Compressor/limiter gain reduction

NOTE Details not specified by AES42-2010.

Status Data Page 2 Bytes 1–8—Microphone Serial Number

Microphone serial number information should be sent in 7-bit ASCII form using only printable characters in the range of 00, and 20 to 7E Hex, and starting in byte 1. The use of nonprintable characters in the range of 01 to 1F Hex is not allowed. Each byte starts with the LSB of the ASCII code in bit 0, through the MSB in bit 6, and has a 0 of reserved usage in bit 7. This allows 12 characters for microphone serial numbers. Fill any unused bytes with zeros.

Status Data Page 2 Byte 9—Microphone Hardware Revision Main Counter

The information is sent as two binary coded decimal (BCD) digits. The L-nibble or lower nibble is sent with its LSB in bit 0, and MSB in bit 3. The U-nibble or upper nibble is sent with its LSB in bit 4, and MSB in bit 7. Numbers less than 10 are coded with a leading 0 in the U-nibble.

Status Data Page 2 Byte 10—Microphone Hardware Revision Index Counter

The information is sent as two binary coded decimal (BCD) digits. The L-nibble or lower nibble is sent with its LSB in bit 0, and MSB in bit 3. The U-nibble or upper nibble is sent with its LSB in bit 4, and MSB in bit 7. Numbers less than 10 are coded with a leading 0 in the U-nibble.

Bytes 9 and 10 together represent the entire hardware revision number with a range of 00.00 to 99.99. Byte 9 is the integer portion while byte 10 is the fractional portion.

Status Data Page 2 Byte 11—Microphone Software Revision Main Counter

The information is sent as two binary coded decimal (BCD) digits. The L-nibble or lower nibble is sent with its LSB in bit 0, and MSB in bit 3. The U-nibble or upper nibble is sent with its LSB in bit 4, and MSB in bit 7. Numbers less than 10 are coded with a leading 0 in the U-nibble.

Status Data Page 2 Byte 12—Microphone Software Revision Index Counter

The information is sent as two binary coded decimal (BCD) digits. The L-nibble or lower nibble is sent with its LSB in bit 0, and MSB in bit 3. The U-nibble or upper nibble is sent with its LSB in bit 4, and MSB in bit 7. Numbers less than 10 are coded with a leading 0 in the U-nibble.

Bytes 11 and 12 together represent the entire software revision number with a range of 00.00 to 99.99. Byte 11 is the integer portion while byte 12 is the fractional portion.

Status Data Page 2 Bytes 13 – 15-Delay in samples

The information is sent as two binary coded decimal (BCD) digits per byte. The L-nibble or lower nibble is sent with its LSB in bit 0, and MSB in bit 3. The U-nibble or upper nibble is sent with its LSB in bit 4, and MSB in bit 7. Numbers less than 10 are coded with a leading 0 in the U-nibble.

Bytes 13 through 15 together represent the entire delay in samples with a range of 000000 to 999999. Byte 13 is the LSB while byte 15 is the MSB.

Status Data Page 2 Bytes 16-22—Reserved

All bits of Page 2 Bytes 16-22 are reserved and should be set to 0.

Status Data Page 2 Byte 23-Compressor/limiter gain reduction

NOTE Details not specified by AES42-2010

Status Data Page 3 Byte 0-Page identifier, peak limiter, overload, mute, & reserved

NOTE Details not specified by AES42-2010

Status Data Page 3 Byte 1

NOTE Details not specified by AES42-2010

Status Data Page 3 Bytes 2 – 22-Reserved – all bits set to 0

Status Data Page 3 Byte 23-Compressor/limiter gain reduction

NOTE Details not specified by AES42-2010

42.6 IEC 60958 Second Edition

This standard is based on three different sources, the AES3 and EBU Tech. 3250-E professional digital audio interconnection standards, and the consumer digital interface specification from Sony and Phillips (SPDIF).

The standard is broken into four parts, 60958-1 Ed, which contains general information on the digital interface; 60958-2 (unchanged from the first edition) on the serial copy management system: 60958-3 Ed2, which contains the consumer interface specific information: and 60958-4 Ed2, which contains information on the professional interface.

Since the professional interface is covered under the section on AES3 above, in this section we will only review how the consumer interface specified in 60958-3 differs from AES3.

Table 42-4 is based on Edition 3 of IEC 60958. It is always advisable to obtain the latest revision of the standard.

Table 42-4. IEC 60958 Edition 3 Standard

Channel Status General Format

Byte 0

Bit 0	0	Contents of the channel status block conform to IEC 60958-3 "consumer use" Standard.
	1	Contents of the channel status block at to the AES3 "professional use" Standard. Ignore the rest of this table. (See Note 1.)
Bit 1	0	Audio words consist of linear PCM samples.
	1	Audio words consist of something other than linear PCM samples.
Bit 2	0	Software copyrighted. (See note 2.)
	1	Software copyright not claimed.

Bits 3–5 Additional format information, depending on the state of bit 1.

If bit 1 = 0, linear PCM mode:

Bit	3 4 5	
State	0 0 0	2 audio channels not using pre-emphasis.
	1 0 0	2 audio channels using 50/15 μs pre-emphasis.
	0 1 0	Reserved (for 2 audio channels using pre-emphasis).
	1 1 0	Reserved (for 2 audio channels using pre-emphasis).
		All other possible states of bits 3–5 are reserved and are not to be used unless defined by the IEC in the future.

If bit 1 = 1, other than linear PCM mode:

Bit	3 4 5	
State	0 0 0	Default state.
		All other possible states of bits 3–5 are reserved and are not to be used unless defined by the IEC in the future.

Bits 6 – 7 Channel Status Mode.

Bit 6 7

State	0 0	Mode 0, Consumer use.
		All other possible states of bits 6–7 are reserved and are not to be used unless defined by the IEC in the future.

Note 1. Other than the use of the Channel Status block of information, the rest of the data format is identical between the AES3 "professional use" Standard and the IEC 60958-3 "consumer use" Standard. The electrical format is different, however. For these reasons it should never be assumed that a "consumer use" receiver would function correctly with a "professional use" transmitter, or vice-versa.

Note 2. If the copyright status is unknown for this application, the state of this bit may alternate at a rate between 4 Hz and 10 Hz.

Channel Status Format for Consumer Use Digital Audio

If Byte 0 bit 1, and bits 6–7 are all 0, then the following applies.

Byte 1—Category Code

Contains the category code indicating the type of equipment generating the signal. Category codes are given in the annexes to the Standard. Bit 0 contains the LSB and bit 7 the MSB. Used in conjunction with the copyright bit to control allowable copying of material.

Byte 2—Source and Channel Number.

Bits 0–3	Source Number	
Bit	0 1 2 3	
State	0 0 0 0	Don't care.
	1 0 0 0	1.
	0 1 0 0	2.
	1 1 0 0	3.
	- - - -	
	1 1 1 1	15.
Bits 4–7	Audio Channel Number.	
Bit	4 5 6 7	
State	0 0 0 0	Don't care.
	1 0 0 0	A (Left channel of stereo).
	0 1 0 0	B (Right channel of stereo).
	1 1 0 0	C.
	- - - -	
	1 1 1 1	O.

Byte 3—Sampling Frequency and Clock Accuracy.

Bits 0–3	Sampling Frequency.	
Bit	0 1 2 3	
State	0 0 0 0	44.1 kHz.
	0 1 0 0	48 kHz.
	1 1 0 0	32 kHz.
	All other possible states of bits 0–3 are reserved and are not to be used unless defined by the IEC in the future.	
Bits 4–5	Clock Accuracy.	
Bit	4 5	
State	0 0	Level II.
	1 0	Level I.
	0 1	Level III.

	1 1	Reserved.
Bits 6–7	Reserved.	
Byte 4—Word Length		
Bit 0	Maximum audio word length.	
	0	Maximum 20 bit audio words.
	1	Maximum 24 bit audio words.
Bits 1–3	Encoded audio word length.	
Bit	1 2 3	Audio word length if bit 0 indicates max. 20 bit length.
State	0 0 0	Length not indicated, default.
	1 0 0	16 bits.
	0 1 0	18 bits.
	0 0 1	19 bits.
	1 0 1	20 bits.
	0 1 1	17 bits.
	All other possible states of bits 1–3 are reserved and are not to be used unless defined by the IEC in the future.	
Bits 4–7	Reserved.	

Note 1. If the auxiliary sample bits are not used they should be set to 0.

Note 2. Generally user data is not used and all bits are set to 0.

Note 3. Channel status is identical for all channels, with the exception of the channel number if not set to all zeros (don't care).

42.6.1 Electrical and Optical Interface

Two types of interface are specified, unbalanced electrical and optical fiber.

Three levels of timing accuracy are specified and indicated in the Channel Status. Level I is the high accuracy mode, requiring a tolerance of ±50ppm. Level II is the normal accuracy mode, requiring a tolerance of ±1000ppm. Level III is the variable pitch mode. An exact frequency range is under discussion, but may be ±12.5%.

By default, receivers should be capable of locking to signals of a

Level II accuracy. If a receiver has a narrower locking range, it must be capable of locking to signals of a Level I accuracy, and must be specified as a Level I receiver. If a receiver is capable of normal operation over the Level III range, it should be specified as a Level III receiver.

42.6.1.1 Unbalanced Line

Connecting cables are unbalanced, shielded, with an impedance of 75 Ω±26.25Ω over the frequency range from 100kHz to 128 times the maximum frame rate.

The line driver has an impedance of 75 Ω±15Ω at the output terminals over the frequency range from 100kHz to 128 times the maximum frame rate. The output level is 0.5 ± 0.1Vp-p, measured across a 75±1%Ω resistor across the output terminals without any cable connected. The rise and fall times measured between the 10% and 90% amplitude points should be less than 0.4UI. The jitter gain from any reference input must be less than 3dB at all frequencies.

The receiver should be basically resistive with an impedance of 75 Ω±5% over the frequency range from 100kHz to 128 times the maximum frame rate. It should correctly interpret the data of a signal ranging from 0. 2 to 0.6Vp-p.

The connector for inputs and outputs is described in 8.6 of Table IV of IEC 60268-11, and popularly known as the RCA connector. Male plugs are used at both ends of the cable. Manufacturers should clearly mark digital inputs and outputs.

42.6.1.2 Optical Connection

This is specified in IEC 61607-1 and IEC 61607-2, and popularly known as the TOSLINK connector.

42.7 AES10 (MADI)

The AES10 Standard describes a serial multichannel audio digital interface, or MADI. The abstract says it uses an asynchronous transmission scheme, but the overall protocol is better described as isochronous. It is based on the AES3 Standard, but allows thirty two, fifty six, or sixty four channels of digital audio at a common sampling rate in the range of 32 to 96kHz, with a resolution of up to 24 bits to be sent over a single 75Ω coaxial cable at distances up to 50m. Transmission over fiber is also possible. Like the other schemes we have looked at it only allows one transmitter and one receiver.

Table 42-5 is based on AES10-2008. It is always advisable to obtain the latest revision of the standard.

MADI used the bit, block, and subframe structure of AES3 with the exception of the subframe preambles. Instead it substitutes four bits according to Table 42-5.

Table 42-5. AES10 MADI

Bit	Name	Description	Sense
0	MADI channel 0	Frame synchronization bit	1 = true
1	MADI channel active	Channel active bit	1 = true
2	MADI channel A or B	AES3 sub-frame 1 or 2	1 = subframe 2
3	MADI channel block sync	Channel block start	1 = true
4-27	AES3 audio data bits	(bit 27 is MSB)	
28	AES3 V	Validity bit	

29 AES3 U	User data bit
30 AES3 C	Status data bit
31 AES3 P	Parity bit (excludes bits Even 0-3)

MADI sends all its active channels in consecutive order starting with channel zero. Each active channel has the active channel bit set to 1. Inactive channels must have all their bits set to 0 including the channel active bit. Inactive channels must always have higher channel numbers than any active channel.

The channels are transmitted serially using a non-return-to-zero inverted (NRZI) polarity free coding. Each 4 bits of the data are turned into 5 bits before encoding.

Each 32 bit channel data is broken down into eight words of 4 bits each following this scheme:

Word	Channel Data Bits			
0	0	1	2	3
1	4	5	6	7
2	8	9	10	11
3	12	13	14	15
4	16	17	18	19
5	20	21	22	23
6	24	25	26	27
7	28	29	30	31

The 4 bit words are turned into 5 bit words as follows:

4 Bit Data	5 Bit Encoded Data
0000	11110
0001	01001
0010	10100
0011	10101
0100	01010
0101	01011

0110	01110
0111	01111
1000	10010
1001	10011
1010	10110
1011	10111
1100	11010
1101	11011
1110	11100
1111	11101

The now 5 bit words are transmitted (left to right) as follows:

Word	Channel Link Bits				
0	0	1	2	3	4
1	5	6	7	8	9
2	10	11	12	13	14
3	15	16	17	18	19
4	20	21	22	23	24
5	25	26	27	28	29
6	30	31	32	33	34
7	35	36	37	38	39

Unlike the coding used for AES3, this coding allows dc on the link.

AES10 uses a synchronization symbol, 11000 10001 transmitted left to right, which is inserted at least once per frame to ensure synchronization of the receiver and transmitter. There are no defined locations for the insertion of this symbol, but it may only be inserted at the 40 bit boundaries between data words. Enough synchronization symbols should be interleaved between channels transmitted, and after the last channel has been transmitted, to fill up the total link capacity.

42.7.1 NRZI Encoding

The 5 bit link channel data is encoded using NRZI polarity free encoding. Each high bit is converted into a transition from the bit before, while each low bit results in no transition. In other words a 1 turns into a 1 to 0 or 0 to 1 transition, while a 0 results in a static 1 or 0.

42.7.2 Sample Frequencies and Rates

MADI allows operation with sampling rates in any of three ranges:

- 32 to 48kHz, ±12.5%, 56 channels.
- 32 to 48kHz nominal, 64 channels.
- 64 to 96kHz, ±12.5%, 28 channels from the nominal frequency.

Higher sampling rates such as 192kHz require the use of multiple audio channels per sample.

Data is transmitted across the link at a constant 125 megabits per second irrespective of the number of channels in use. The data transfer rate is 100 megabits per second. The difference is due to the 4 data bit to 5 link bit encoding used.

Actual data rates used will vary. Fifty six channels at 48kHz + 12.5% or 28 channels at 96kHz + 12.5% results in a data rate of 96.768 megabits per second, while 56 channels at 32kHz – 12.5% results in a data rate of 50.176 megabits per second.

AES3 has been extended to sampling rates of 4 or 8 times, but no packing scheme has yet been proposed to handle these higher rates in AES10 (MADI).

42.7.3 Synchronization

Unlike AES3, MADI does not carry synchronization information. Therefore a separate AES3 signal must be provided to both the transmitter and receiver for synchronization purposes.

A MADI transmitter must start each frame within 5% of the sample period timing of the external reference signal. A MADI receiver must accept frames that start within 25% of the sample period timing of the external reference signal.

42.7.4 Electrical Characteristics

Either 75Ω coax or optical fiber is allowed for the transmission media. Optical interfacing is described below.

The line driver has an impedance of 75Ω ±2Ω average output level when terminated into 75Ω is 0V±0.1 V. The peak-to-peak output voltage is between 0.3 V and 0.6V into 75Ω. Rise and fall times between the 20% and 80% amplitude points must be no longer than 3ns, and no shorter than 1 ns with a relative timing difference to the average of the amplitude points of no more than ±0.5ns.

Interestingly there is no input impedance specified for the receiver, although the example schematic shows a 75Ω termination.

When a signal meeting the limits shown in Fig. 42-14 is applied to the input of a MADI receiver, it must correctly interpret it.

Cabling to interconnecting MADI devices must be 75 Ω±2Ω and have a loss of less than 0.1dB/m over the range from 1–100MHz. Cables are equipped with 75Ω BNC-type male connectors and have a 50m maximum length. Chassis connectors are female.

At the receive end of the cable the eye pattern must be no worse than what is shown in Fig. 42-14. Equalization is not allowed.

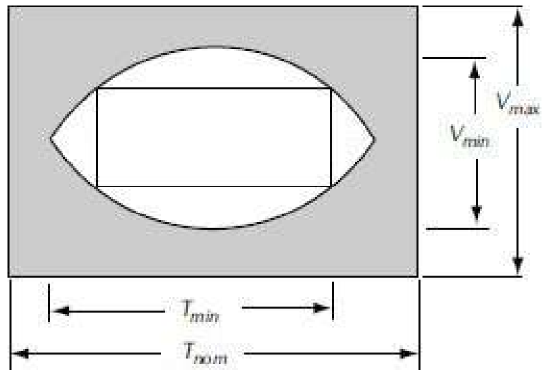


Figure 42-14. AES10 eye pattern for minimum and maximum input signals where t_{nom} = 8ns; t_{min} = 6ns; V_{max} = 0.6V; V_{min} = 0.15V. The MADI receiver must correctly interpret signals within this eye diagram as applied to its input.

The cable shield must be grounded to the chassis at the transmitter. If the shield is not grounded directly to the chassis at the receiver, it must be grounded above 30MHz. This can be achieved by capacitively coupling the shield to the chassis through a suitable low inductance capacitor of around 1.0nF.

42.7.5 Optical Interface

Graded-index fiber with a core diameter of 62.5nm, nominal cladding diameter of 125nm, and a numerical aperture of 0.275 is to be used with ST1 connectors. This will allow links of up to 2000m.

42.8 Harmon BLU link

All of the interconnect schemes we have looked at so far have been point to point, and not networked. BLU link is a good example of a simple yet useful networking scheme that was developed for a family of signal-processing devices from BSS Audio and later adopted for other devices from the Harmon family of companies.

Unlike the standards-based digital audio interconnect methods discussed so far, the protocol for BLU link is not published and is available only in products from a single manufacturer, Harmon. Nonetheless it is in wide use, and needs to be examined from an applications viewpoint.

Each BLU link component has network in and out connectors for interconnecting the devices. Wiring is done in a ring configuration. Up to 256 audio channels at 48kHz or 128 audio channels at 96kHz (both at 24 bit) can be put on the network, and each channel travels in both directions around the loop for redundancy. This allows continued operation is the loop is broken.

An output is connected to an input with a Cat 5e or higher data cable of up to 100m in length. By using special fiber converters that distance can be extended. Up to 60 devices can be part of a ring.

Even though BLU link uses Cat 5e cabling and the same “RJ-45” connectors as used for Ethernet networking, it is important to note that BLU link is not using the Ethernet protocol, just the same cable and connectors as used by Ethernet.

Referring to Fig. 42-15, there are a couple of things to note. First, the physical wiring must always be from output to input. Although signal flows in both directions over each cable, the output and input terminology shows the direction of primary signal flow.

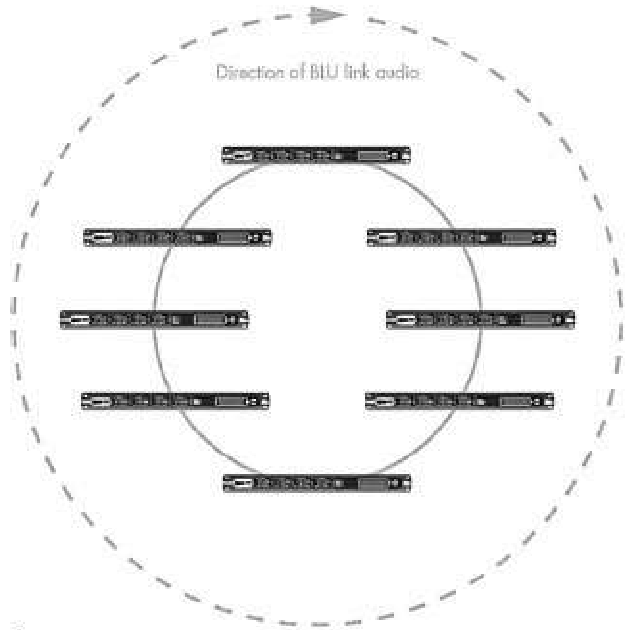


Figure 42-15. Harmon BLU link. The physical connection is a loop. Signals travel both directions around the physical loop so if the loop is broken, audio still flows.

42.9 Nexus

The Stage Tec Nexus is another proprietary digital audio networking system. It is a very high-quality system using fiber optic interconnections, providing a very flexible system. Redundant interconnections allow very high reliability. A very wide variety of input and output devices are available to insert into the Nexus frames. Both analog and digital inputs and outputs are available.

One of the most interesting aspects of the Nexus system is the ability of the programmable devices in a system to learn what they should be doing from other devices in the system. If a device fails in a running system, when the replacement device is plugged in, it determines what it should be doing from the other devices in the network. The instructions for all the devices are stored in multiple places in the system to enable this capability.

42.10 IEEE 1394 (FireWire)

FireWire is an attractive networking scheme since it provides both isochronous and asynchronous capabilities in the same network. However, this same diversity of capabilities that makes it attractive as a home networking scheme has also lead to a wide variety of incompatible protocols that are all carried over IEEE 1394.

Equipment from different manufacturers, and sometimes even different models from the same manufacturer are not compatible. James Snider, chairman of the 1394 Trade Association wrote in March 2001 that users of 1394 enabled equipment “want to know exactly what will work together and what will not, so they do not have unreasonable performance expectations.” The 1394 Trade Association is working on this issue, however, the issues are not totally resolved.

FireWire has a maximum distance limitation for a given link of 4.5m and has a bandwidth limitation of 400MBit/s for the entire network. In this sense it is similar to an Ethernet repeater-based network. FireWire’s proponents claim it can be cheaper than Ethernet, but market forces have dropped the cost of Ethernet and increased the capabilities to the point one could question if FireWire will ever catch up.

Lastly, 1394 has been mostly applied to consumer video and gaming application to this point. While it can certainly carry audio, there are few if any professional digital audio devices currently using this protocol.

AES58 defines a method of sending 32 bit generic data, including professional audio across a IEEE 1394 link.

42.11 Ethernet

Ethernet is the most common digital networking standard in the world, with over 50 million nodes installed. The huge volumes in the computer industry are constantly driving the price down and the capabilities upward. Many professional audio devices use Ethernet for control and programming, and both AMX and Crestron have embraced it for audio/video system control.

Yet with all these advantages, Ethernet per se is poorly suited to carry real-time audio since it is by nature an asynchronous system. Kevin Gross then of Peak Audio decided that there had to be a way to overcome the limitations of Ethernet for transmission of real-time information such as audio and video. His solution, called CobraNet®, was granted a patent, and has been licensed by many major professional audio companies for inclusion in their products including: Biamp, Creative Audio, Crest Audio, Crown, Digigram, EAW, LCS, Peavey, QSC, Rane, and Whirlwind.

More recent entrants into digital audio networking include Aviom, EtherSound, and Dante. All of the above make use of at least some portion of Ethernet technology.

Before we can examine these audio networking technologies, we need to get a good understanding of Ethernet.

42.11.1 Ethernet History

In 1972 Robert Metcalf and his colleagues at the Xerox Palo Alto Research Center (PARC) developed a networking system called the Alto Aloha Network to interconnect Xerox Altos computers. Metcalf changed the name to Ethernet in 1973. While the Altos is long gone, Ethernet has gone on to become the most popular networking system in the world, Fig. 42-16.

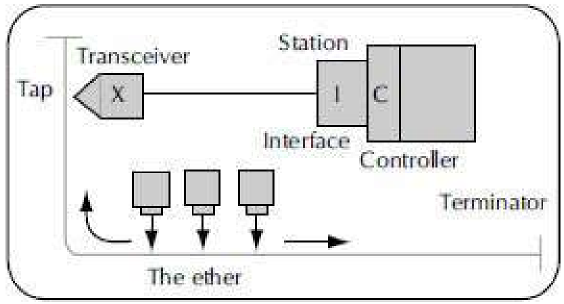


Figure 42-16. Robert Metcalf's first drawing of what became known as Ethernet.

This system had a number of key attributes. It used a shared media, in this case a common coaxial cable. This meant that the available bandwidth was shared among all the stations on the cable. If any one station transmitted, all the other stations would receive the signal. Only one station could transmit at any instant in time and get the data through uncorrupted. If more than one station attempted to transmit at the same time, this was called a collision, and resulted in garbled data.

In a shared media Ethernet network there will be collisions as a normal part of operation. As a result a mechanism for preventing collisions, detecting those it could not prevent, and recovering from them was required.

This mechanism is called *carrier-sense multiple access with collision detection* (CSMA/CD). In other words, while any station can transmit at any time (multiple access), before any station can transmit it has to make sure no other station is transmitting (carrier-sense). If no other station is transmitting, it can start to transmit. However, since it is possible for two or more stations to attempt to transmit at the same time, each transmitting station must listen for another attempted transmission at the same time (a collision). If a station transmitting detects a collision, the station transmits a bit sequence called a jam to insure all transmitting stations detect that a collision has occurred, then is silent for a random time before attempting to transmit again. Of course no retransmission can be attempted if another station is transmitting. If a second collision is detected, the delay before retransmission is attempted again increases. After a number of tries the attempt to transmit fails.

Now since the signals travel at the speed of light (approximately) down the coax cable, and since a station at one end of the cable has to be able to detect a collision with a transmission from a station at the other end of the cable, two requirements had to be imposed. First, there was a limitation on how long the cable could be. This was imposed to limit the time it could take for a transmission from a station at one end of the cable to reach the most distant station. Second, there was a minimum length imposed on the data packet transmitted. This made sure that stations that were distant from each other would have time to realize that a collision had occurred. If the cable were too long, or the packet were too short, and the stations at the ends of the cable were to both transmit at the same time, it would be possible for them to both finish transmitting before they saw the packet from the other station, and never realize that a collision had occurred, Fig. 42-17.

Destination address (6 bytes)	Source Address (6 bytes)	Protocol (2 bytes)	Payload (46–1500 bytes)	FCS (4 bytes)

Figure 42-17. Ethernet packet format.

42.11.2 Ethernet Packet Format

Every Ethernet device in the world has a globally unique media access control or MAC address. Manufacturers apply to the Institute of Electrical and Electronics Engineers (IEEE) and are assigned a block of MAC addresses for their use. Each manufacturer is responsible to make sure that each and every device it ships has a unique MAC address within that range. When it has used up 90% of its addresses it can apply for an additional block of addresses.

A MAC address is 48 bits, or 6 bytes long, which allows for 281,474,976,710,656 unique addresses. While Ethernet is extremely popular, we have not begun to run out of possible MAC addresses.

An Ethernet packet starts with the MAC address of the destination. That is followed by the MAC address of the station sending the packet. Next come 2 bytes called the EtherType number or protocol identifier, which identify the protocol used for the payload. Again the IEEE assigns these numbers. The protocol identifier assigned for CobraNet®, for example, is 8819 in hexadecimal notation.

The data payload can range from a minimum size of 46 bytes to a maximum size of 1500 bytes. The protocol identifier specifies the content of the payload and how it is to be interpreted. Data of less than 46 bytes must be extended or padded to 64 bytes, while data of more than 1500 bytes must be broken into multiple packets for transmission.

The frame check sequence (FCS) is a 4 byte long cyclic redundancy check (CRC) calculated by the transmitting station based on the contents of the rest of the Ethernet packet (destination address, source address, protocol, and data fields). The receiving station also calculates the FCS and compares it with the received FCS. If they match, the data received is assumed to have been received without corruption. There is a 99.9% probability that even a 1 bit error will be detected.

As you can see, the smallest possible Ethernet packet is 64 bytes long, and the longest is 1518 bytes long.

42.11.3 Network Diameter

The maximum allowable network diameter, Fig. 42-18, that will permit Ethernet's collision detection scheme to work is dependent on:

- The minimum packet size (64 bytes),
- The data rate (these last two together determine the time duration of the minimum size packet),
- The quality of the cable (which determines the speed of propagation down the cable).

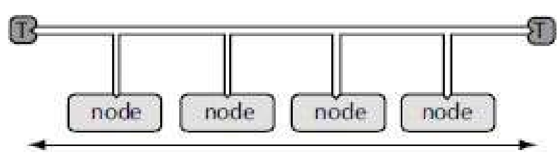


Figure 42-18. Ethernet maximum allowable network diameter.

42.11.4 Ethernet Varieties

In 1980 the IEEE standardized Ethernet as IEEE 802.3. This initial standard was based on the use of 10mm 50Ω coaxial cable. Many variations quickly appeared.

- 10Base5—This was the original Ethernet, also called thicknet or thick Ethernet because of the large diameter coaxial cable used. It ran at 10MBit/s, baseband, with a maximum segment size of 500m (hence 10Base5).

- 10Base2—Designed as a less expensive Ethernet, it was called thinnet or thin Ethernet due to the thinner RG-58 50Ω coaxial cable used. It ran at 10MBit/s, baseband, with a maximum segment length of 200m.
- 1Base2—A slower variant of thinnet. It ran at 1MBit/s, baseband, with a maximum segment length of 200m.
- 10Broad36—Very rare, this ran over a RF cable plant similar to cable TV distribution systems, and was built with cable TV distribution components.

All of these variants suffer from a common problem. Since they use a shared media that had to physically connect to every station in the network, a problem at any point along the backbone could disable the entire network. Clearly a different approach was needed to protect the shared media from disruption.

In 1990 a new technology called 10Base-T was introduced to solve these problems. Instead of the vulnerable shared media being strung all over the entire facility, it was concentrated into a box called a repeater hub, Fig. 42-19. The media was still shared, but protected. It had an allowable network diameter of 2000m, and used the same packet structure. Each station was connected to the hub with twisted pair cable, with two pairs used. One pair carried the signal from the station to the hub, while the other pair carried the signal from the hub to the station. Category 3 (Cat 3) unshielded twisted pair (UTP) was used with transformer isolation at both ends of each pair. The maximum length of a single cable run was restricted to 100m. Longer runs up to 2000m were possible using a fiber version called 10BaseF. Cat 3 cable was more durable and less expensive than the coax formerly required. Of greatest importance, a problem with a cable only affected a single station, and would not bring down the entire network.

Since then, 100BaseT or Fast Ethernet has been introduced. It runs at ten times the data rate of 10Base-T, or 100MBit/s. Since the data rate is ten times as high as 10Base-T, and the minimum packet size is the same, the maximum network diameter had to be reduced to 200m. CobraNet® uses Fast Ethernet ports but can be transported over Gigabit Ethernet between switches.

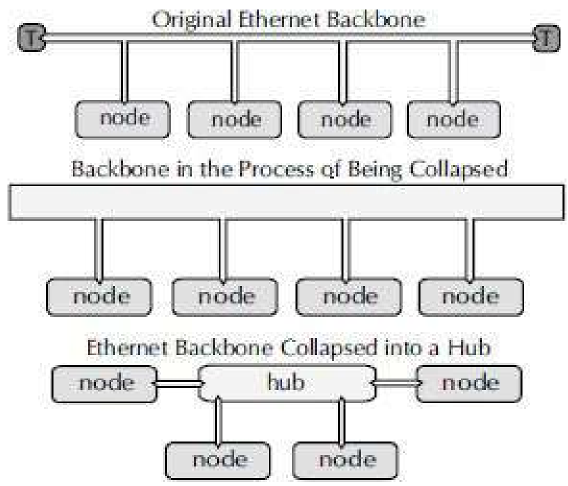


Figure 42-19. Transforming the Ethernet backbone into a repeater hub.

Within Fast Ethernet there are several varieties. 100Base-T4 uses all 4 pairs of a Cat 3 UTP cable. 100Base-TX uses 2 pairs of a Cat 5 cable. This is the most common variety of Fast Ethernet. Both of these varieties allow single cable runs of 100m. 100Base-FX uses multimode fiber, and allows single runs of up to 2000m. A version of Fast Ethernet to run over single-mode fiber has not been Standardized, but many manufacturers sell their own versions, which allow distances of as much as 100,000m in a single run.

Many Fast Ethernet devices sold today not only support 100Base-TX, but also 10Base-T. Such a dual speed port is commonly called a

10/100 Ethernet port. It will negotiate automatically with any Ethernet device hooked to it and connect at the highest speed both ends of the link support. The technique for this negotiation is described below.

Gigabit Ethernet is now available, and the price has dropped so much it is more and more replacing Fast Ethernet. As you might have guessed it runs at a rate ten times as fast as 100BaseT, or 1000MBit/s. The first versions ran over optical fiber, but now a version that runs over Cat 5e UTP cabling is available. It does, however, use all four pairs in the cable. Gigabit Ethernet increases the minimum packet size from 64 bytes to 512 bytes in order to allow the network diameter to stay at 200m. Ethernet ports that support 10/100/1000MBit/s speeds and autonegotiate to match the highest speed the connected device can support are now common.

Within Gigabit Ethernet there are also several varieties. 1000Base-LX (L for long wavelength) can be used with either multimode or single-mode optical fiber. 1000Base-SX (S for short wavelength) is used with multimode fiber only. 1000Base-SX is less expensive than 1000Base-LX. 1000Base-LH (LH for long haul) is not an IEEE standard, but is supported by many manufacturers. Manufacturers make different versions depending on the distance to be covered. 1000Base-T runs over Cat 5 cable using all four pairs. The maximum single cable run is 100m.

A version of Ethernet that will run at ten times the speed of Gigabit Ethernet is available and the price has been dropping.

Several manufacturers power their products over Ethernet cabling. There is now an IEEE Standard for Power over Ethernet (PoE) and most manufacturers sending power to their products over the Ethernet cabling have gone to this standard.

Wireless Ethernet to the IEEE 802.11 Standard has become very popular and inexpensive. It provides a variable data rate based on distance and environmental conditions. The best case data rate for 802.11ac (the latest as of this writing) is 6.77Gbits/s for an 8 antenna access point, but single antenna data rates are only 433MBit/s.

42.11.5 Ethernet Topology

The Ethernet topology shown in Fig. 42-19 as a collapsed backbone is commonly called a star topology, since every station connects back to the common hub. It is also permissible to tie multiple stars together in a star of stars, Figs. 42-20 and 42-21.

Using fiber to interconnect the stars can increase the distance between clusters of stars, Fig. 42-22.

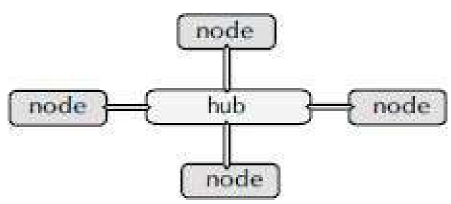


Figure 42-20. Ethernet star topology.

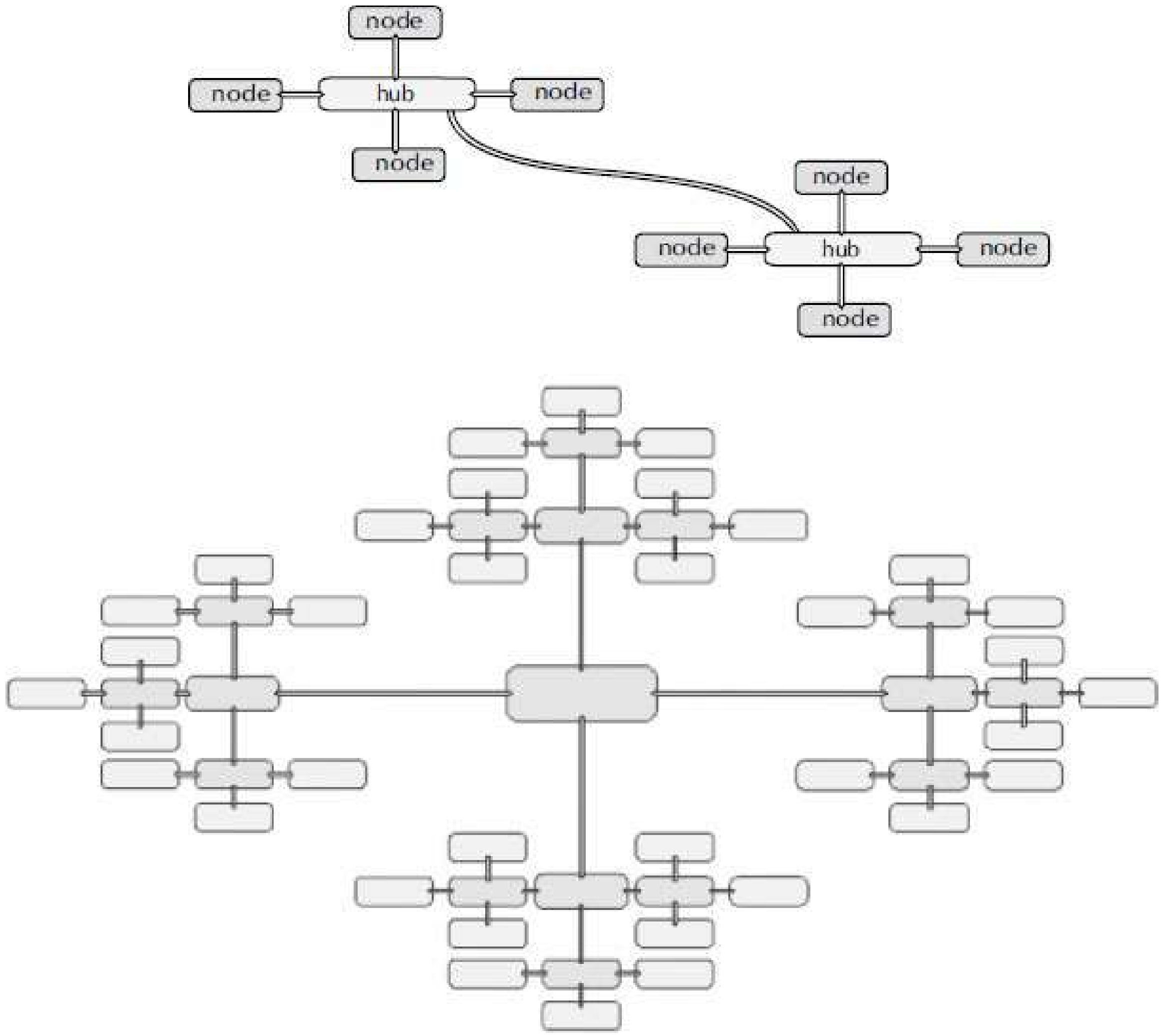


Figure 42-21. Ethernet star of stars topology examples.

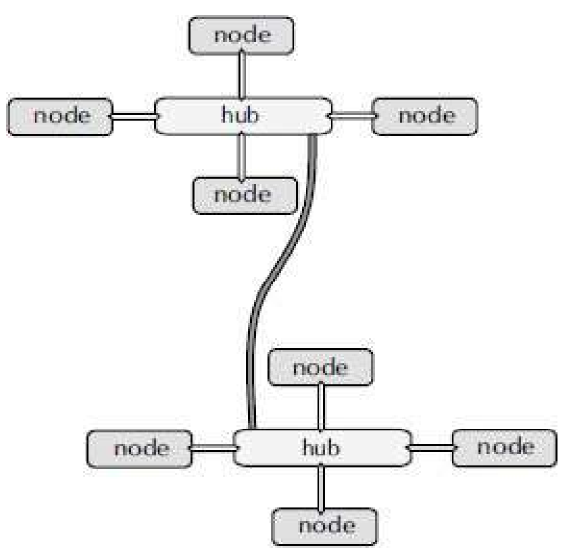


Figure 42-22. Fiber used to interconnect two stars. This allows breaking the copper cable 100m limitation. Now the only limitation is network diameter. As you can see from Figs. 42-18, 42-19, and 42-20, quite large networks can be made with simple wiring using the star of stars approach. In all these examples, there are no loops. This is because a loop in Ethernet, unless special techniques are used, results in something called a *broadcast storm*, which is sort of the data equivalent of runaway feedback in an audio system.

42.11.6 Ethernet Equipment

This will become clearer when we examine the internal functions of the repeater hubs we have been talking about. A hub has ports that connect either to stations or other hubs. Any data that comes in a port is immediately sent out all the other ports *except* the port it came in on, Fig. 42-23. An audio analogy would be a mix-minus system.

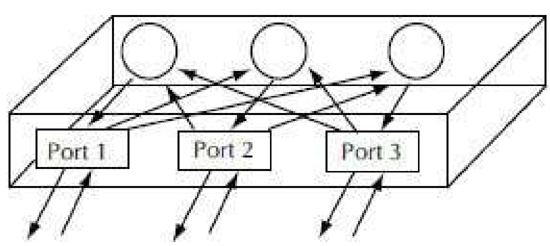


Figure 42-23. Repeater hub functional diagram.

One of the factors keeping the size of the network from growing is that all of these star and star of stars topologies still have the same network diameter limitation. One way to build a bigger network is to isolate data in one star from that in another, and only pass between stars packets that need to reach stations in the other star. Collisions that occur in a given star are not passed to the other stars since only complete packets addressed to a station in the other star are passed on. This isolates each star into a collision domain of its own, so the network diameter limitation only applies within a given collision domain.

The device that provides this function between a pair of collision domains is called a bridge. As the technology became cheaper multiport bridges started to appear that were called switches. As switches become popular and bridges fade from use, you will sometimes see a bridge referred to as a two-port switch, Fig. 42-24.

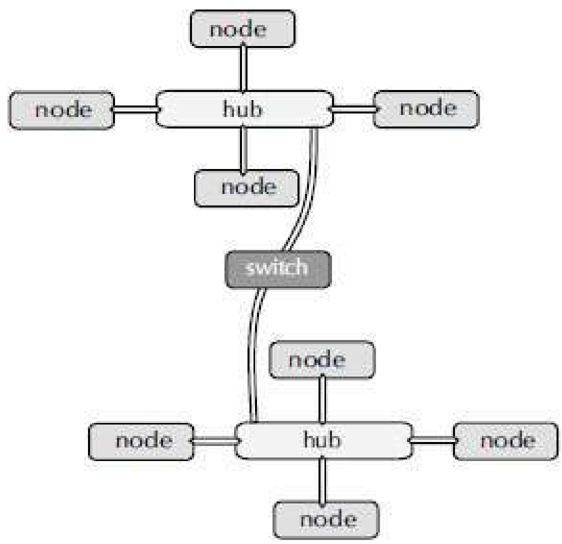


Figure 42-24. An Ethernet switch being used to isolate two collision domains.

Ethernet switches receive a packet into memory. They examine the destination address and decide which of their ports has attached to it the station with the address in question. Then if the destination is not on the same port as the packet was received from, the switch forwards the packet to only the correct port. Packets where the destination address is on the same port as the packet was received from are discarded.

Switches determine which addresses are connected to each port by recording the source address of every packet received, and associating that address with the port from which it was received. This information is assembled in a look up table (LUT) in the switch. Then as each packet is received, the switch checks to see if the destination address is in the LUT. If it is, the switch knows where to send that packet and only sends it out the appropriate port. If a destination address is not in the LUT, the switch sends that packet out every port but the one from which it was received. Since most Ethernet stations respond to packets received addressed to them, when the response is sent the switch learns which port that address is on. Thereafter packets addressed to that station are only sent out the correct port.

If a given MAC address is found on a different port than was contained in the LUT, the LUT is corrected. If no packet is received from a given MAC address within a timeout window of perhaps 5 minutes, its entry in the LUT is deleted. These characteristics allow the switch to adapt and learn as network changes are made.

Packets intended to go out a given port are never allowed to collide inside the switch. Instead each outgoing packet is stored in a first in first out (FIFO) buffer memory assigned to a given port, and transmitted one at a time out the port.

While most data passing through a switch behaves as described above, there is one type of packet that does not. Most data packets are addressed to a specific destination MAC address. This is called unicast addressing. There is a specific address called the multicast or broadcast address. Packets with this address in their destination field are sent to all stations. Therefore, these packets are sent out all ports of a switch except the port they came in on.

Switches are not the shared media of the early coaxial cable Ethernet varieties, or the newer repeater hubs. Instead by storing the packets, examining the addresses, selectively passing the packets on, and FIFO buffering the outputs, they break the network diameter limitation.

Switches have another difference from repeater hubs. Repeater hubs and stations connected to them operate in half duplex mode. In other words a given station can only receive or transmit at different times. If a station that is transmitting in half duplex mode sees a received signal, that tells it a collision has occurred. Since switches store and buffer the packets, they can operate in full duplex mode with other switches or with stations which can operate full duplex. When a station is connected to a switch in full duplex mode it can receive at the same time as it transmits and know that a collision can't occur since it is talking to a full duplex device which does not allow collisions to occur internally, Fig. 42-25.

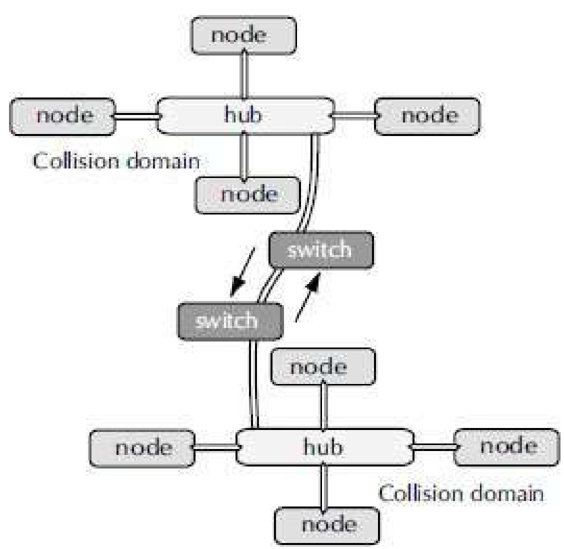


Figure 42-25. Two Ethernet switches showing full duplex operation between them, and the isolation of two collision domains.

Full duplex operation has the added benefit of doubling the communications bandwidth over a half duplex link. A half duplex fast Ethernet connection has 100MBit/s of available bandwidth which must be split and shared between the packets going each direction on that link. This is because if packets were going both directions at once, that by definition would be a collision. A full duplex link on the other hand has no problem allowing packets to flow in both directions at once, so a fast Ethernet link has 100MBit/s capability in each direction.

Of course a repeater hub-based fast Ethernet network has only 100MBit/s of total bandwidth available for the entire network since it uses a shared media. A network based entirely on fast Ethernet switches has 100MBit/s of bandwidth available in each direction on each link that makes up the network, assuming that all the stations are capable of full duplex operation. When you combine no collisions with full duplex operation, a switched network can run much faster than a repeater hub-based network.

The internal packet routing function inside a switch is called the switch fabric or switch cloud. Switches which contain enough packet routing capability in their cloud to never run out, even if all ports are receiving the maximum possible amount of data, are known as “nonblocking” switches.

Proper Ethernet network design includes ensuring that no packet may go through more than 7 switches on its way from the source to the destination.

When switches were first introduced their expense limited their application to the few situations which required their capabilities. Today the price of switches has come down until they are hardly any more expensive than repeater hubs. As a result the repeater hub is becoming a vanishing part of Ethernet history, Fig. 42-26.

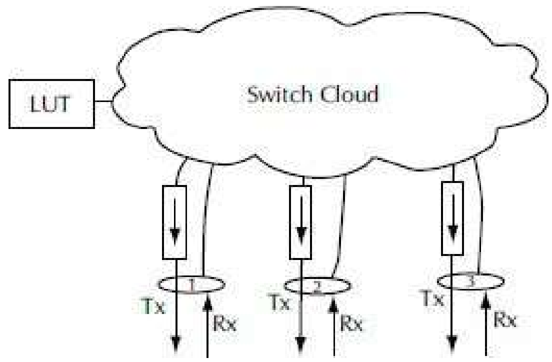


Figure 42-26. Switch functional diagram.

42.11.7 Ethernet Connection Negotiation

It is important to understand how different Ethernet devices negotiate connections between themselves in order to understand why some combinations of devices will work and others won't.

If a 10MBit/s Ethernet device is not transmitting data, its output stops. After a period of no data transmissions, it will begin sending normal link pulses (NLPs). These allow the device at the other end of the link to know that the connection is still good, and it serves to identify the device as a 10MBit/s device.

100MBit/s devices on the other hand always send a signal even when no data is being transmitted. This signal is called a carrier, and serves to identify the device as a 100MBit/s device.

10/100 Ethernet devices often use a technique called autonegotiation to establish the capabilities of the device at the other end of the link, before the link is established. This process determines if the other device is capable of full or half duplex operation, and if it can connect at 10MBit/s, 100MBit/s, or Gigabit speeds. Data is conveyed using fast link pulses (FLPs), which are merely sequences of NLPs that form a message.

If at the end of the autonegotiation process a 10MBit/s link is established, both devices will send NLPs when idle. If a 100MBit connection was established, both devices transmit carrier signals.

Autonegotiating devices also utilize parallel detection. This enables a link to be established with a nonne-gotiating fixed speed device, and for a link to be established before a device detects the FLPs. The state diagram in Fig. 42-27 shows how the different possible end conditions are reached. Notice that a 100MBit device can never parallel detect into a full duplex link.

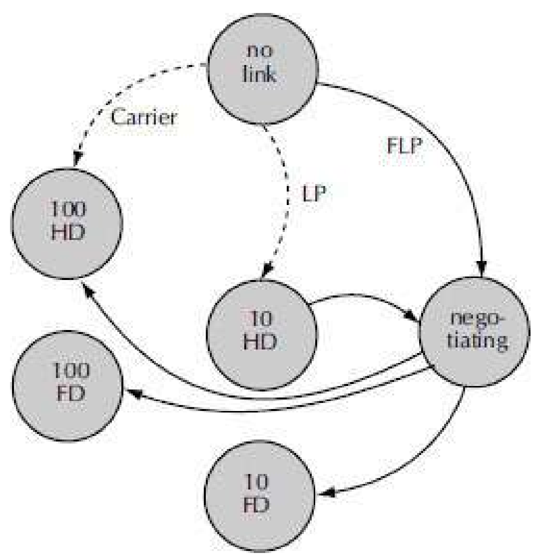


Figure 42-27. Ethernet autonegotiation state diagram. 10 = 10MBit/s, 100 = 100MBit/s, HD = half duplex, FD = full duplex. The dashed lines show the parallel detection that can take place while waiting for the FLPs to be recognized. Note that if parallel detection sets up a 100MBit half duplex connection, then a full duplex connection can never be established.

Fiber optic links do not pass the FLPs needed for autonegotiation though they do pass the carrier. One consequence of this is that if full duplex operation over a fiber link is desired, either manual configuration is required, or an intelligent media converter is required. Such a converter includes circuitry to autonegotiate the link at each end of the fiber.

If a 10/100 NIC were to connect to a 10MBit/s repeater hub, the 10MBit/s hub sends NLPs that are detected by the NIC. Seeing the NLPs, the 10/100 NIC goes to 10MBit/s half duplex operation, and link is established. This is correct since hubs are half duplex devices.

If a 10/100 NIC were to connect to a 100MBit/s repeater hub, the 100MBit/s hub sends a carrier that is detected by the NIC. Seeing the carrier, the 10/100 NIC goes to 100MBit/s half duplex operation, and link is established. This is correct since hubs are half duplex devices.

If a 10/100 NIC were to connect to a 10/100MBit/s switch, the switch sends FLPs that are detected by the NIC. After interpreting the FLPs, the 10/100 NIC goes to 100MBit/s full duplex operation, and link is established. This is correct since 100MBit/s is the highest common rate, and switches can operate full duplex.

A media converter can be thought of as a two-port repeater that coverts from one type of media to another. The most common such conversion is from copper to fiber. There are two basic types of media converters, simple and intelligent.

Simple media converters have no intelligence and just convert electrical signals to light and back. These simple media converters can’t pass or detect FLPs. Therefore, they can’t pass the signals needed for autonegotiation from one end to the other, nor are they capable of autonegotiating with the ports at each end on their own.

Intelligent media converters add electronics at each end of the fiber link that are able to generate and interpret FLPs. As a result such a converter can either autonegotiate with the port at each end, or be manually configured. They are also capable of both half and full duplex operation.

42.11.8 Managed Switches

All switches have the capabilities described so far. Some switches add significant additional capabilities. Switches with just the basic capabilities are known as *unmanaged switches*, and have become very inexpensive. Unmanaged switches operate automatically, and do not require special settings for their operation. Managed switches provide the capability to control the switch's internal settings.

Common control techniques include a dedicated serial port for control, and Telnet, Web access using a normal Web browser, or Simple Network Management Protocol (SNMP). The last three control methods function over the network. Some managed switches provide all four methods of control. The control capabilities available using each method will often differ. The methods that work over the network will usually require that first the switch is accessed via the serial port, and an IP number assigned to the switch. After that the other control methods can reach the switch over the network at the assigned IP address.

Not all managed switches will provide all the additional capabilities that will be mentioned. Check with the manufacturer of the switch to determine its exact capabilities.

42.11.9 Virtual Local Area Network (VLAN)

Virtual Local Area Network (VLAN) capability allows certain switch ports to be isolated from the other ports. This allows dividing up a larger switch into several virtual smaller switches. While this capability may not matter if all the data is unicast, if there is any multicast or broadcast traffic, there might be significant benefit to isolating that traffic to just certain ports. Some switches may allow data from several VLANs to share a common link to another switch without the data being mingled. Both switches must support the same method for doing this for it to work. Most today use a technique called *tagging* to allow isolated VLANs to share a common physical link between switches.

42.11.10 Quality of Service (QoS)

Quality of service (QoS) allows priority to be given to certain data trying to leave a switch over other data. For example, if we are sending audio over Ethernet using CobraNet®, we would not want there to be any dropouts in the audio if there was a momentary spike in normal computer data traffic through the switch. Such a dropout could occur if a surge in computer data traffic took up bandwidth needed for audio transmission and delayed the reception of the audio packets.

Several means can be used to specify to the switch which traffic is to be given priority. Priority can be given to traffic on a certain VLAN, or that received from certain ports, or that received from certain MAC addresses, or even traffic containing a specific protocol identifier.

42.11.11 Routing

Ethernet switches normally don't examine the payload portion of the Ethernet packet. Routers are capable of looking inside the payload and routing Ethernet packets based on Internet Protocol (IP) addresses that might be found inside some Ethernet payloads. Such a router, or a routing function built into some switches, can allow packets to flow between normally independent networks or VLANs. This can be very useful, for example, to allow a central SNMP management system to monitor and control all the network devices in a facility even if they are in independent isolated networks or VLANs.

42.11.12 Fault Tolerance

42.11.12.1 Trunking

Trunking or link aggregation allows two or more links between switches to be combined to increase the bandwidth between the switches. Both switches must support trunking for this to work. While the algorithm used to share the traffic between the links works for many types of data, it does not for all possible types of data. You may find situations where adding a second link and activating trunking between two switches does not provide any significant increase in available bandwidth, Fig. 42-28.

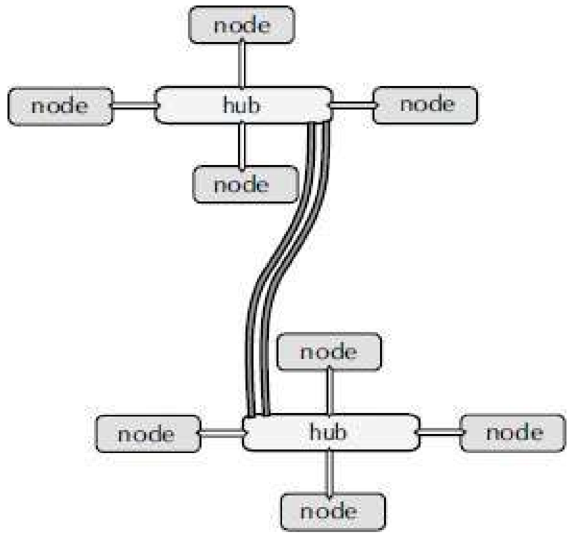


Figure 42-28. Example of trunking between two switches.

Trunking does provide increased fault tolerance, particularly if the links aggregated run through different physical paths between the two switches. If one link is lost, the other link or links will continue to carry traffic between the switches.

42.11.12.2 Spanning Tree

Spanning tree provides automatic protection from inadvertent loops in a network's topology. The cost for this protection is a delay in the activation of a connection made to a port where spanning tree is activated while the switch tries to determine if there is a path out that port that eventually returns to the same switch. If it finds no such path, it will activate the port. The delay might be on the order of 30 seconds to a minute. If it finds a path back to itself on the port, it will disable that port. Whenever a connection to a port is made or lost, and the port has spanning tree active, the switch will reexamine all the ports for loops and activate those where loops are not found.

In any network, damage to the cabling is one of the more common causes of network failures. Spanning tree can be used to reduce the impact of such failures on network operation, Fig. 42-29.

When managed switches with spanning tree capability are used it is common to deliberately build the network with loops. The switches will find the loops and disable enough of the links between switches to insure the network topology is a star of stars and stable. If one of the active links is later disabled, perhaps due to physical damage to the cable or the failure of another switch, then one or more of the currently disabled links will automatically be restored to operation. This provides an inexpensive way to increase the reliability of a network.

Fig. 42-30 shows one possible network topology that can be stable if spanning tree is used. Such a network design can be quite robust, and accommodate multiple failures while maintaining operation.

One difficulty with designing a network that uses spanning tree is that we can't know which links will be disabled and which will stay active. This makes it difficult to predict the amount of traffic a given link will carry.

While spanning tree used with the correct network topology can increase system reliability, it does not respond instantly to failures or changes in network topology. At times it may take several minutes for operation to be restored after a failure.

42.11.12.3 Meshing

At this time meshing is only available on some Hewlett Packard (HP) Procurve switches. Meshing is an attempt to combine the best portions of trunking and spanning tree into a new protocol. Unlike spanning tree, meshing does not disable any links. Instead it keeps track of packets and prevents them from being recirculated around loops. When there are multiple possible routes for a packet to take to its destination, meshing attempts to send the packet by the most direct route.

One of the most significant advantages of meshing is that recovery from failures of links or switches is far faster than spanning tree, and may be accomplished in seconds rather than minutes.

42.11.12.4 Rapid Spanning Tree

More recently a new protocol has brought much of the advantages of meshing and other proprietary technologies into the general market. It allows restoration of a network typically in seconds rather than minutes.

42.11.13 Core Switching

At its simplest a core switch can be thought of as the central switch in a star of stars configuration. Data that needs only to travel between devices local to each other is switched through the edge switches and never goes through the core switch. Core switches often run at ten times the data rate of the edge switches. For example if the edge switches are fast Ethernet, they will each have a gigabit uplink port that connects back to the gigabit Ethernet core switch.

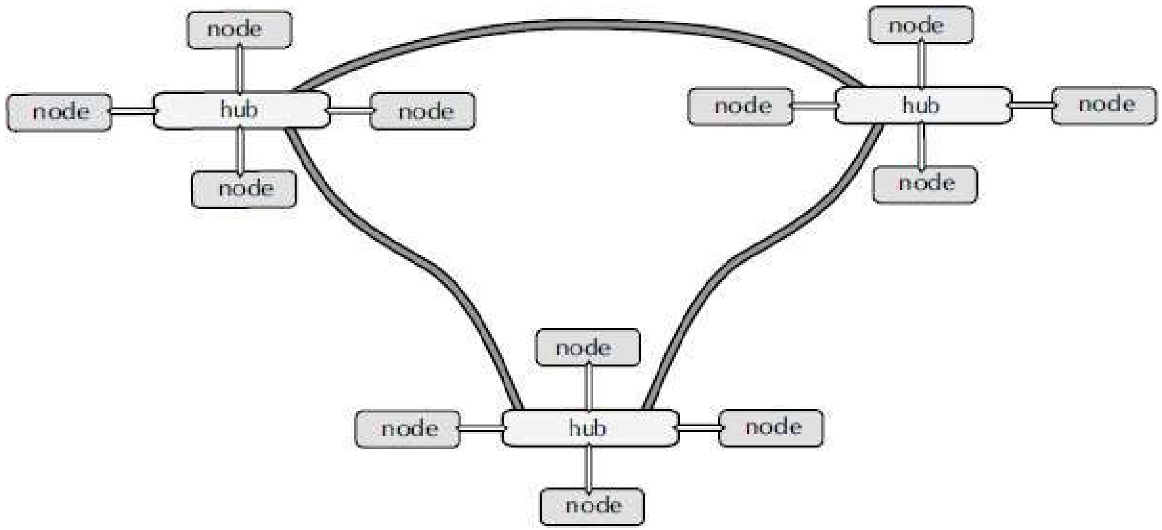


Figure 42-29. A loop around three switches. Spanning tree would disable one of the three links between the switches, and allow the network to be stable.

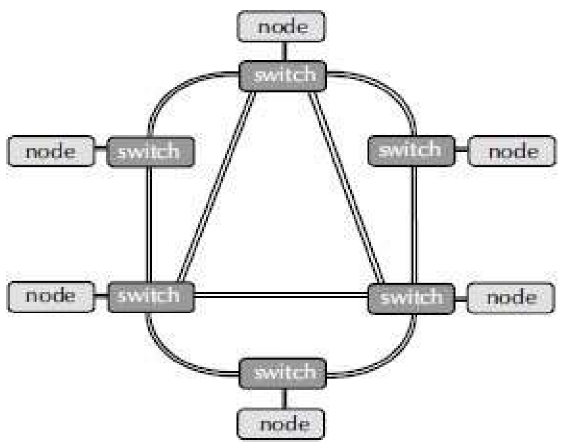


Figure 42-30. Multiple loops around six switches. This is a typical topology used to increase the fault tolerance of a network. Spanning tree would disable enough of the links between the switches to allow the network to be stable.

Besides allowing ten times the data traffic, another reason to use the next higher-speed protocol in the core switch is that the latency through the higher-speed link and switch is only 1/10 as long as if the higher speed was not used.

Some core switches will be equipped with routing capabilities to allow easy central control of all the VLANs using SNMP management.

Core switches are often built to higher-quality standards than ordinary switches since such a core switch can be a single point of failure in the network.

To prevent a single point of failure and greatly increase the fault tolerance of the network, it is possible to use a pair of core switches, each of which connects to all of the edge switches. The network will continue full operation even if one of the core switches or any link to a core switch was to fail.

42.11.14 Ethernet Wiring

Proper design of an Ethernet cable plant is important for reliable operation, ease of maintenance, and maximum performance.

A typical Ethernet network cable path or link is shown in Fig. 42-31. The items that make up the cable plant include:

- Cabling connecting nodes—this can be Cat 5 or fiber optic cable.
- Wiring closet patch panels.
- Station cables—the cable that runs from node to wall plate.
- Wall plates—the data or information outlet close to the node.

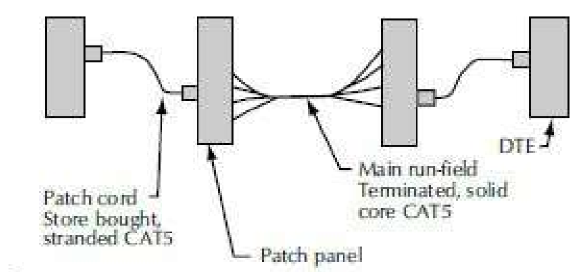


Figure 42-31. Ethernet typical cable plant showing the entire link

from one Ethernet device to another.

It is considered good design practice to include the intermediate patch points as shown. This gives the cable plant operator flexibility in accommodating expansion and configuration changes.

There are two main types of cables used in Ethernet networks: Cat 5 cable and fiber optic cable. The following sections will describe these cable types, as well as the issues associated with each.

42.11.14.1 UTP Cable Grades

Unshielded twisted pair (UTP) cables are graded in several categories.

- Quad: nontwisted four conductor formerly used for telephone premise wiring.
- Category 1: No performance criteria UTP.
- Category 2: Rated to 1MHz (old telephone twisted pair).
- Category 3: Rated to 16MHz (10Base-T and 100Base-T4 Ethernet, current FCC required minimum for telephone).
- Category 4: Rated to 20MHz (token-ring).
- Category 5: Rated to 100MHz—now withdrawn as a Standard and replaced by Cat 5e.
- Category 5e: Improved Cat 5 with tighter tolerances. (100Base-TX and 1000Base-T Ethernet).
- Category 6: Rated to 250MHz.
- Category 7: Shielded cabling mostly used in Europe.

When used with quality equipment there usually is not a lot of advantage for fast Ethernet networks in using cable with a rating beyond Cat 5e. Future higher speed networks, or marginal equipment on fast Ethernet may benefit from improved cable.

Unless specified differently by the manufacturer, most UTP has a minimum bend radius of 4 times the cable diameter or about 1 inch.

Cat 5e is inexpensive unshielded twisted pair (UTP) data grade cable. It is very similar to ubiquitous telephone cable but the pairs are more tightly twisted. It should be noted that not all Cat 5e cable is UTP. Shielded Cat 5 also exists but is rare due to its greater cost and much shorter distance limitations than UTP Cat 5e.

42.11.14.2 Distance Limitations

On fast Ethernet systems, Cat 5e cable runs are limited to 100m due to signal radiation and attenuation considerations. A Cat 5e run in excess of 100m may be overly sensitive to electromagnetic interference (EMI).

42.11.15 Connectors

Cat 5 cable is terminated with an RJ-45 connector. Strictly speaking this nomenclature is incorrect since it designates a particular telephone usage of the connector rather than the connector itself. Since 8 position 8 contact nonkeyed modular connector is difficult to say and write, we are stuck with the common usage of RJ-45, Fig. 42-32.

Figure 42-32. "RJ-45" plug. This is an 8 position 8 contact nonkeyed modular connector originally developed for telephone applications. It is the connector used for all UTP cables in Ethernet networks.

There are two different types of contacts in RJ-45 connectors. There is the *bent tyne* contact, intended for use with solid core Cat 5, and then there is the *aligned tyne* contact used with stranded Cat 5 cable. Errors can occur when using an incorrect cable/connector combination. Fig. 42-33 shows an end-on view of a single contact in a modular connector. The aligned tyne contact (on left) must be able to pierce through the center of the wire, therefore it can only be used on stranded wire. The bent tyne contact has the two or three tynes offset from each other to straddle the conductor; therefore, it can be used on solid or stranded wire.

Cable openings in modular connectors can be shaped for flat, oval, or round cable. Cat 5 cable does not usually fit properly into connectors made specifically for flat telephone cable, Fig. 42-34.

Cheap modular connectors may not have proper gold plating on the contacts, but instead only have a gold flash. Without proper plating, the connectors may quickly wear and corrode, causing unreliable connections.

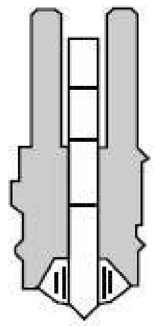


Figure 42-33. Different style contacts found in RJ-45 plugs. On the left is the aligned tyne contact style such as found on original telephone connectors. Since it must pierce through the wire, it must only be used on stranded wire. On the right is the bent tyne contact. It has two or three tynes which are offset from each other and straddle the conductor. As a result it can be used on either stranded or solid conductor wire.

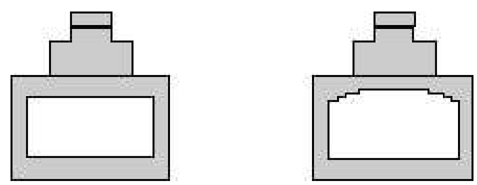


Figure 42-34. Different style cable openings found in RJ-45 plugs. On the left is the opening for flat cable such as found on original telephone connectors. On the right is the opening for round cable needed for Cat 5e cable.

AMP makes quality modular connectors, but the secondary crimp point is located in a different position from everyone else's connectors. Fig. 42-35 shows a standard crimp die and an AMP plug. Point "A" is the primary crimp point, and should fold the primary strain relief tab in the plug down so that it locks against the cable jacket. At the opposite end of the plug, the contacts are pressed down into the individual conductors. The "B" secondary crimp point secures the individual conductors so that they do not pull out of the contacts.

AMP puts this crimp in a different location from all other manufacturers. If AMP connectors are used in a standard crimper they will either jam, bend, or break the crimp die. If standard connectors are used in an AMP crimper, the die will usually break. Once either type of plug is properly crimped onto the wire, they are interchangeable and will work properly in any mating jack, Fig. 42-35.

Some plugs are made with inserts that guide the wires. These can make the job of properly assembling the connector easier. Some connectors made with inserts may also provide better performance than Cat 5, Figs. 42-36 and 42-37.

Pairing, Color Codes, and, Terminations

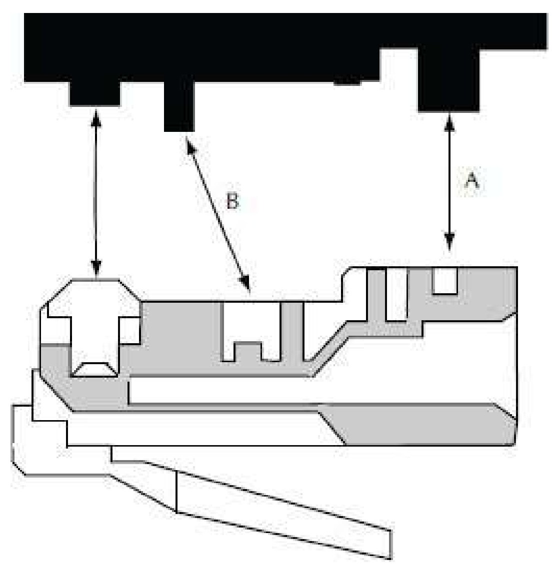


Figure 42-35. A standard crimp die above an AMP connector. Note the misalignment of the secondary strain relief crimp point B. An AMP connector will jam in a standard crimper, while a standard RJ-45 will usually break an AMP crimp die.

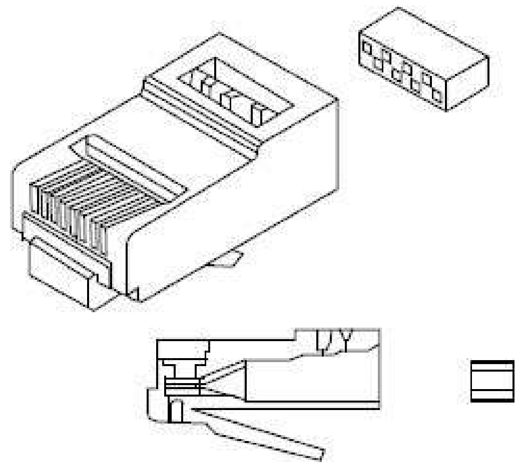


Figure 42-36. An RJ-45 connector that uses an insert to guide the individual conductors into the connector. Such a connector can be easier to properly assemble on a cable. Some connectors made this way provide performance well beyond Cat 5.

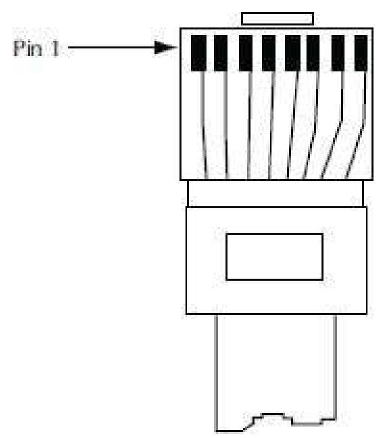


Figure 42-37. RJ-45 pin numbering.

Cat 5 cable consists of four twisted pairs of wires. To minimize the crosstalk between the pairs, each pair is twisted at a slightly different rate. For fast Ethernet, one pair is used to transmit (pins 1 and 2) and another pair is used to receive (pins 3 and 6). The remaining two pairs are terminated but unused by fast Ethernet. Although only two of the four twisted pairs are used for fast Ethernet, it is important that *all pairs be terminated*, and that the proper wires be twisted together. Standards set forth by EIA/TIA 568A/568B and AT&T 258A define the acceptable wiring and color-coding schemes for Cat 5e cables. These are different from the USOC wiring Standards used in telecommunications, Figs. 42-38 and 42-39.

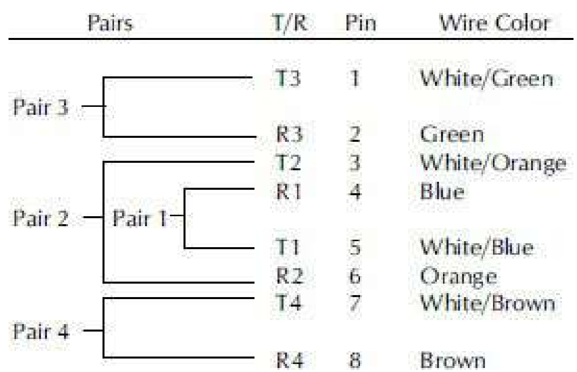


Figure 42-38. Standard EIA/TIA T568A (also called ISDN, previously called EIA). One of the wiring schemes used for Ethernet.

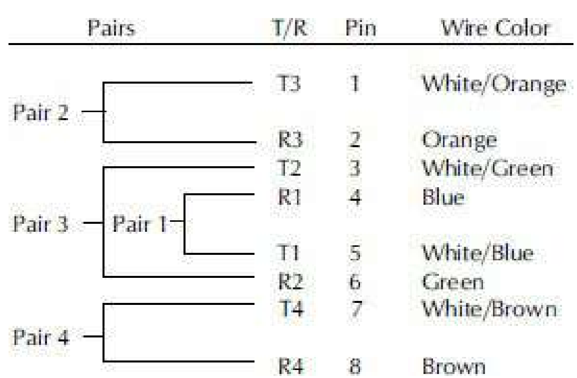


Figure 42-39. Standard EIA/TIA T568B (also called AT&T specification, previously called 258A). One of the wiring schemes used for Ethernet.

Note that there are two conflicting color code standards for data use of the RJ-45 connector. Both work just fine, but to avoid problems make sure that one of the standards is selected and used uniformly throughout a facility.

Often when installers accustomed to telephone wiring install data cabling they will incorrectly use the telephone USOC (Universal Service Ordering Code) wiring scheme. This will result in a network that either does not work, or has very high error rates, Fig. 42-40.

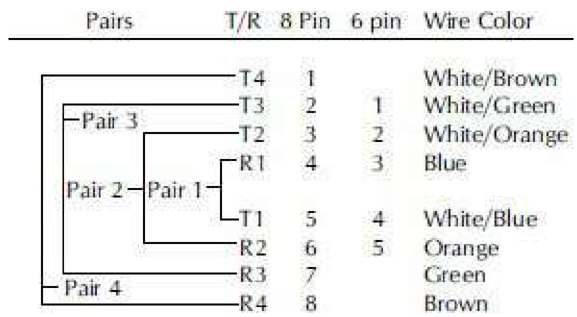


Figure 42-40. USOC (Universal Service Order Code), the wiring scheme used for telephones. This must not be used for data! The eight contact connector uses all pairs for four lines. The six contact connector uses only center one-three pairs for one, two, or three line phones.

Normal Ethernet cable wiring such as shown in Fig. 42-41 is used for interconnecting unlike devices. In other words it is used to connect the Network Interface Card (NIC) in a station to a switch or repeater hub. Connections between like devices such as a pair of NICs, or between switches, repeater hubs, or switch to repeater hub require a “crossover” cable wired per Fig. 42-42. This is because the data transmit pair must connect to the receive input, and vice versa.

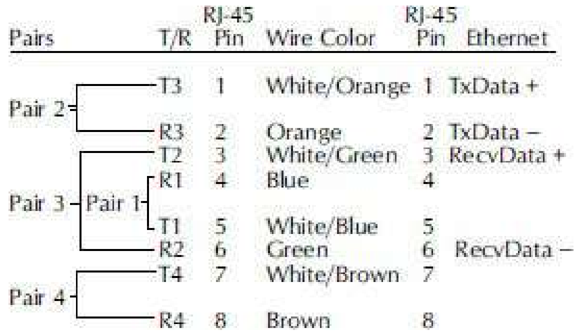


Figure 42-41. Ethernet Standard (T568B colors) patch cord wiring used for most interconnects. Ethernet usage shown for 10Base-T and 100Base-T. Gigabit Ethernet uses all the pairs.

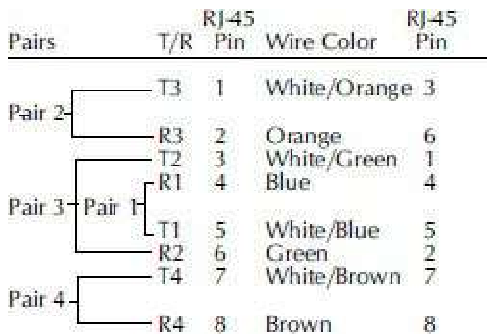


Figure 42-42. Ethernet standard (T568B colors) crossover cord. Pairs 2 and 3 are reversed end to end. Used for connections between like devices (NICs, switches, or repeater hubs).

It is very easy to tell the difference between a crossover cable and a straight-through cable by looking at the conductors in the RJ-45 connectors. If the wiring is identical at both ends, you are holding a straight- through cable, if it is different, you most likely have a crossover cable.

Some hubs and switches have uplink ports that can eliminate the need for crossover cables. Such a port is wired with pairs 2 and 3 reversed internally. Make sure that when connecting two switches or repeater hubs so equipped, you only use the uplink port at one end. Another caution is that often such an uplink port is not an independent port but is wired internally to one of the normal ports. In such a case make sure that only one of the pair of ports is used.

Some switches employ an autoselect crossover feature. This allows the use of either a straight-through or a crossover cable on any port. The switch automatically senses which cable type is in use and adjusts the electronics to suit the cable.

Stranded patch cable sometimes has different colors.

Pair 1 Green and Red

Pair 2 Yellow and Black

Pair 3 Blue and Grey

Pair 4 Brown and Grey

42.11.16 Fiber Optic Cable

There are two basic varieties of fiber optic cable, single-mode and multimode. Both are used in Ethernet network designs. Two fibers are needed to make an Ethernet connection, one fiber for transmit, and one for receive, Fig. 42-43.

Multimode fiber is built of two types of glass arranged in a concentric manner. Multimode fiber allows many modes, or paths, of light to propagate down the fiber optic path. The relatively large core of a multimode fiber allows good coupling from inexpensive LED light sources, and the use of inexpensive couplers and connectors, Fig. 42-44.

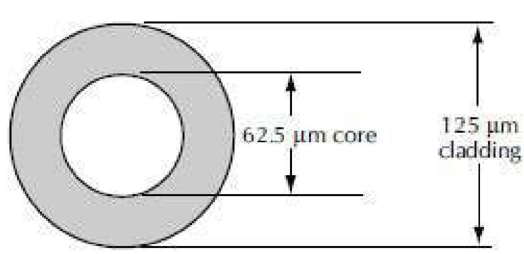


Figure 42-43. Cross-section through a multimode fiber optic

cable.

Figure 42-44. Possible light paths down a multimode fiber optic cable. Note that there are multiple possible paths for the light to take. This is why this is called multimode cable.

Two sizes of multimode fiber are available. 62.5/125µm is used primarily in data communications, and 50/100µm is used primarily in telecommunications applications. The standard for transmission of 100Mbit Ethernet over 62.5/125µm multimode fiber is called 100Base-FX. 100Base-FX has a 2km distance limitation.

Single-mode fiber optic cable is built from a single type of glass. The cores range from 8µm to 10µm, with 8/125µm being the most commonly used. There is only a single path of light through the fiber, Fig. 42-45.

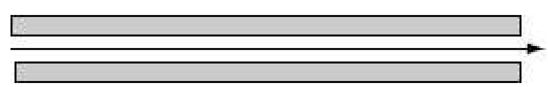


Figure 42-45. Possible light paths down a single-mode fiber optic cable. Note that there is only one possible path for the light to take. This is why this is called single-mode cable.

While single-mode fiber cable costs approximately the same as a multimode cable, the cost of the optical transmitters and receivers is significantly more for a single-mode installation than multimode. Single-mode fiber has a core diameter that is so small that only a single mode of light is propagated. This eliminates the main limitation to bandwidth, but makes coupling light into the fiber more difficult.

Although multimode fiber cable has a specific distance limitation of 2km, distance limitations of single-mode fiber vary according to the proprietary system in use. All are in excess of 2km with some allowing 100km. There is currently no Ethernet standard for single-mode fiber.

42.11.16.1 Fiber Optic Connectors

There are two common types of fiber optic connectors, SC and ST, Fig. 42-46. The ST, or straight tip, connector is the most common connector used with fiber optic cable, although this is no longer the case for use with Ethernet. It is barrel shaped, similar to a BNC connector, and was developed by AT&T. A newer connector, the SC, is becoming more and more popular. It has a squared face and is thought to be easier to connect in a confined space. The SC is the connector type found on most Ethernet switch fiber modules and is the connector of choice for 100 Mbit and gigabit Ethernet. A duplex version of the SC connector is also available, which is keyed to prevent the TX and RX fibers being incorrectly connected.

Figure 42-46. The two most common fiber optic connectors. The ST on the right has been the most popular. Today most Ethernet fiber optic interfaces come equipped for the SC on the left.

There are two more fiber connectors that we may see more of in the future. These are the MTRJ and MTP. They are both duplex connectors and are approximately the size of an RJ-45 connector.

42.11.16.2 Cabling and Network Performance

A number of factors can degrade the performance of an Ethernet network, and among these is a poor cable plant. Cabling problems and a susceptibility to EMI can actually lead to packet loss. The following sections present cabling considerations that will help to insure a high-quality cable plant installation.

The Cat 5 specifications require that no more than ½ inch of the pairs be untwisted at each termination. It is good practice to never strip more of the outer jacket of the cable than is required, and to keep the cable pairs twisted at their factory twist rates until the point they must separate to enter the terminations.

As with audio cabling, there are certain proximity specifications to be aware of when designing your network cable routes. Table 42-6 lists some UTP proximity guidelines. For fiber optic cable runs, proximity is not a concern due to fiber's inherent immunity to EMI and RFI.

Table 42-6. Ethernet UTP Proximity Specifications

Condition	<2kVA	2-5kVA	>5kVA
Unshielded power lines or electrical equipment in proximity to open or nonmetal pathways	5in (12.7cm)	12 in (30.5cm)	24in (61cm)
Unshielded power lines or electrical equipment in proximity to grounded metal conduit pathway	2.5in (6.4cm)	6in (15.2cm)	12 in (30.5cm)
Power lines enclosed in a grounded metal conduit (or equivalent shielding)	N/A	6 in (15.2cm)	12 in (30.5cm)

in proximity to grounded metal conduit pathway			
Transformers and electric motors	40in (1.02m)	40in (1.02m)	40in (1.02m)
Fluorescent lighting	12 in (30.5 cm)	12 in (30.5cm)	12 in (30.5cm)

Fiber is insensitive to electromagnetic fields and does not require separation

42.11.17 Cable Installation

42.11.17.1 Cable Ties and UTP

Another factor that can degrade the installation quality is snug cable ties. Ties should never be pulled tight enough to deform or dent the outer jacket of the UTP cable. Doing so produces a slight change in the cable impedance at the point under the tie, which can lead to poor network performance. If tight ties are used at even intervals down the cable length, the performance degradation is even worse.

For best performance with minimum alien crosstalk between cables, they should not be bundled or combed into a straight and neat harness, but instead be allowed to lie randomly and loosely next to each other.

42.11.17.2 Pull Force and Bend Radius

A common myth is that fiber optic cable is fragile. In fact, an optical fiber has greater tensile strength than copper or steel fibers of the same diameter. It is flexible, bends easily, and resists most of the corrosive elements that attack copper cable. Some optical cables can withstand pulling forces of more than 150 pounds! The fact is, Cat 5e cable may be more fragile than optical cables: tight cable ties, excessive untwisting at the connector, and sharp bends can all degrade the cable's performance until it no longer meets Cat 5e performance requirements. While fiber may have a reputation for being more fragile than it really is, it still has limitations, and as such, care should be taken when installing both Cat 5e and fiber optic cables. Here are some guidelines for Cat 5e and fiber optic bend radius and pull force limitations.

42.11.17.3 Cat 5e

All UTP cables have pull force limitations much lower than those tolerated in the audio industry. If more than 25lbs of force is applied to Cat 5e cable during installation, it may no longer meet specification. Like most audio cables, UTP cables also have minimum bend radius limitations. Generic Cat 5e allows a minimum bend radius of four times the cable diameter or 1 in for a ¼ inch diameter cable. Unless specified otherwise by the manufacturer, it is fairly safe to use this as a guideline. Note that this is a minimum bend *radius* and not a minimum bend *diameter*.

42.11.17.4 Fiber Optic Cable

The bend radius and pull force limitations of fiber vary greatly based on the type and number of fibers used. If no minimum bend radius is specified, one is usually safe in assuming a minimum radius of ten times the outside diameter of the cable. For pulling force, limitations begin at around 50lbs and can exceed 150lbs. In general, it is recommended that you check with the fiber manufacturer for specifications on the specific cable used in your installation.

42.11.17.5 Cable Testing

All network cable infrastructure, both copper and fiber, should be tested prior to use, and after any suspected damage. The tester used should certify the performance of the link as meeting the Cat 5, Cat 5e, Cat 6, or whatever performance level you thought you had bought.

Inexpensive Cat 5 testers are often just continuity checkers and are worse than useless since they can provide a false sense of security that the cabling is fine when in fact it may be horrible.

A tester that can correctly certify a link as meeting all of the Cat 5 specifications will cost thousands of dollars, and testers capable of certifying to higher levels are more expensive.

While there are dedicated fiber testers, many of the quality Cat 5 testers can accept fiber testing modules.

42.12 CobraNet®

CobraNet® is a technology developed by Peak Audio and now sold by Cirrus Logic, Inc., for distributing real-time, uncompressed, digital audio over Ethernet networks. The basic technology has applications far beyond audio distribution, including video and other real-time signal distribution.

CobraNet® includes specialized Ethernet interface hardware, a communications protocol that allows isochronous operation over Ethernet, and firmware running on the interface that implements the protocol. It can operate on either a switched network or a dedicated repeater network.

To the basic Ethernet capabilities, CobraNet® adds transportation of isochronous data, sample clock generation and distribution, and control and monitoring functions.

The CobraNet® interface performs synchronous to isochronous and isochronous to synchronous conversions as well as the data formatting required for transporting real time digital audio over the network.

A CobraNet® interface provides conversion from synchronous to isochronous and back, and formats the data to meet Ethernet requirements. This allows it to provide real-time digital audio across the network.

As shown in Fig. 42-47, CobraNet® can transport audio data, and carry and use control information as well as allowing normal Ethernet traffic over the same network connection. Simple Network Management Protocol (SNMP) can be used for control and monitoring. In most cases normal Ethernet traffic and CobraNet® traffic can share the same physical network.

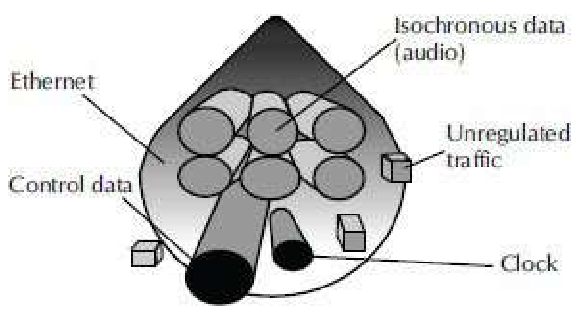


Figure 42-47. CobraNet® Data Services showing the different types of data flowing through the Ethernet network.

42.12.1 CobraNet® Terminology

CobraNet® Interface. The hardware or hardware design with associated firmware provided by Cirrus Logic to CobraNet® licensees and affiliates.

CobraNet® Device. A product that contains at least one CobraNet® interface.

Conductor. The particular CobraNet® interface selected to provide the master clock and transmission arbitration for the network. The other CobraNet® interfaces in the network function as performers.

Audio Channel. A 48kHz sampled digital audio signal of 16, 20 or 24 bit depth.

Bundle. The smallest unit for routing audio across the network. Each bundle is transmitted as a single Ethernet packet every isochronous cycle, and can carry from zero to eight audio channels. Each bundle is numbered in the range from 1 to 65,535. A given bundle can only be transmitted by a single CobraNet® interface. There are two basic types of bundles.

Multicast Bundle. Bundles 1 through 255 are multicast bundles and are sent using the multicast MAC destination address. If a transmitter is set to a multicast bundle number it will always transmit regardless of whether a receiver is set to the same bundle number. Multiple receivers can all pick up a multicast bundle.

Unicast Bundle. Bundles 256 through 65,279 are unicast bundles and are sent using the specific MAC destination address of the receiver set to the same bundle number. Only a single receiver can receive each of these bundles. If no receiver is set for this bundle number the bundle will not be transmitted.

42.12.2 Protocol

CobraNet® operates at the data link layer (OSI Level 2). It uses three distinct packet types, all of which are identified in the Ethernet packet by the unique protocol identifier (8819 hex) assigned by the IEEE to Cirrus Logic. CobraNet® is a local area network (LAN) technology and does not utilize Internet Protocol (IP), which is most important in wide area networks (WAN).

42.12.3 Beat Packet

Beat packets are sent with the multicast destination MAC address 01:60:2B:FF:FF:00. They contain the clock, network operating parameters, and transmission permissions. The beat packet is sent by the Conductor and indicates the start of the isochronous cycle. Because the beat packet carries the clock for the network, it is sensitive to delay variations in its delivery to all the other CobraNet® interfaces. Failure to meet the delay variation specification can keep the other devices from locking their local clocks to the master clock in the Conductor. The beat packet is usually small on the order of 100 bytes, but grows with the number of active bundles.

42.12.4 Isochronous Data Packet

One isochronous data packet is transmitted for each bundle each isochronous cycle, and carries the audio data. It can be addressed to either unicast or multicast destination addresses depending on the bundle number. Since the CobraNet® interfaces buffer the data, out of order delivery within an isochronous cycle is acceptable. To reduce the impact of the Ethernet packet structure overhead on the total bandwidth consumed, data packets are usually large on the order of 1000 bytes.

42.12.5 Reservation Packet

Reservation packets are sent with the multicast destination MAC address 01:60:2B:FF:FF:01. CobraNet® devices usually send a reservation packet once per second. This packet is never large.

42.12.6 Timing and Performance

In order for CobraNet® to provide real-time audio delivery, certain maximum delay and delay variation requirements must be put on the performance of the Ethernet network, Table 42-7.

If the network loses a beat packet it will cause an interruption in proper operation of the entire CobraNet® network. If an isochronous data packet is lost, a 1⅓ ms dropout will occur only in the audio carried by that particular bundle. A single such dropout may be inaudible or may make a "tick" in the audio. Large numbers of dropouts may sound like distortion.

42.12.7 Bundle Identification

Audio is carried over CobraNet® networks in bundles. Bundles may contain from zero to eight audio channels. Each bundle consists of a stream of packets of one of three types specified in Table 42-8.

Table 42-7. CobraNet® Packet Timing and Performance Requirements for the Ethernet Network

Parameter	Minimum	Maximum	Typical	Comments
Isochronous Cycle Interval			1333 µs	Future CobraNet® revisions may allow other cycle interval options.
Beat Packet Length	5.12 µs	121.4 µs	≤10 µs	Beat packet grows as bundle count increases.
Data Packet Length	5.12 µs	121.4 µs	100 µs	Size dependent on audio resolution and number of audio channels carried in bundle.
Reservation Packet Length	5.12 µs	121.4 µs	10 µs	
Inter-Packet Spacing	0.96 µs		5 µs	
Beat Packet Delay Variation	0 µs	250 µs		Normal delay distribution assumed.
Forwarding Delay	0 µs	400 µs		Assume maximal packet length when calculating store and forward delay (if applicable). Includes delay variation, i.e., 750 µs forwarding delay + 250 µs maximal positive excursion due to delay variation = 1000 µs. A higher forwarding delay can be tolerated on networks with small delay variation. If the forwarding delay specification is exceeded, additional delay is automatically added to the audio in increments of 64 sample periods (1 ⅓ ms).

Make sure your Ethernet switch vendor will guarantee that their switches in the configuration you propose will meet the above delay variation and forwarding delay specifications

Table 42-8. CobraNet® Bundle Types

Hexadecimal Bundle Number	Decimal Bundle Number	Designation	Usage	Transmission Addressing	Transmission Mode
0	0	Null	Unused bundle. Disables transmission/ reception when selected.	Never transmitted.	Never transmitted.
1–FF	1–255	Multicast	Publicly available bundles. Each bundle is transmitted by a single unit and may be received by any number of units.	Always multicast.	Always transmitted.
100–FEFF	256–65279	Unicast	Publicly available bundles. Each bundle is transmitted by a single unit. If the default unicast mode setting is used, it will only be received by a single unit.	Generally unicast, but may multicast if *txUnicastMode* variable is so adjusted.	Only transmitted when at least one receiver is identified via reverse reservation.
FF00–FFFF	65280–65535	Private	Individual transmitters locally allocate private bundles. The bundle number is conditioned on the transmitter's MAC. There are 256 of these bundles per transmitter thus the total number of private bundles is virtually unlimited.	Generally unicast, but may multicast if *txUnicastMode* variable is so adjusted.	Only transmitted when at least one receiver is identified via reverse reservation.

The bundle number specifies the type of bundle. There can only be a single transmitter for any given multicast or unicast bundle number at a time on the same network or VLAN. The three bundle types have different characteristics. Multicast bundles will be routed to every port in a switched network, or to every port in a VLAN, so use them sparingly. It is generally suggested that no more than four multicast bundles be used at any one time in a given switched network or VLAN within a switched network

42.12.8 Multicast Bundles

In a given network or VLAN there can only be a single instance of a given multicast bundle number at a time. The conductor will only allow one CobraNet® transmitter to be active during any isochronous cycle on a given multicast bundle number.

Multicast bundles are always multicast addressed, and are always transmitted even if no receiver has selected that bundle. Since they are multicast, they will appear on every port of a network or VLAN, even on switched networks. Therefore, the receiver does not have to submit a reverse reservation request to the conductor in order to receive a multicast bundle since the bundle will always appear at its input.

Caution must be used with multicast bundles and switched networks so as not to overwhelm the ports with multicast traffic. It is generally suggested to not use more than four multicast bundles on a given switched network or VLAN at the same time.

Multicast bundles can serve as a common denominator to allow interoperability with CobraNet® devices, which can only be configured from their front panel switches.

42.12.9 Unicast Bundles

In a given network or VLAN there can only be a single instance of a given unicast bundle number at a time. The conductor will only allow one CobraNet® transmitter to be active during any isochronous cycle on a given unicast bundle number.

Unicast bundles may be either unicast or multicast addressed based on the transmitter's reception of one or more reverse reservation requests for its bundle number. The *txUnicastMode* variable is used to control the transmitter's ability to switch to multicast on a unicast bundle number. The default setting of the *txUnicast-Mode* variable disables the ability to transmit multicast on a unicast bundle number. With the default setting, if more than one receiver requests a given unicast bundle number, only the first receiver to get its reverse reservation request in will get that bundle. With the default setting, unicast bundles can't be used for point to multipoint routing, instead multicast bundles must be used.

Some CobraNet® devices allow the same audio to be transmitted on more than one bundle at a time. This can provide an alternative way for a single CobraNet® device to unicast to as many as four receiving devices at the same time.

Unicast bundles are only transmitted if a receiver has requested that bundle. This allows a receiver to select which of many sources it wishes to receive, and only the source selected will transmit onto the network.

42.12.10 Private Bundles

Individual CobraNet® transmitters® control their own private bundles. Unlike multicast and unicast bundles, there may be more than one private bundle with the same bundle number on a network or VLAN at the same time. This is because a private bundle is specified using the transmitter's unique MAC address in addition to the bundle number.

Private bundles may be either unicast or multicast addressed based on the transmitter's reception of one or more reverse reservation requests for its bundle number and MAC address. The *txUnicastMode* variable is used to control the transmitter's ability to switch to multicast on a private bundle number. The default setting of the *txUnicastMode* variable disables the ability to transmit multicast on a private bundle number. With the default setting, if more than one receiver requests a given private bundle number from the same transmitter, only the first receiver to get its reverse reservation request in will get that bundle. With the default setting, private bundles can't be used for point to multipoint routing, instead multicast bundles must be used.

Some CobraNet® devices allow the same audio to be transmitted on more than one bundle at a time. This can provide an alternative way for a single CobraNet® device to send private bundles to as many as four receiving devices at the same time.

Private bundles are only transmitted if a receiver has requested that bundle. This allows a receiver to select which of many sources it wishes to receive, and only the source selected will transmit onto the network.

42.12.11 Bundle Assignments

Over CobraNet®, all audio channels are packaged into groups called bundles for transmission over the Ethernet network. The usual assignment is eight audio channels of 20 bit depth into one bundle. This is the maximum size possible, although using less audio channels is possible. In general for most efficient utilization of network bandwidth, maximum-size bundles are suggested. In the rest of this section we will be talking about maximum-size bundles. If 24 bit audio channels are used the maximum is seven audio channels packaged into a single bundle due to the maximum allowable Ethernet packet size.

A CobraNet® system is coordinated by one of the devices called the conductor. When two or more CobraNet® devices are interconnected properly, one of the devices will be elected the network conductor based on a priority scheme. The conductor indicator will light on the CobraNet® device that is serving as the conductor.

Each CobraNet® device has the ability to send and receive a fixed number of bundles. The bundle number tells the CobraNet® conductor which specific CobraNet® device is trying to communicate with which other CobraNet® device(s) over the network. Use of bundle numbers removes the necessity of the user having to tell the devices the Ethernet hardware (MAC) addresses of the other devices with which it is trying to communicate. As long as the CobraNet® devices are all set to the same bundle number, the CobraNet® system takes care of all the rest of the technical details of setting up an audio path over Ethernet between the devices.

A given bundle may have only one transmitter that places it onto the network. Unicast bundles may have only a single receiver. Multicast bundles may have multiple receivers.

In an ordinary Ethernet data network it is possible to mix both repeater hubs and switches and have the network continue to work. This is not the case with CobraNet®! For a CobraNet® network, you must either use all repeater hubs, or all switches in the network. This is because the CobraNet® protocol changes depending on which type of network it is operating over. However, non-CobraNet® devices may be attached to a switched CobraNet® network via repeater hubs.

On a repeater hub-based network, there is a fixed maximum of eight bundles per network. Any bundle may be placed onto the network from any port, and will appear at every other port on the network. The bundles usually used in a repeater hub network are numbered in the range from 1 to 255 decimal, and are called multicast bundles. Such bundles are always transmitted in a multicast mode, and may be received by any of the CobraNet® devices on the network.

As long as the limit of eight total bundles is not exceeded, it does not matter which channel numbers in the range of 1 to 65,279 are used.

It is not suggested to mix ordinary computer data on a repeater network with CobraNet®, as this could result in dropouts in the audio.

On a switched network, there is no fixed maximum number of bundles possible. The number will be determined by the network design. Again, bundles from 1 to 255 decimal are multicast bundles and, since they are multicast, will usually be sent to every port in the network. It is not suggested to use more than four multicast bundles in a switched CobraNet® network. There are special cases where more could be used, which we will go into later.

Bundles from 256 to 65,279 decimal are called unicast bundles. These are addressed to a single destination unit and are usually sent unicast. A switch will send these channels only out the ports leading to the CobraNet® device to which they are addressed. Unlike multicast bundles, unicast bundles will not be transmitted unless a receiver is requesting that bundle. This allows destination controlled routing, where the receiver selects one of several possible transmitters to receive, and only the selected transmitter is activated.

It is possible to have far more than eight total bundles active on a switched network if most of those channels are sent unicast using unicast bundles. A given port on a fast Ethernet switch can only send eight bundles out without running out of bandwidth. Those bundles will consist of every multicast bundle on the network, plus any unicast bundle addressed to a CobraNet® device connected either directly or through other switches to this port on the switch.

Some switches have gigabit Ethernet ports in addition to the fast Ethernet ports. The gigabit ports can be used to transfer data between switches with 10 times the bandwidth of a fast Ethernet port and can carry ten times as many bundles as fast Ethernet can. Gigabit Ethernet also transfers data at ten times the speed of fast Ethernet, and thus can have as little as 1/10 the forwarding delay. This can become very important in larger networks.

Unlike repeater hub-based networks, CobraNet® over a switched network does allow coexistence with ordinary computer data on the same network, because there are no collisions with the audio. There is the possibility that CobraNet® traffic on the network will cause problems for 10Mbit/s Network Interface Cards (NICs) used for computer data traffic. Recall that multicast bundles are sent to all switch ports in the same network. Since 8 bundles will fill a fast Ethernet (100Mbit/s) switch port, if that port is connected to a 10Mbit/s NIC (most fast Ethernet switch ports are dual speed 10/100 ports) then it is easy to see that multicast data from CobraNet® can saturate the 10Mbit NIC and make it drop the computer data packets it needs.

There are several possible solutions: one easy solution is to upgrade the NIC to 100Mbit/s full duplex.

1. Another possibility is to use little if any multicast bundles.
2. Most managed switches have multicast filtering features. These

allow you to exclude multicast traffic from a specified port. If your data is carried by the Internet protocol (IP), it is usually safe to filter all multicast traffic except the FF:FF:FF:FF:FF:FF destination address used by the address resolution protocol (ARP) associated with IP.

3. Obviously separate physical networks for audio and data will solve the problem. Separate networks can also be created using VLANs, which are supported by most managed switches. All traffic in a given VLAN, even multicast traffic, is isolated to only those ports which are part of the VLAN. You can typically partition up to eight different VLANs, and assign ports to them as you wish. Uplink ports used to connect two switches can be connected to multiple VLANs, and the traffic from those VLANs is multiplexed onto that link, and then demultiplexed at the other end.

VLANs can also be used in some cases when you need to use more multicast bundles than is allowable on a given CobraNet® network. By splitting the network into two virtual networks you have the ability to run twice as many multicast bundles.

Another solution that can be used with some CobraNet® devices, is transmitting the same audio information on two, three, or four unicast bundles to specific destinations instead of a single multicast bundle. Please note that not all CobraNet® devices have this capability. Some devices can only transmit two bundles, while others can transmit four. Some devices only accept eight audio inputs, while others accept sixteen. Obviously if a device accepts sixteen audio inputs and can only transmit two bundles, it can’t use this technique.

Also be aware that different CobraNet® devices can receive different numbers of bundles, and select only certain audio channels from those bundles to use or output.

Follow this procedure when designing a CobraNet® network:

1. Make a list of all the audio sources and their locations.
2. For each source, list the destination(s) to which it needs to go.
3. Group the audio sources at a location into bundles with no more than eight audio channels in a given bundle (or seven if 24 bit).
4. Determine if each bundle can be unicast, or if it must be multicast.
5. Make sure you don't have more than four multicast bundles in a network. If you need more than four multicast bundles:

   - Consider using multiple switched networks or VLANs.
   - Consider transmitting several unicast bundles instead of one multicast bundle.
   - Use the following rules to see if you can send more than four multicast bundles on a given network or VLAN:
   - Carefully map the number of bundles sent to each port of the system. The total of multicast and unicast Bundles arriving at each switch port may not exceed eight.
   - If a half-duplex device that can only transmit two bundles, and is set to transmit using both its bundles is part of the network, then you must make sure that the network conductor is not transmitting a multicast bundle. This may require changing the default conductor priority of one or more devices in the system to assure this condition is met.
   - Map the bundles carried by every link in the system to make sure that the limit of 8 bundles each direction on a given fast

Ethernet connection is not exceeded.

42.12.12 CobraCAD®

Fortunately there is an easier way to do steps 5 and 6 above. CobraCAD® can be downloaded for free from the Cirrus Logic Web site. CobraCAD® is a new software tool that provides a simple graphical user interface for the design and configuration of CobraNet® networks.

It allows you to draw your proposed CobraNet® network design using any of the CobraNet® devices on the market as of when the version you are using was released. You may also use any of a large selection of Ethernet switches.

After drawing the physical Ethernet interconnections, you next draw the bundle connections between the CobraNet® devices.

Then just press the Design Check button, and CobraCAD® will perform a design rule check. Designs that pass this check are extremely likely to work in the real world. There are still a few things CobraCAD® can't check for, so be sure to read the information and disclaimers in the Help system.

You may also want to check Cirrus Logic's CobraNet® Web site at: http://www.cobranet.info/ for the most recent version.

42.12.13 CobraNet® Hardware

At the heart of the CobraNet® interface, as shown in Fig. 42-48, is the digital signal processor, or DSP. It runs the software that together with the hardware provide all CobraNet® and Ethernet functions. It stores all the audio and Ethernet information as needed in stacks in the SRAM, converts the incoming synchronous audio into isochronous packets for transmission over the network, and converts isochronous packets from the network back into synchronous audio outputs. The DSP provides all interface functions to and from the device in which it is installed, and controls all other parts of the CobraNet® interface including the sample clock.

The sample clock is a voltage controlled crystal oscillator (VCXO), which is under the control of the DSP. If the CobraNet® interface is serving as the conductor, the sample clock is fixed in frequency and serves as the master clock for the network. In all other interfaces on the network, the sample clock is adjusted by the DSP so that it locks to the frequency of the network master clock.

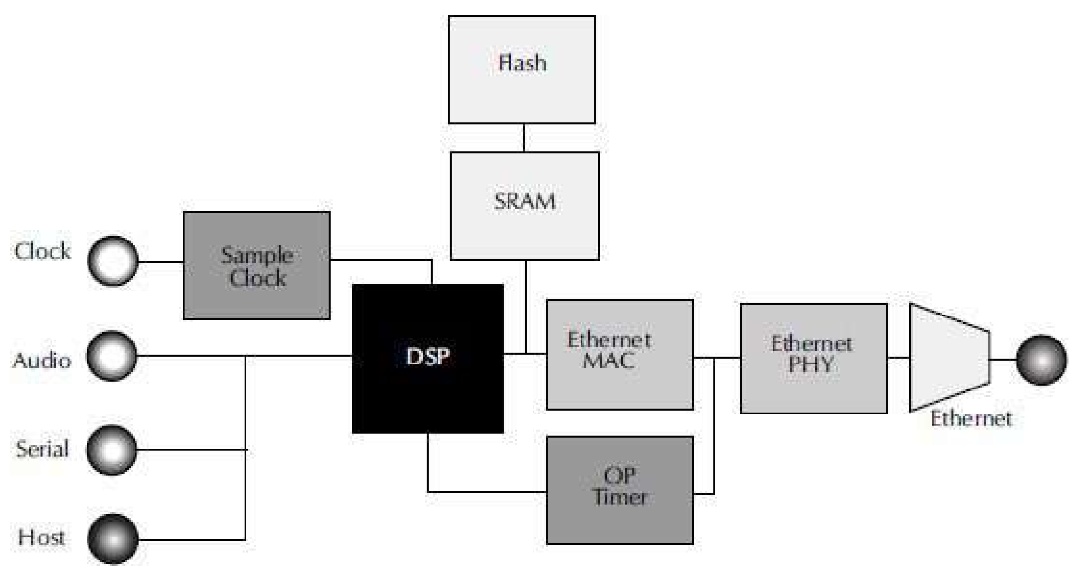


Figure 42-48. CobraNet® interface hardware. From the right there is the Ethernet network connection, and the isolation transformers. The PHY is a chip that acts as the Ethernet physical interface. The MAC is the Ethernet media access controller. The PHY and MAC chips together with the transformers constitute a standard fast Ethernet interface. The Flash memory serves as nonvolatile storage for the DSP firmware and the management

variable settings. SRAM provides memory for the DSP processor. All the audio and Ethernet buffers are located here. The DSP (digital signal processor) is the heart of the CobraNet® interface and provides all control and processing functions. The sample clock is controlled by the DSP and either serves as the master clock or locks to the Conductor's master clock over the network. The OP timer controls transmission of packets onto the network.

The CobraNet® interface provides its clock signal to the device it is part of, but can also receive a clock signal from the device and use that signal as the network master clock if the interface is the conductor.

The CobraNet® interface can provide up to thirty two synchronous digital audio signals to the device, and accept up to thirty two synchronous digital audio signals from the device for transmission across the network.

The serial port can accept serial data which is then bridged across the network and appears at all other CobraNet® devices on the network.

The host interface allows bidirectional communication and control between the CobraNet® interface DSP and the processor of the host device in which the interface is located. Detailed information on the connections and signals on the CobraNet® interface to the host are available in the CobraNet® Technical Datasheet found in pdf form on the Cirrus Logic's CobraNet® Web site at: http://www.cobranet.info/.

42.13 Aviom

Aviom uses the Physical Layer of Ethernet. In other words it is transported over Cat 5e cable with RJ-45 connectors. It does not use any other parts of Ethernet, but instead uses its own protocol. Aviom says on its Web site, “A-Net manages data differently than Ethernet does,” which has advantages as well as disadvantages. Aviom at this time offers two different versions of its technology, Pro16 and Pro64. Pro16 is limited to point to point connections while Pro64 is more flexible. Pro64 allows up to sixty four audio channels and lets all devices see all sixty four channels. The Aviom protocols are low latency and simple allowing inexpensive but effective products such as its personal monitor mixers, which have revolutionized personal in-ear monitoring onstage and in studios. A single Cat 5e cable to a small box by each musician carries sixteen audio channels and power, and allows the musician to make his own monitor mix exactly as he wishes.

42.14 EtherSound

EtherSound, as the name implies, does comply with the 802.3 Ethernet standard, but EtherSound networks are usually not built the same way Ethernet networks are. Ethernet networks are built using a star or star of stars topology, where each edge device connects to a switch port. Other than the simple case of a two device network, Ethernet edge devices do not directly connect to each other. Other than rare Ethernet edge devices with redundant ports, most Ethernet edge devices have only a single Ethernet port. Ethernet devices are never wired in a daisy-chain or cascade. EtherSound, on the other hand, provides in and out Ethernet ports on its edge devices, and in many cases builds networks by daisy-chaining its edge devices. This can result in simpler network designs, but also means that if a device fails in the middle of a daisy-chain it splits the rest of the devices into two isolated chains. EtherSound can also use switches in a more conventional star topology, but then devices downstream of the switch can't send audio back to the devices before the switch. The devices can be wired in a ring for fault-tolerance, and daisy-chain, star, and ring topologies can be mixed in the same network.

EtherSound has low latency and can support multiple sampling rates mixed in the same network. In order to get this low latency, EtherSound traffic must not be mixed with ordinary Ethernet traffic on the same network or VLAN. In EtherSound, 96kHz streams occupy two EtherSound channels, while 192kHz streams take four.

EtherSound ES-Giga has been introduced which expands the channel count from 64 to 256 channels per direction over a Gigabit Ethernet link.

42.15 Dante

A new entry into the digital audio networking world is Dante from Audinate. Unlike the other real-time digital audio networking protocols, Dante makes use of the new IEEE 1588 real-time clocking standard to solve many of the issues facing those who would use Ethernet for audio transport. Dante also uses the standard UDP/IP data transport standards. This allows it to use standard Ethernet ports on a computer, for example, instead of requiring dedicated hardware to interface a computer to the audio network. Dante supports multiple latencies, sampling rates, and bit depths in the same network. Unlike CobraNet which is a layer 2 transport limited to local area networks, Dante is a layer 3/IP transport allowing signals to be routed between subnets. Dante has added support for the new AES67 layer 3 protocol which was designed to allow interoperability between audio networking solutions from different manufacturers. It is expected that by the time this book is published many other companies will have added AES67 support to their audio networking offerings, thus bringing about a de facto industry standard.

42.16 Q-Lan

QSC Audio has introduced Q-Lan, a new layer 3 Ethernet-based digital audio networking scheme in some of its new products. It allows standard Ethernet ports on a computer to serve as audio transport ports, and does not require dedicated hardware for audio I/O. It is designed to take full advantage of Gigabit Ethernet and other advances in Ethernet technology, and to stay fully compatible with Ethernet as it evolves. Among the advantages it brings to audio networking are high channel counts, low latency, and the ability to operate over many switch hops. It uses automatic configuration techniques to greatly simplify the process of setting up an audio network and make it fast and easy.

Q-Lan supports redundant Ethernet networks with up to 512 audio channels into and 512 audio channels out of a single network connection, and no limit on the total audio channels on the network. Audio is carried at 32 bit resolution, and the latency across the network is 1ms, Fig. 42-49.

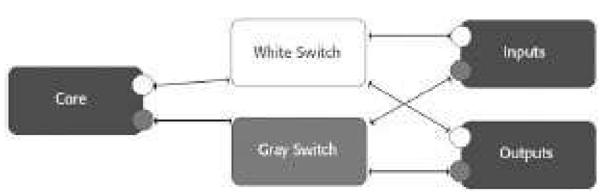


Figure 42-49. Q-Lan network with redundant Ethernet switches.

Q-Lan also supports redundant DSP processing cores and redundant input and output devices. Automatic changeover on failure is supported for both the DSP cores and the I/O devices. Redundant pairs of audio inputs and outputs are wired in parallel. Incoming audio drives both inputs of a redundant pair. The system relay disconnects the inactive audio output of a redundant pair so only one output actually drives the load, Fig. 42-50.

Ethernet switches for use with Q-Lan must be Gigabit Ethernet, and non-blocking. The switch must provide at least 80Kbyte total egress buffering available for audio class service with a minimum of 40Kbyte per port for this class. Forwarding decision time must be 10μs or less.

Q-Lan will operate on a layer 2 Ethernet network, but can take advantage of layer 3 capabilities such as IGMP, WAN multicast management, wirespeed IP routing and support for internet routing protocols if the network supports them.

Even though the Ethernet Standard does not support it, many vendors support “jumbo packets” which are larger than allowed by the Standard. If support for jumbo packets is enabled on a switch the network latencies for all packets will be increased beyond what is acceptable for Q-Lan. Fortunately most switches ship with jumbo packet support disabled by default.

Q-Lan requires DiffServ which is only available on managed Ethernet switches. Managed Gigabit Ethernet switches have become so common that the price differential from unmanaged Gigabit switches has become very small.

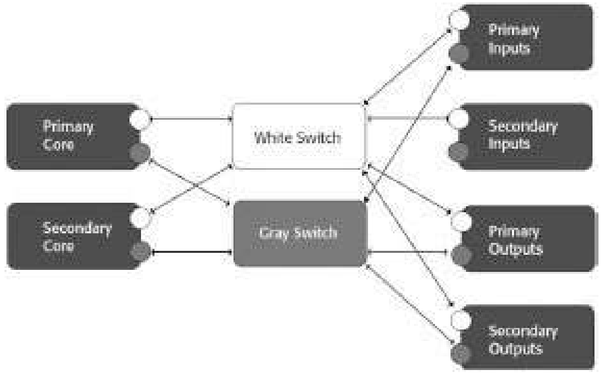


Figure 42-50. Fully redundant Q-Lan network with redundant DSP cores, redundant Ethernet switches, and redundant I/O devices.

Bibliography

AES-2id-2012—AES information document for digital audio engineering—Guidelines for the use of the AES3 interface.

AES3-2009 (AES3-1-2009, AES3-2-2009, AES3-2-2009, AES#-4-2009)—AES Recommended Practice for Digital Audio Engineering—Serial transmission format for two-channel linearly represented digital audio data.

AES-3id-2001 (r2006)—AES information document for Digital audio engineering—Transmission of AES3 formatted data by unbalanced coaxial cable.

AES5-2005 8 (r2013)—AES recommended practice for professional digital audio—Preferred sampling frequencies for applications employing pulse-code modulation.

AES10-2008—AES Recommended Practice for Digital Audio Engineering—Serial Multichannel Audio Digital Interface (MADI).

AES11-2009—AES Recommended Practice for digital audio engineering—Synchronization of digital audio equipment in studio operations.

AES18-1996 (r2002)—AES Recommended practice for digital audio engineering —Format for the user data channel of the AES digital audio interface.

AES42-2010—AES standard for acoustics—Digital interface for microphones.

IEC 60958-1 Ed3—Digital audio interface—Part 1 General.

IEC 60958-2 Ed3—Digital audio interface—Part 2 Serial copy management system.

IEC 60958-3 Ed3—Digital audio interface—Part 3 Consumer applications.

IEC 60958-4 Ed3—Digital audio interface—Part 4 Professional applications.

IEEE Std. 802.3, 2000 edition—Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications.

Brent Harshbarger, Digital Audio Networking Demystified, *Pro A/V Magazine*, 2007-08-01.

D. G. Kirby, Twisted Pair Cables for AES/EBU Digital Audio Signals, *J. Audio Eng. Soc.*, Vol. 43, no. 3, March 1995.

Kevin Gross, QSC Q-Sys Q-Lan White Paper—9 October 2009—Rev A.

Kevin Gross, CobraNet® Technology Datasheet, 11 April 2001, Version 1.1, Peak Audio.

Kevin P. Gross and Tom Holtzen, Controlling and Monitoring Audio Systems with Simple Network Management Protocol (SNMP), preprint 4760 (A-1), presented at the 105th AES Convention 1998 September 26–29, San Francisco, CA.

Robert A. Finger, AES3-1992: The Revised Two-Channel Digital Audio Interface, *J. Audio Engi. Soc.*, Vol. 40, no. 3, March 1992.

Chapter 43

Personal Monitor Systems

by Gino Sigismondi

43.1 Background

The emergence of modern sound reinforcement systems for music in the 1960s brought with it the need for performers to be able to better hear themselves onstage. Prior to the days of arena concerts and stacks of Marshall™ amplifiers, it wasn't that difficult for performers to hear vocals through the main PA loudspeakers. Most concerts were held in smaller venues, with a few notable exceptions. When the Beatles played Shea Stadium in 1964, the only PA was for voice; guitars were only as loud as the guitar amplifiers. Of course, the crowd noise was so loud even the audience couldn't hear what was going on, let alone the band! As rock and roll shows continued to get bigger and louder, it became increasingly difficult for performers to hear what they were doing. The obvious solution was to turn some of the loudspeakers around so they faced the band. A further refinement came in the form of wedge-shaped speakers that could be placed on the floor, facing up at the band, finally giving singers the ability to hear themselves at a decent volume, Fig. 43-1. With the size of stages increasing, it became difficult to hear everything, not just the vocals. Drums could be on risers 15 feet in the air, and guitar amps were occasionally stowed away under the stage. These changes required the use of a monitor console—a separate mixer used for the sole purpose of creating multiple monitor mixes for the performers—to accommodate all the additional inputs as well as create separate mixes for each performer. Today, even the smallest music clubs offer at least two or three separate monitor mixes, and it is not uncommon for local bands to carry their own monitor rig capable of handling four or more mixes. Many national touring acts routinely employ upwards of sixteen stereo mixes, Fig. 43-2.

Figure 43-1. Floor loudspeaker monitor wedge. Courtesy Shure Incorporated.

Figure 43-2. Large frame monitor console. Courtesy Shure Incorporated.

The problems created by traditional monitor systems are numerous; the next section examines them in detail. Suffice it to say, a better way to monitor needed to be found. Drummers have used headphones for years to monitor click tracks (metronomes) and loops. Theoretically, if all performers could wear headphones, the need for monitor wedges would be eliminated. Essentially, headphones were the first personal monitors—a closed system that doesn't affect or depend on the monitoring requirements of the other performers. Unfortunately, they tend to be cumbersome and not very attractive. The adoption of transducer designs from hearing aid technology allows performers to use earphones, essentially headphones reduced to a size that fit comfortably in the ear. Professional musicians, including Peter Gabriel and The Grateful Dead, were among the first to employ this new technology. The other major contribution to the development of personal monitors is the growth of wireless microphone systems. Hardwired monitor systems are fine for drummers and keyboardists that stay relatively stationary, but other musicians require greater mobility. Wireless personal monitor systems, essentially wireless microphone systems in reverse, allow the performer complete freedom of movement. A stationary transmitter broadcasts the monitor mix. The performer wears a small receiver to pick up the mix. The first personal monitor systems were prohibitively expensive; only major touring acts could afford them. As with any new technology, as usage becomes more widespread, prices begin to drop. Current personal monitor systems have reached a point where they are within many performers' budgets.

43.2 Personal Monitor System Advantages

The traditional, floor wedge monitor system is fraught with problems. Performers, especially singers, find it difficult to hear things clearly. Feedback is a constant and annoying issue. And the monitor engineer forever battles to keep up with the needs of the individual performers. Anyone who has performed live has probably dealt with a poor monitor system, but even a great system has many limitations due to the laws of physics, and those laws bend for no one. The concept of in-ear monitoring rose from the desire to create an onstage listening experience that could overcome the limitations imposed by a traditional floor monitor system.

Many parallels exist between personal monitors and a traditional floor wedge setup. The purpose of any monitor system is to allow performers to hear themselves. The sounds to be monitored need to be converted to electronic signals for input to the monitor system. This is usually accomplished via microphones, although in the case of electronic instruments such as keyboards and electronic drums, the signals can be input directly to a mixing console. The various signals are then combined at a mixer, and output to either power amplifiers and loudspeakers or to the inputs of personal monitor systems. More recently, personal monitoring mixers that allow performers to mix themselves on stage have become quite prevalent. Any amount of signal processing, such as equalizers or dynamics processing (compressors, limiters, etc.) can be added inbetween. A hardwired personal monitor system is similar (in signal flow terms) to a traditional wedge system, since the belt pack is basically a power amplifier, and the earphones are tiny loudspeakers. A wireless personal monitor system, however, adds a few more components, specifically a transmitter and receiver, Fig. 43-3. From the output of the mixer, the audio signal goes to a transmitter, which converts it to a radio frequency (RF) signal. A belt-pack receiver, worn by the performer, picks up the RF signal and converts it back to an audio signal. At this stage the audio is then amplified and output to the earphones.

The term personal monitors is derived from several factors, but basically revolves around the concept of taking a monitor mix and tailoring it to each performer's specific needs, without affecting the performance or listening conditions of the others. The concept is broader than that of in-ear monitoring, which states where the monitors are positioned, but gives no further information on the experience.

A. Sennheiser Evolution 300 series IEM.
Courtesy Sennheiser Electronic Corporation.

B. Shure PSM 900 system.
Courtesy Shure Incorporated.

Figure 43-3. Two wireless personal monitor systems.

The four most prominent benefits when using personal monitors are:

- Improved sound quality.
- Portability.
- Onstage mobility.
- Personal control.

43.3 Sound Quality

There are several factors that, when taken as a whole, result in improved sound quality with personal monitor systems. These factors include adequate volume for the performers, gain-before-feedback, hearing conservation, reduced vocal strain, and less interference with the audience mix.

43.3.1 Adequate Volume

The most common request given to monitor engineers is “Can you turn me up?” (Sometimes not phrased quite so politely.) Unfortunately, it is not always quite that simple. Many factors can limit how loud a signal can be brought up when using traditional floor monitors: size of the power amplifiers, power handling of the speakers, and most importantly, potential acoustic gain (see Gain-Before-Feedback below). Another factor that makes hearing oneself difficult is the noise level onstage. Many times, vocalists rely solely on stage monitors, unlike guitarists, bassists, and keyboardists whose instruments are generally amplified to begin with. Drummers, of course, are acoustically loud without amplification. Volume wars are not uncommon as musicians struggle to hear themselves over the ever-increasing din. The clarity of the vocals is often obscured as other instruments are added to the monitor mix, which becomes increasingly necessary if fewer mixes are available. Keyboards, acoustic guitars, and other instruments that rely on the monitors often compete with the vocals for sonic space. A personal monitor system, which isolates the user from crushing stage volumes and poor room acoustics, allows the musician to achieve a studio like quality in the onstage listening experience. Professional, isolating earphones, when used properly, provide more than 20dB of reduction in background noise level. The monitor mix can then be tailored to individual taste without fighting against otherwise uncontrollable factors.

43.3.2 Gain-Before-Feedback

More amplification and more loudspeakers can be used to achieve higher monitoring levels with traditional stage wedges, but eventually the laws of physics come into play. The concept of gain-before-feedback relates to how loud a microphone can be turned up before feedback occurs. Closely related is PAG, or potential acoustic gain. The PAG equation is a mathematical formula that can be used to predict how much gain is available in a sound system before reaching the feedback threshold, simply by plugging in known factors such as source-to-microphone distance and microphone-to-loudspeaker distance, Fig. 43-4. Simply stated, the farther away a sound source is from the microphone, or the closer the microphone is to the loudspeaker, or the farther away the loudspeaker is from the listener, then the less available gain-before-feedback. Now picture a typical stage. The microphone is generally close to the performer's mouth (or instrument); that's good. The microphone is close (relatively) to the monitor loudspeaker; that's bad. The monitor loudspeaker is far (relatively) from the performer's ears; that's also bad. Feedback occurs whenever the sound entering a microphone is reproduced by a loudspeaker and "heard" by the same microphone again. To achieve a decent monitoring level requires quite a bit of available gain. But given the above situation, two major factors drastically reduce the available gain-before-feedback. Compounding the problem is the issue of NOM, or number of open microphones. Every time you double the number of open microphones, the available gain-before-feedback drops by 3dB. With four open microphones onstage instead of one, the available gain has dropped by 6dB.

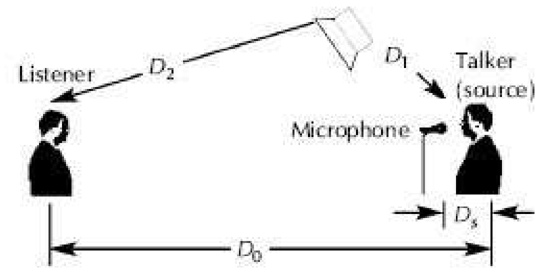


Figure 43-4. PAG values.

Solutions? The PAG equation assumes omnidirectional microphones, so using cardioid or even supercar-dioid pattern microphones will help; just don't point them at the speakers. Also, the equation assumes that the sound system has a perfectly flat frequency response. The most commonly employed tool for reducing feedback due to response problems is the graphic equalizer. Since some frequencies will feed back before others, an equalizer allows a skilled user to reduce the monitor system's output of those troublesome frequencies. This technique results in approximately 3–9dB of additional gain, assuming the microphone position doesn't change. It is common practice for some monitor engineers to attempt to equalize the monitor system to the point where there is no feedback, even with a microphone pointed right into the speaker cone. Unfortunately, the fidelity of the monitor is often completely destroyed in an effort to eliminate feedback using equalizers. Even after equalization has flattened the response of the monitor system, PAG again becomes the limiting factor. At this point, the microphone can't be moved much closer to the sound source, and moving the loudspeaker closer to the performer's ears also makes it closer to the microphone, negating any useful effect on PAG.

Personal monitoring completely removes PAG and gain-before-feedback issues. The "loudspeakers" are now sealed inside the ear canal, isolated from the microphone. With the feedback loop broken, it is possible to achieve as much volume as necessary—which leads to the next topic.

43.3.3 Hearing Conservation

If there's an overriding theme in switching performers to personal monitors, it's so that they can hear themselves better. But it doesn't do any good if eventually they can't hear at all. As mentioned earlier, volume wars on stage are a universal problem. Prolonged exposure to extremely high sound pressure levels can quickly cause hearing to deteriorate. Some performers have taken to wearing ear plugs to protect their hearing, but even the best ear plugs cause some alteration of frequency response. Personal monitors offer a level of hearing protection equal to that of ear plugs, but with the additional benefit of tiny loudspeakers in the plugs. The monitoring level is now in the hands of the performer. If it seems to be too loud, there is no excuse for not turning the monitors down to a comfortable level. The use of an onboard limiter is strongly recommended to prevent high-level transients from causing permanent damage. In larger, complex monitor rigs, compression and/or limiting is often employed to offer a greater degree of control and protection. These types of processors are more or less standard on newer consoles, and should be taken advantage of when possible.

NOTE: Using a personal monitor system does not guarantee that the user will not or cannot suffer hearing damage. These systems are capable of producing levels in excess of 130dB SPL. Prolonged exposure to these kinds of levels can cause hearing damage. It is up to the individual user to be responsible for protecting his or her own hearing. Please see the section Safe Listening with Personal Monitors for more information.

Reduced Vocal Strain. Closely related to the volume issue, the ability to hear more clearly reduces vocal strain for singers. In order to compensate for a monitor system that does not provide adequate vocal reinforcement, many singers will force themselves to sing with more power than is normal or healthy. Anyone who makes a living with their voice knows that once you lose it, you lose your livelihood. Every precaution should be taken to protect your instrument, and personal monitors are a key ingredient in helping vocalists continue to sing for years to come. (See Adequate Volume, previously discussed.)

Interference with the Audience Mix. The benefits of personal monitors extend beyond those available to the performer. An unfortunate side-effect of wedge monitors is spill from the stage into the audience area. Although directional at high frequencies, speaker cabinets radiate low-frequency information in a more or less omnidirectional manner. This situation aggravates the already complex task facing the FOH (front-of-house) engineer, who must fight against loud stage volumes when creating the audience mix. The excessive low frequencies coming off the backs of the monitors make the house mix sound muddy and can severely restrict the intelligibility of the vocals, especially in smaller venues. But eliminate the wedges, and the sound clears up considerably.

43.3.4 Portability

Portability is an important consideration for performing groups that travel, and for installations where the sound system or the band performance area is struck after every event. Consider the average monitor system that includes three or four monitor wedges at roughly 40 pounds each, and one or more power amplifiers at 50 pounds—this would be a relatively small monitor rig. A complete personal monitor system, on the other hand, fits in a briefcase. Purely an aesthetic consideration, removing wedges and bulky speaker cables from the stage improves the overall appearance. This is of particular importance to corporate/wedding bands and church groups, where a professional, unobtrusive presentation is as important as sound quality. Personal monitors result in a clean, professional-looking stage environment.

43.3.5 Mobility

Monitor wedges produce a *sweet spot* on stage, a place where everything sounds pretty good, Fig. 43.5. If you move a foot to the left or right, suddenly things do not sound as good anymore. The relatively directional nature of loudspeakers, especially at high frequencies, is responsible for this effect. Using personal monitors, though, is like using headphones—the sound goes where you go. The consistent nature of personal monitors also translates from venue to venue. When using wedges, room acoustics play a large part in the overall quality of the sound. Since professional earphones form a seal against ambient noise, acoustics are removed from the equation. In theory, given the same band with the same members, the monitor settings could remain virtually unchanged, and the mix will sound the same every night.

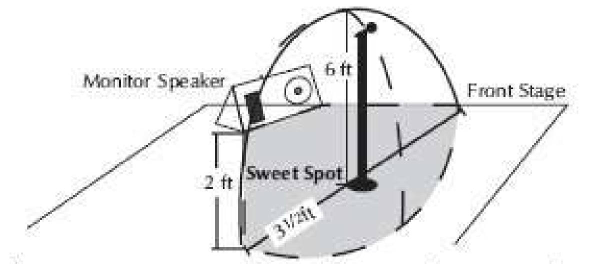


Figure 43-5. Sweet spot created by a wedge monitor loudspeaker.

43.3.6 Personal Control

Perhaps the most practical benefit to personal monitors is the ability for performers to have direct control over what they are hearing. While still relying on the sound engineer to make fine adjustments, personal monitor systems give the performer some ability to make broad adjustments, such as overall volume, pan, or the ability to choose different mixes. If everything in the mix needs to be louder, instead of giving a series of complex hand gestures to the monitor engineer, the performer can raise the overall volume directly from the belt pack.

Many professional systems utilize a *dual-mono* scheme, where the belt pack combines the left and right audio channels of a stereo system and sends the combined signal to both sides of the earphones, Fig. 43-6. The inputs to the system should now be treated as "Mix 1" and "Mix 2" instead of left and right. The balance control on the receiver acts as a mix control, allowing the performer to choose between two mixes, or listen to a combination of both mixes with control over the level of each. Panning to the left gradually increases the level of Mix 1 in both ears, while reducing the level of Mix 2, and vice versa. This feature is referred to by different names, such as MixMode™ (Shure) or FOCUS (Sennheiser), but the function is basically the same. Less expensive, mono-only systems can offer a similar type of control by providing multiple inputs at the transmitter, with a separate volume control for each. Consequently, the transmitter should be located near the performer for quick mix adjustments.

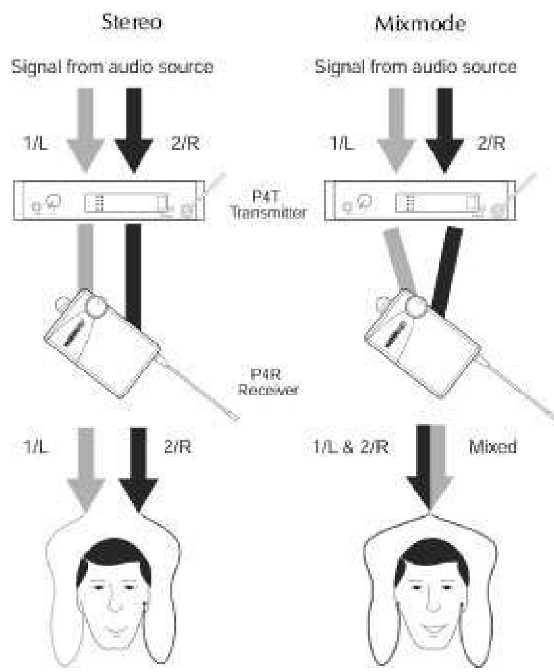


Figure 43-6. How dual-mono works. Courtesy Shure Incorporated.

Putting a small, outboard mixer near the performer increases the amount of control, Fig. 43-7. By giving control of the monitor mix to the performer, the sound engineer can spend more time concentrating on making the band sound good for the audience instead of worrying about making the band happy. Distributed mixing, covered in section 43.8.2, allows for simple routing of multiple audio channels to a large number of personal mixers. Some digital consoles even allow for personal mixing to be done via mobile devices.

Figure 43-7. Aviom A360 Personal Mixer. Courtesy Aviom, Inc.

The cost of transitioning to personal monitors has recently dropped dramatically. A basic system costs as much, if not less than, a typical monitor wedge, power amplifier, and graphic equalizer combination. Expanding a system is also more cost effective. When providing additional wedges for reproducing the same mix, a limited number can be added before the load on the amplifier is too great, and another amp is required. With a wireless personal monitor system, however, the number of receivers monitoring that same mix is unlimited. Additional receivers do not load the transmitter, so feel free to add as many receivers as necessary without adding more transmitters. For bands that haul their own PA, transportation costs may be reduced as well. Less gear means a smaller truck, and possibly one less roadie.

43.4 Choosing a System

Given the personal nature of in-ear monitoring, choosing the right system is an important step. Several choices are available. Present as well as future needs should be taken into account before making an investment.

Personal monitor systems come in two basic varieties—wireless or hardwired. A hardwired system requires the performer be tethered to a cable, which is not necessarily a negative. Drummers and keyboard players who stay relatively stationary, or even backup singers, can take advantage of the lower cost and greater simplicity of a hardwired personal monitor system. Simply connect the monitor sends to the inputs of the hardwired system and dial up a mix. Hardwired systems also work worldwide without the hassle of finding clear frequencies or dealing with local wireless laws and codes. Lastly, if several performers require the same mix, hardwired systems with sufficiently high input impedance can be daisy-chained together without significant signal loss. Alternately, a distribution amplifier could be used to route a signal to multiple hardwired systems. A distribution amplifier takes a single input and splits it to multiple outputs, often with individual level control.

Wireless equipment, by nature, requires special considerations and attention to detail. But the advantages many times outweigh the increased cost and complexity. One of the main benefits of personal monitors is a consistent mix no matter where the performer stands; going wireless allows full exploitation of this advantage. Additionally, when several performers require the same mix, hooking them up is even easier. As many wireless receivers as necessary can monitor the same mix with no adverse effects.

Secondly, consider the travel requirements, if any, of the users. Most wireless equipment, whether it is a microphone system or personal monitors, transmits on unused television channels. Since these unoccupied channels will be different in every city, it is imperative that appropriate frequencies are chosen. For a group that performs only in one metropolitan area, or for a permanent installation, once a good frequency is chosen, there should be no need to change it. However, for touring acts, the ability to change operating frequencies is essential.

The following important specifications for selecting wireless microphones also apply when selecting a personal monitor system:

- Frequency range.
- Tuning range (bandwidth).
- Number of selectable frequencies.
- Maximum number of compatible frequencies.

Other Wireless Considerations

As the amount of available spectrum in the UHF television band has continued to shrink, carefully selecting and coordinating frequencies amongst all wireless devices is critical. Most wireless systems, including personal monitors, have built-in scanning, which allows the system to quickly and reliably set itself to a clear frequency. While appropriate for moderately-sized applications, this method assumes no other wireless devices (microphones, intercom, etc.). When microphones, personal monitors, and possibly intercoms, are all used simultaneously, they must be coordinated as a group. Software tools, such as Shure Wireless Workbench 6 or Professional Wireless Systems IAS, are essential tools for large scale frequency coordination.

It is often desirable to separate personal monitor systems and wireless microphones into different frequency ranges within the UHF spectrum. This is known as “band planning.” Each of these types of systems has some limited range of frequencies it can tune through, known as the “band,” and isolating high-power personal monitor transmitters will typically allow a larger number of wireless systems to operate reliably than if they were all overlapping.

Finally, even when using software, obtaining scan data within the venue will provide the most accurate results for coordination purposes. This way, it is possible to see exactly what sort of interfering RF signals are present, and can be accounted for during coordination. PC-based scanning devices, such as the WinRadio, or the handheld TTi scanner, can export scan data in a format that is usable by certain software applications. Other devices, like the Shure AXT600 Spectrum Manager, can tie directly into Wireless Workbench as a real-time scanner.

43.5 Configuring a Personal Monitor System

Choosing the proper system requires some advance planning to determine the monitoring requirements of the situation. At a minimum, the three questions below require answers:

- How many mixes does the situation require?
- Will the monitor mix be stereo or mono?
- How many monitor mixes can be supplied by the mixing console?

This information directly relates to the equipment needed to satisfy the in-ear monitoring requirements of the performers. The following example details the thought process involved in deciding how to configure a system.

43.5.1 How Many Mixes Are Required?

The answer to this question depends on how many performers there are, and their ability to agree on what they want to hear in the monitors. For example, typical rock band instrumentation consists of drums, bass, guitar, keys, lead vocal, and two backup vocals provided by the guitar player and keyboardist. In a perfect world, everyone would want to listen to the same mix, so the answer to this question would be one mix. However, most real-world scenarios require more than one monitor mix. An inexpensive configuration uses two mixes, one consisting of vocals, the other of instruments. Using a system that features dual-mono operation, the performers individually choose how much of each mix they wish to hear, Fig. 43-8. This scenario is a cost-effective way to get into personal monitors, yet still requires a fairly good degree of cooperation among band members.

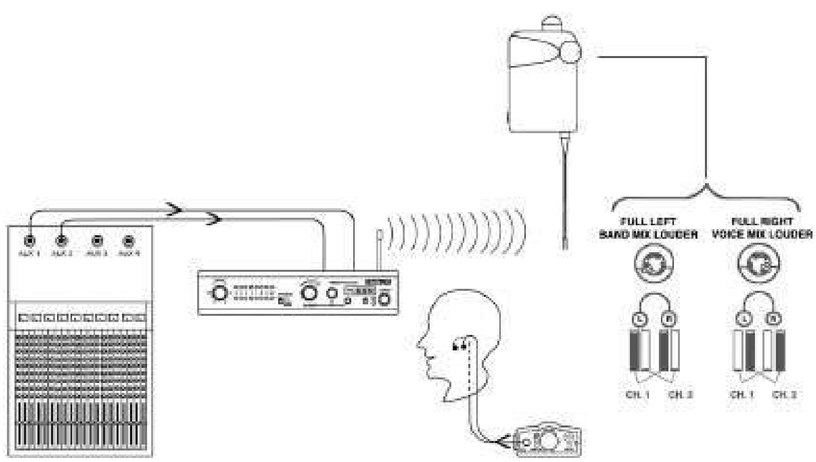


Figure 43-8. Two mixes, dual-mono.

Another scenario gives the drummer a separate mix, Fig. 43-9. This option works well for two reasons:

1. Drummers, in general, will want to hear considerably more drums in the monitors than other band members.

2. For bands who perform on small stages the drums are so loud that they are easily heard acoustically (with no additional sound reinforcement). Therefore, drums may not even be necessary in the other mixes. Now there are three mixes—the vocal mix, the instruments (minus drums), and the drummer's mix.

Up to this point, it is assumed that the vocalists are able to agree on a mix of the vocal microphones. While forcing singers to share the same mix encourages a good vocal blend, this theory commonly falls apart in practice. Often, separating out the lead vocalist to an individual mix will address this issue, and this can be handled in one of two ways. First, place some of the backup vocal mics in the instruments mix, and adjust the vocal mix to satisfy the lead singer, even if that means adding some instruments to the vocal mix. This scenario results in:

- An individual mix for the lead singer.
- A mix for the guitarist and keyboardist that includes their vocals.
- A drum mix (at this point the bass player can drop in wherever he or she wants, often on the drummer's mix).

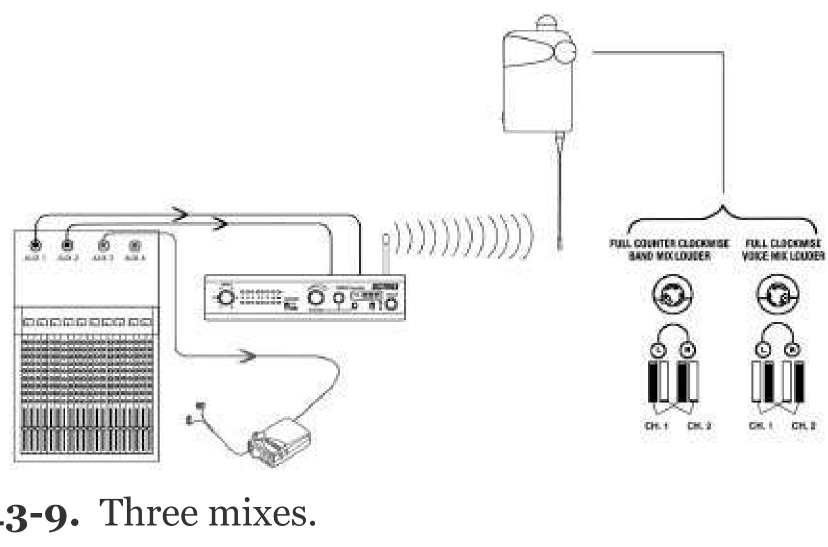


Figure 43-9. Three mixes.

The second option is to create a fourth mix for the lead singer, without affecting the other three. This configuration allows the guitarist and keyboardist to retain control between their vocals and instruments, while giving the lead singer a completely customized mix. Does the bass player need a separate mix? That is number five. Adding a horn section? That could easily be a sixth mix. More mixes can be added until one of two limitations is reached; either the mixer runs out of outputs, or the maximum number of compatible frequencies for the wireless monitor system has been reached. Depending on the application, at this point it may be desirable to use a distributed mixing system to put the mix responsibility in the hands of the performers.

43.5.2 Stereo or Mono?

Most personal monitor systems allow for monitoring in either stereo or mono. At first glance, stereo may seem the obvious choice, since we hear in stereo, and almost every piece of consumer audio equipment offers at least stereo, if not multichannel surround, capabilities. While it may not be applicable to all situations, especially with a limited number of mixes available, a monitor mix created in stereo can more accurately re-create a realistic listening environment. We spend our entire lives listening in stereo; logically, a stereo monitor mix increases the perception of a natural sound-stage. Monitoring in stereo can also allow for lower overall listening levels. Imagine a group with two guitar players sharing the same mix. Both instruments are occupying the same frequency spectrum, and in order for each guitarist to hear, they are constantly requesting their own level be turned up. When monitoring in mono, the brain differentiates sounds based only on amplitude and timbre.

Therefore, when two sounds have roughly the same timbre, the only clue the brain has for perception is amplitude, or level. Stereo monitoring adds another dimension, localization. If the guitars are panned, even slightly, from center, each sound occupies its own "space." The brain uses these localization cues as part of its perception of the sound. Research has shown that if the signals are spread across the stereo spectrum, the overall level of each signal can be lower, due to the brain's ability to identify sounds based on their location.

Stereo, by its very nature, requires two channels of audio. What this means for personal monitor users is two sends from the mixer to create a stereo monitor mix —twice as many as it takes to do a mono mix, Fig. 43-10. Stereo monitoring can rapidly devour auxiliary sends; if the mixer has four sends, only two stereo mixes are possible, versus four mono. Some stereo transmitters can be operated in a dual-mono mode, which provides two mono mixes instead of one stereo. This capability can be a great way to save money. For situations that only require one mix, such as solo performer, mono-only systems are another cost-effective option. Strongly consider a system that includes a microphone input that will allow the performer to connect a microphone or instrument directly to the monitor system.

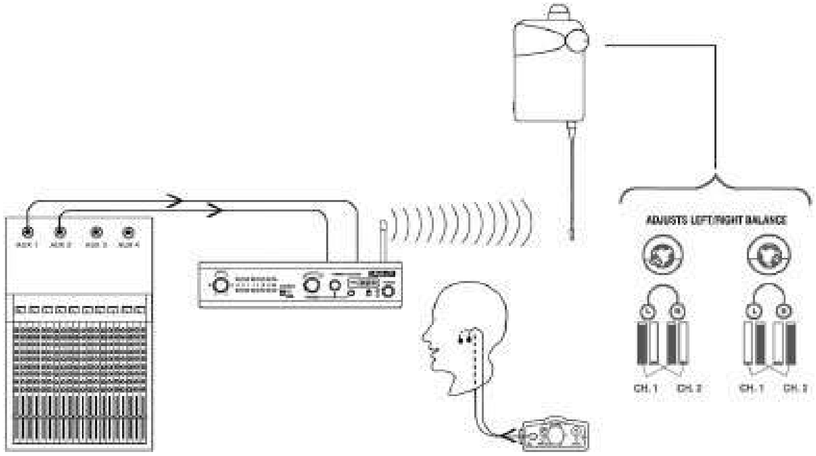


Figure 43-10. One stereo mix.

43.5.3 How Many Mixes Are Available from the Mixing Console?

Monitor mixes are typically created using auxiliary sends from the front-of-house (audience) console, or a dedicated monitor console if it's available. A typical small-format console will have at least four auxiliary sends. Whether or not all these are all available for monitors is another matter. Aux sends are also used for effects (reverb, delay, etc.). At any rate, *available auxiliary sends are the final determinant for the number of possible monitor mixes*. If the answer to question 1 (number of required mixes) is greater than the answer to question number 3 (number of mixes available), there are two options: reconfigure the required monitor mixes to facilitate sharing mixes with the existing mixer, or get a new mixer.

43.5.4 How Many Components Are Needed?

After answering the above questions, plug the numbers into the following equations to determine exactly how many of each component are needed and choose a system that can handle these requirements.

Stereo Mixes:
Number of transmitters = number of desired mixes.
Number of aux sends = 2 (number of transmitters), (ex. 4 mixes requires 4 transmitters and 8 aux sends).

Dual-Mono Mixes:
Number of transmitters = (number of desired mixes)/2
Number of required aux sends = 2(number of transmitters) (ex. 4 mixes requires 2 transmitters and 4 aux sends).

Mono Mixes:
Number of transmitters = number of desired mixes.
Number of aux sends = number of transmitters (ex. 4 mixes requires 4 transmitters and 4 aux sends).
Number of receivers = number of performers

43.6 Earphones

43.6.1 Earphone Options

The key to successful personal monitoring lies in the quality of the earphone. All the premium components in the monitoring signal path will be rendered ineffective by a low-quality earphone. A good earphone must combine full-range audio fidelity with good isolation, comfort, and inconspicuous appearance. The types of earphones available include inexpensive Walkman®-type ear-buds, custom molded earphones, and universal earphones. Each type has its advantages and disadvantages. While relatively affordable, ear-buds have the poorest isolation, are not really designed to withstand the rigors of a working musician's environment, and are likely to fall out of the ear. On the other end of the spectrum, custom molded earphones offer exceptional sound quality and isolation, a considerably higher price tag, and are difficult to try before buying since they are made specifically for one person's ears. The procedure for getting custom molds involves a visit to an audiologist. The audiologist makes an impression of the ear canals by placing a dam inside the ear to protect the eardrum, and fills them with a silicone-based material that conforms exactly to the dimensions of the ear canal. The impressions are then used to create the custom molded earphones. Another visit to the audiologist is required for a final fitting. Manufacturers of custom molded earphones include Ultimate Ears, Sensaphonics, and Future Sonics, Fig. 43-11

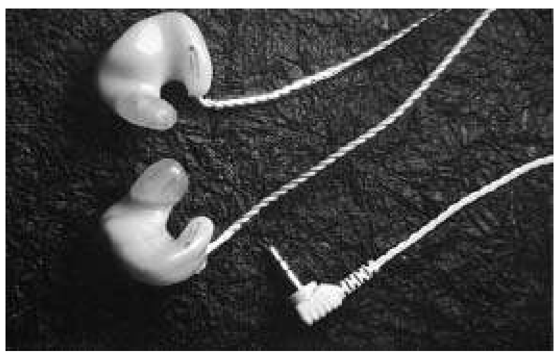


Figure 43-11. Custom molded earphones. Courtesy Sensaphonics.

A third type of earphone is the universal fit, Fig. 43-12. Universal earphones combine the superior isolation and fidelity of custom molded designs with the out-of-the-box readiness of ear-buds. The universal nature of this design is attributed to the interchangeable sleeves that are used to adapt a standard size earphone to any size and shape of ear canal. This design allows the user to audition the various sleeves to see which works best, as well as being able to demo the earphones before a purchase is made. The different earphone sleeve options include foam, rubber flex sleeves, rubber flange tips, and custom molded. The foam sleeves resemble regular foam earplugs, but with a small hole in the center of the foam lined with a tube of plastic. They offer excellent isolation and good low-frequency performance. On the downside, they eventually get dirty and worn, and need to be replaced. Proper insertion of the foams also takes longer—relative to the other options—since the earphone needs to be held in place while the foam expands. For quick insertion and removal of the earphones, flexible rubber sleeves may be a good choice. Made of soft, flexible plastic, flex sleeves resemble a mushroom cap and are usually available in different sizes. While the seal is usually not as tight as with the foams, rubber sleeves are washable and reusable. The triple-flange sleeves have three rings (or flanges) around a central rubber tube. They are sometimes referred to as Christmas trees based on their shape. The pros and cons are similar to that of the flex sleeves, but they have a different comfort factor that some users may find more to their liking. The fourth, and most expensive, option is custom sleeves. The custom sleeves combine the relative ease of insertion and permanency of flex sleeves with the superior (depending on the preference of the user) isolation of the foams. The process for obtaining custom sleeves for universal earphones is very similar to that of getting custom molded earphones; a visit to an audiologist is required to get impressions made. Custom sleeves also give the user many of the same benefits as custom molded earphones, but usually at a lower cost, and with the added benefit of being able to interchange earphones with the sleeves if they get lost, stolen, or are in need of repair. A final option, for users of the ear-bud type of earphone, is a rubber boot that fits over the earphone. This option typically has the poorest isolation.

Figure 43-12. Shure SE846 universal earphone. Courtesy Shure Incorporated.

If there is ever a problem with a universal earphone, another set can be substituted with no negative repercussions. A custom molded earphone does not allow for this kind of versatility; if one needs repair, the only alternative is to have a backup to use in the interim.

> *IMPORTANT NOTE: There are several brands of custom molded earplugs with internal filters that have relatively flat frequency response and different levels of attenuation. Although it may be physically possible to make universal earphones fit into the plugs with the filter removed, this is not advised. The location of the earphone shaft in the ear canal is crucial to obtaining proper frequency response, and most earplugs will prevent them from getting in the proper position. Once again, custom molded earplugs are NOT an acceptable alternative to custom sleeves.*

43.6.2 Earphone Transducers

The internal workings of earphones vary as well. There are two basic types of transducer used in earphone design—dynamic and balanced armature.

The dynamic types work on the same principle as dynamic microphones or most common loudspeakers. A thin diaphragm is attached to a coil of wire suspended in a magnetic field. Diaphragm materials include Mylar (in the case of dynamic microphones) or paper (for loudspeakers). As current is applied to the coil, which is suspended in a permanent magnetic field, it vibrates in sympathy with the variations in voltage. The coil then forces the diaphragm to vibrate, which disturbs the surrounding air molecules, causing the variations in air pressure we interpret as sound. The presence of the magnet-voice coil assembly dictates a physically larger earphone. Dynamic transducers are commonly used in the ear-bud types, but recent technological advances have allowed them to be implemented in universal designs. They are also found in some custom molded earphones.

Originally implemented in the hearing aid industry, the balanced armature transducer combines smaller size with higher sensitivity. A horseshoe-shaped metal arm has a coil wrapped around one end and the other suspended between the north and south poles of a magnet. When alternating current is applied to the coil, the opposite arm (the one suspended in the magnetic field) is drawn towards either pole of the magnet, Fig. 43-13. The vibrations are then transferred to the diaphragm, also known as the *reed*, usually a thin layer of foil. Balanced armature transducers are similar to the elements used in controlled magnetic microphones. In addition to the increased sensitivity, they typically offer better high-frequency response. Achieving a good seal between the earphone and the ear canal is crucial to obtaining proper frequency response.

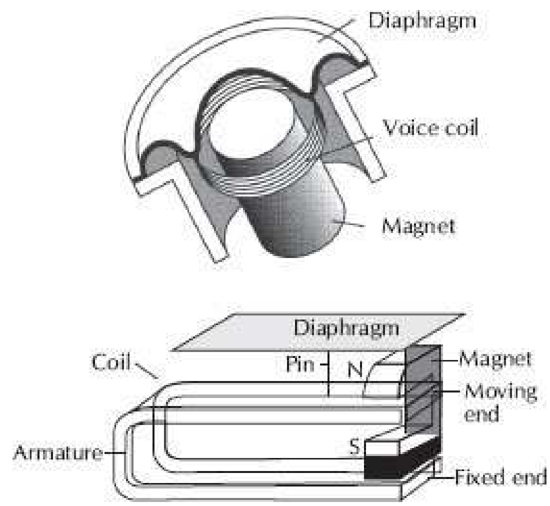


Figure 43-13. Dynamic and balanced armature transducer. Courtesy Shure Incorporated.

A further subdivision occurs with the use of multiple transducers. Dual transducer (*dual driver*) earphones are the most common. Another example of a loudspeaker with dual-transducer design is one with a horn (or tweeter) for high-frequency reproduction and a woofer for low-frequency sounds. The frequency spectrum is divided in two by a crossover network. Each driver only has to reproduce the frequency range for which it has been optimized. Dual driver earphones work on a similar principle—each earphone contains a tweeter and a woofer optimized for high- and low-frequency performance, respectively. Additionally, a passive crossover is built into the cable to divide the audio signal into multiple-frequency bands. The end result is usually much better low end, as well as extended high-frequency response. The additional efficiency at low frequencies may be of particular interest to bassists and drummers. Many companies have introduced triple-driver earphones and beyond, as well as hybrid earphones that combine both dynamic and balanced armature transducers in a single earphone. Although more doesn't always equal better, a well designed multi-driver earphones can offer the extended bass response of dual-driver earphones, but with better mid-range clarity. In the end, earphone choice is largely a personal decision, and often requires experimentation on the part of the user to find the solution that works best for them.

43.6.3 The Occluded Ear

One final note for users who are new to earphones. When the ear canal is acoustically sealed (occluded), the auditory experience is different from normal listening. For those performers who have spent many years using traditional floor monitors, an adjustment period may be necessary. A common side effect for vocalists is under-singing. The sudden shock of hearing oneself without straining causes some vocalists to sing softer than they normally would, making it difficult for the FOH engineer to get the vocals loud enough in the house mix. Remember, the FOH engineer is still fighting the laws of PAG, so singers still needs to project.

Another side effect of the occluded ear is a buildup of low frequencies in the ear canal. Sealing off the ear canal such as with an earplug, causes the bones of the inner ear to resonate due to sound pressure levels building up in the back of the mouth. This resonance usually occurs below 500Hz and results in a hollow sound that may affect vocalists and horn players. Recent studies have shown, however, that ear molds that penetrate deeper into the ear canal (beyond the second bend) actually reduce the occlusion effect. The deeper seal reduces vibration of the bony areas of the ear canal.

43.6.4 Ambient Earphones

Some users of isolating earphones complain of feeling closed off or too isolated from the audience or performance environment. While isolating earphones provide the best solution in terms of hearing protection, many performers would appreciate the ability to recover some natural ambience. There are several ways in which this can be accomplished, the most common being ambient microphones. Ambient microphones are typically placed at fixed locations, nowhere near the listener's ear, and the levels are controlled by the sound engineer instead of the performer. Additionally, the directional cues provided by ambient microphones (assuming a left/right stereo pair) are dependent on the performer facing the audience. If the performer turns around, the ambient cues will be reversed.

More natural results can be obtained by using a newer technology known as *ambient earphones*. An ambient earphone allows the performer, by either acoustic or electronic means, to add acoustic ambience to the in-ear mix. Passive ambient earphones have a port, essentially a hole in the ear mold, which allows ambient sound to enter the ear canal. Active ambient earphones use minuscule condenser microphones mounted directly to the earphones. The microphones connect to a secondary device that provides the user with a control to blend the desired amount of ambience into the personal monitor mix. Since these microphones are located right at the ear, directional cues remain constant and natural. Ambient earphones not only provide a more realistic listening experience, but also ease between-song communication amongst performers. A challenge is presented to the monitor engineer charged with providing a mix for the performer, in that the engineer will not hear exactly what the performer hears when making mix adjustments, unless he or she stands in exactly the same spot on stage. For some performers, though, the benefit of recovering some of the lost “live” feel is worth the trade-off. The performer should be made to understand this trade-off when requesting changes to the monitor mix

43.7 Applications for Personal Monitors

Configuring personal monitor systems and making them work is a relatively simple process, but the ways in which they can be configured are almost limitless. This section takes a look at several typical system set-up scenarios. Personal monitor systems are equally useful for performance and rehearsal, and their benefits extend from small nightclub settings to large arena tours to houses of worship.

43.7.1 Rehearsals

For groups that already own a mixer, implementing a system for rehearsals is a simple process. There are a number of ways to get signal into the system, depending on how many mixes are necessary. To create a simple stereo mix, simply connect the main outputs of the mixer directly to the monitor system inputs. (Note that this works just as well for mono systems). Auxiliary sends can also be used if separate mixes are desired. For bands that carry their own PA system (or at least their own mixer), this method allows them to create a monitor mix during rehearsal, and duplicate it during a performance. No adjustment needs to be made for the acoustic properties of the performance environment.

43.7.1.1 Performance, Club/Corporate/Wedding Bands—No

Monitor Mixer

The majority of performing groups do not have the benefit of a dedicated monitor mixer. In this situation, monitor mixes are created using the auxiliary sends of the main mixing console. The number of available mixes is limited primarily by the capabilities of the mixer. At a basic level, most personal monitor systems can provide at least one stereo or two mono mixes. Therefore, any mixer employed should be capable of providing at least two dedicated, prefader auxiliary sends. Prefader sends are unaffected by changes made to the main fader mix. Postfader sends change level based on the positions of the channel faders. They are usually used for effects. Although postfader sends can be used for monitors, it can be somewhat distracting for the performers to hear level changes caused by fader moves.

For users that only have two auxiliary sends available, the best choice is a system that allows a dual-mono operating mode, since this allows for the most flexibility. Hookup is straightforward—just connect Aux Send 1 of the console to the left input and Aux Send 2 to the right input. (Use Aux 3 and 4 if those are the prefader sends—all consoles are different!) Then, depending on who is listening to what, create the mixes by turning up the auxiliary sends on the desired channels. A few common two-mix setups are listed below.

Each performer can choose which mix they want to listen to by adjusting the balance control on the receiver. Be sure the receivers are set for dual-mono operation, or each mix will only be heard on the left or right side, but not in both ears. Also remember that any number of receivers can monitor the same transmitter.

Some performers may prefer to listen to the house mix, so they can monitor exactly what the audience is hearing. Keep in mind that this may not always produce the desired results. Rarely will what sounds good in the ear canal sound equally as good through a PA system in a less-than-perfect acoustic environment. Many times, a vocal that seems to blend just right for an in-ear mix will get completely lost through the PA, especially in a small room when live instruments are used. This technique may be appropriate for electronic bands, where the majority of instruments is input directly to the mixer. The only sound in the room is that created by the sound system.

The more auxiliary sends in a console, the greater number of monitor mixes that is possible. See Tables 43-1 to 43-3 for more examples.

Table 43-1. Three Monitor Mixes (MixMode™)

Option 1

Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Vocal mix	Band mix	Dedicated drum mix	Unused

Option 2

Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Lead vocal	Everything else	Dedicated drum mix	Unused

Option 3

Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Front mix	Backline mix	"Ego" mix (band-leader gets whatever he or she wants)	Unused

Table 43-2. Four Monitor Mixes (MixMode™—Using Only Aux Sends and PSM Loop Jacks)

Option 1			
Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Vocal mix	Band mix	Horn mix	Band mix (looped from PSM Right Loop Out Jack)

Table 43-3. Four Monitor Mixes (MixMode™)

Option 1			
Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Lead vocal-ist's mix	Guitarist's mix	Bassist's mix	Drummer's mix
Option 2			
Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
Vocal mix	Band mix	Horn mix	Vocal/band mix
Option 3			
Aux 1 Out (PSM 1 Left)	Aux 2 Out (PSM 1 Right)	Aux 3 Out (PSM 2 Left)	PSM 2 Right
"EGO" mix (lead vocal/ instrument only)	"Ego" mix (everything else)	Band mix	Dedicated drum mix

43.7.1.2 Club Level Bands—With Monitor Console

At this point, a typical small format FOH console has reached its limit for monitoring purposes. Bands that have graduated to the next level of performance (larger, more prestigious clubs and theaters or small tours) may find themselves in a position to take advantage of a dedicated monitor console. Most monitor boards are capable of providing at least eight mono (or four stereo) mixes. It now becomes practical for each band member to have his or her own dedicated mix. System hookup is again very simple—the various mix outputs from the monitor console are connected directly to the personal monitor system. Stereo monitoring is a much more viable option due to the large number of mixes available, as well as the presence of a skilled monitor engineer to adjust the mixes to the point of perfection.

Some performers even carry their own monitor console. Due to the consistent nature of personal monitors, a band with the same instrumentation and performers for every show can leave the monitor mix dialed-in on its console. Since venue acoustics can be completely disregarded, a few minor adjustments are all that is typically necessary during sound check.

43.7.1.3 Professional Touring System

When budget is no longer a consideration, personal monitoring can be exploited to its fullest capabilities. Many systems used by professional artists on large-scale tours often employ greater than sixteen stereo mixes.

A completely separate, totally personalized mix is provided for every performer onstage. Large frame monitor consoles are a requirement. For example, to provide sixteen stereo mixes requires a monitor console with thirty two outputs.

Effects processing is generally employed to a much larger extent than with a smaller system.

When operating a large number of wireless personal monitor systems, R-related issues become much more important. Frequency coordination must be done carefully to avoid interaction between systems as well as outside interference. Depending on the extent of the touring, a frequency agile system is desirable, if not required. Proper antenna combining, to reduce the number of transmitter antennas in close proximity, is a necessity. Directional antennas may also be used to increase range and reduce the chances of drop-outs due to multipath interference. It should be noted that high-gain directional antennas should be used with caution. Many tours have log periodic, or more commonly, helical antennas, located right at the mix position. This puts the antennas typically no more than 20 feet from the performers. A directional antenna is not needed for distances that short, and in some cases can actually make performance worse instead of better, by overloading the receiver front-end. Standard, half-wave, omnidirectional antennas, placed for good line-of-sight, are usually adequate. For larger arena shows, with giant stages and long thrusts, a directional antenna is more appropriate.

43.7.2 Mixing for Personal Monitors

Mixing for personal monitors may require a different approach than one used for a traditional monitor system. Often, the requirements for the performers are very involved, and can require a greater degree of attentiveness from the monitor engineer. In particular, many small nightclub sound systems typically provide monitors for the sole purpose of vocal reinforcement. An in-ear monitor system, due to its isolating nature, usually demands other sound sources be added to the monitor mix, especially if the instrumentalists choose to reduce their overall stage volume. Some performers may prefer a more active mix, such as hearing solos boosted, or certain vocal parts emphasized. This luxury usually requires a dedicated monitor console and sound engineer. The FOH engineer has enough responsibility mixing for the audience, and generally only alters the monitor mix on request from the performers. In most situations, except for the upper echelon of touring professionals, this approach is perfectly acceptable and still far superior to using wedges.

For performers who are mixing for themselves, there are other considerations. One of the advantages of having a professional sound engineer or monitor engineer is years of experience in mixing sound. This skill cannot be learned overnight. For bands that are new to personal monitors, there is a strong temptation to try to create a mix that rivals professionally produced recordings for in-ears. While this is certainly possible with a trained sound engineer and the right equipment, it is unlikely that someone unfamiliar with the basic concepts behind mixing will be able to successfully imitate a professional mix.

A common mistake made by in-ear monitor novices is to put everything possible into the mix. Here’s an alternative to the everything-in-the-mix method:

1. Put the earphones on and turn the system on. DO NOT put any instruments in the mix yet.
2. Try to play a song. While performing, determine what instruments need more reinforcement.
3. Begin bringing instruments into the mix, one at a time. Usually, vocals come first since those are often the only unamplified instruments onstage.
4. Only turn things up as loud as necessary, and resist the temptation to add instruments to the mix that can be heard acoustically.

A note on monitor mixing: performers now have an unprecedented level of personal control over what they are hearing. The temptation to make oneself the loudest thing in the mix is great, but this may not be the best for the situation. Proper blending with the other members of the ensemble will be next to impossible if the mix is skewed too far from reality. Consider big bands that normally play acoustically, or a vocal choir. These types of ensembles create their blend by listening to each other, not just themselves. If the lead trumpet player uses a personal monitor system, and cranks the trumpet up three times louder than everything else, there is no accurate reflection for the musician on whether he or she is playing too loud or too soft. Remember, great bands mix themselves—they don't rely entirely on the sound tech to get it right.

43.7.3 Stereo Wireless Transmission

Stereo personal monitoring systems employ a form of transmission known as MPX (multiplex). Stereo multiplexed wireless transmission has a limited frequency response of 50Hz–15kHz. This frequency response limitation has been in place since the FCC approved MPX in 1961. Audio engineers mixing stereo wireless transmissions for on-stage talent wearing in-ear monitors should be aware of the operating principles of MPX stereo to achieve the desired results at the receiver.

Stereo wireless transmitters use a steep cut filter, or brick-wall filter, prior to modulation, centered at 19kHz to create a safe haven for the required pilot tone. MPX encoders in stereo wireless transmitters use a 19kHz pilot tone to inform receivers that the transmission is encoded in stereo. If the receiver does not sense a 19kHz pilot tone, it will only demodulate a mono signal. Moreover, if the 19kHz pilot tone is not stable, stereo imaging degrades at the receiver. Most importantly, if in-ear monitor receivers do not sense stable 19kHz pilot tones, they will mute (this is called *tone-key squelch*, a circuit designed to keep the receiver muted when the corresponding transmitter is turned off). Problems are created due to the extensive EQ capabilities of modern mixing consoles, which offer high-frequency shelving equalization from as low as 10kHz to as high 12, 15, and even 16kHz. Digital mixing consoles offer parametric filtering that can center on practically any frequency and boost by as much as 18dB. With a multichannel mixing board, it is easy enough to create a counteractive frequency response at the frequency of interest—19kHz. In stereo wireless, there are two pieces of information actually being transmitted, the mono or sum signal (left + right) and the difference (left – right) channel, each occupying a 15kHz-wide swath of spectrum. The 19kHz pilot tone is centered in between these two signals, Fig. 43-14.

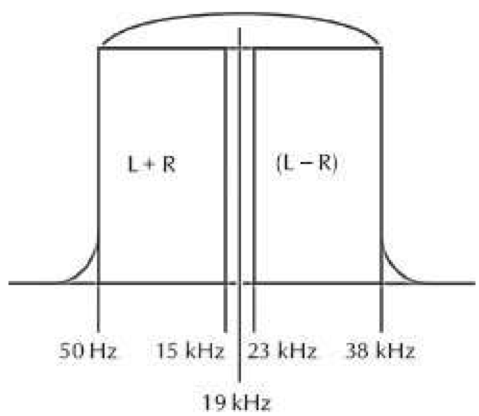


Figure 43-14. Stereo MPX encoding.

The stereo image is restored in the receiver by adding the sum and difference signals to create the left channel, and subtracting them to derive the right channel.

$$(L + R) + (L - R) = 2L \qquad (43\text{-}1)$$

$$(L + R) - (L - R) = 2R \qquad (43\text{-}2)$$

This system ensures mono compatibility, since the received signal will simply collapse to mono when the pilot tone is lost. Only the $L + R$ sum signal remains.

However, since the 19kHz pilot tone resides in the audio band, it can easily be compromised by the program material. The result of these high-frequency components getting into the modulator can cause, at best, degradation of stereo separation and distortion, and in worst-case situations, muting of the receiver. Add the high-frequency shelf used in the pre-emphasis curves prior to the companding circuits in stereo transmitters (a form of noise reduction), and it is easy to see how a small high-frequency boost on a channel strip can have a huge effect on what is heard after the RF link. If the audio signal modulates the pilot tone, stereo reception and the resultant sound quality will be poor. If upper harmonics of musical instruments aggravate the $(L - R)$ sidebands (especially in a transient manner—tambourines, triangles, high hats, click tracks, etc.), stereo separation can degrade, frequency response can be compromised, and even dynamic interactions between one channel and another can be detected.

Several simple practices go a long way toward improving stereo transmission:

- Refrain from extreme stereo panning. Instead of panning hard left and right, try the 10 o'clock and 2 o'clock positions.
- Use equalization sparingly prior to stereo transmission for smoother MPX encoding.

Newer systems may use digital stereo encoders that, combined with better filtering in the receiver, are less susceptible to the effects of high frequency content, and also offer wider stereo imaging.

43.7.4 Personal Monitors for Houses of Worship and Sound Contractors

The advantages of using personal monitors extend beyond those of just the performers. The above examples illustrate the benefits to the performer, and from a strictly music industry-oriented point of view. This section will discuss how personal monitors can be a useful tool for the sound contractor, specifically as they apply to modern houses of worship.

Musical performances have become an integral part of the modern worship service. Praise teams and contemporary music groups, while bringing new levels of excitement to traditional church services, also bring with them the problems of an average rock band. Most prominent among these problems are volume wars. Drums naturally tend to be the loudest thing on stage. The guitarist, in order to hear himself better, turns his amplifier up louder. The singers then need more monitor level to compete with the rest of the band. And then the cycle begins again. In any live sound situation, church or otherwise, loud stage volumes can distract from the overall sound in the audience. Try an easy experiment at the next sound check. When the band is satisfied with the monitor mix, turn off the audience PA and just listen to the sound coming off the stage. It's probably loud enough that the main sound system doesn't need to be turned on! To compound matters, the "backwash" off the floor monitors consists primarily of low-frequency information that muddies-up the audience mix. While this situation creates headaches for most sound engineers, it is even worse in the church environment. The majority of Sunday morning service attendees are not looking for an extremely loud rock and roll concert, but in some cases the congregation mix gets this loud just so it can be heard over the stage monitors. If the main system is off, and it's still too loud, what can be done? Turn down the floor monitors and the band complains—not to mention how terrible it will sound.

With the band using personal monitors, these problems evaporate. Traditional floor monitors can be completely eliminated. For part two of our experiment, turn off the stage monitors while the band is playing. Notice how much clearer the audience mix becomes? This is how it would sound if the band were using personal monitors. Also, personal monitors are not just for vocalists. Drummers with in-ear monitors tend to play quieter. When the loudest instrument on stage gets quieter, everything else can follow suit. Some churches take this a step further by using electronic drums, which create little, if any, acoustic noise. Bass, keyboard, and electric guitar can also be taken directly into the mixer if the players are using personal monitors, eliminating the need for onstage amplifiers. The end result is a cleaner, more controlled congregation mix, and musicians can have very loud monitors without affecting the congregation.

Secondly, consider the feedback issue. Feedback occurs when the sound created at the microphone comes out of a loudspeaker, and reenters the microphone. The closer the loudspeaker is to the microphone, the greater the chance for feedback. Eliminating the floor monitor also eliminates the worst possible feedback loop. With the "loudspeakers" sealed inside the ear canal, there is no chance for the signal to reenter the microphone. No equalizer or feedback reducer will ever be as effective as personal monitors at eliminating feedback on the stage.

Many other uses are possible for personal monitors. Choir directors could use them for cues, or to hear the pastor more clearly. Pastors who desire monitor reinforcement of their speech microphones, a sure-fire recipe for feedback, will find this a much better solution. Organists located at the rear of the sanctuary could use them to better hear the choir located up front, or also to receive cues. The advantages extend well beyond the benefits to the performer, and increase the overall quality of the service and the worship experience.

43.8 Expanding the Personal Monitor System

43.8.1 Personal Monitor Mixers

Personal monitoring gives the performer an unprecedented level of control. But for the performer who desires more than simple volume and pan operation, an additional mixer may be implemented. Personal monitor mixers are especially useful for bands who have a limited number of available monitor mixes, or who do not have a monitor engineer, or anyone at all to run sound. In a perfect world, all performers would be happy listening to the exact same mix; in reality, everyone may want to hear something different. A small mixer located near the performers allows them to customize their mix to hear exactly what they desire. Theoretically, any mixer can double as a personal monitor mixer, but most lack one key feature; the input signals need to find their way to the main (FOH) mixer somehow. Large sound systems with separate monitor consoles use transformer-isolated splitters to send the signals to two places, but these are prohibitively expensive for most working bands and small clubs. Y-cables can be used to split microphone signals, but they can get messy and are somewhat unreliable. A few manufacturers produce mixers with integrated microphone splitters. These range from basic four channel mixers with only volume and pan controls for creating a single mix to larger monitor consoles that can provide four or more stereo mixes along with fader control and parametric equalization.

43.8.2 Distributed Mixing

Distributed mixing is the direct result of advances in the area of digital audio networking. By converting analog audio signals to digital, audio can be routed to many locations without degradation or appreciable signal loss. Unlike with analog personal mixers, cabling is far simpler. Typically, analog outputs from a mixing console connect to an analog-to-digital converter. Multiple channels of digital audio can then be routed from the A/D converter to personal mixing stations located by each performer, using readily available Ethernet (Cat-5e/Cat-6) cable, thus eliminating a rat's nest of microphone cables or the large, unwieldy cable snakes required for analog audio distribution. Cat-5 cable is inexpensive and readily available. The mixing station provides an analog headphone output that can drive a set of isolating earphones directly, or better yet, connect to either a hardwired or wireless personal monitor system. If nothing else, the personal monitor system offers the advantage of a limiter for some degree of hearing protection, as well as a volume control at the performer's hip. The mixers supplied with most distributed systems do not always have a limiter. Most systems provide eight or sixteen channels of audio, allowing each performer to create his or her own custom mix, independent of other performers and without the intervention of a sound engineer. Note that giving this level of control to the performers will probably require some training in the basics of mixing to be successful (see Creating a Basic Monitor Mix above).

A further development in the area of distributed mixing can be directly attributed to the proliferation of smart, mobile technology. Some digital consoles offer built-in Wi-Fi, or the ability to connect to a wireless router, to allow mixing on mobile devices. This frees the mix engineer to make adjustments from anywhere in the room. A logical extension of this capability is to provide each band member with a mobile device that connects via Wi-Fi to the console, enabling each performer to adjust their own mix. In this case, no audio is routed to the mobile device, it is only used as a control surface. The transmitters can then be located near the mixing console for direct connection. This is a distinct advantage over hardwired distributed mixing systems, which require the transmitter be located near the performer, since each individual mixing station needs a direct connection to the transmitter in question. Additionally, most performers come equipped with their own mobile device, reducing the need for expensive mixing stations, Fig. 43-15.

Figure 43-15. QMix Aux-Mix Control Software. Courtesy Presonus.

Both of these technologies go a long way towards solving one of the major challenges for newcomers to personal monitor systems, which is the "infrastructure" that is required to get each performer the desired mix. The wireless components are really only half the equation. Just as a traditional floor wedge monitor system requires a mixing console to provide the right mix of sources to the performers, the same holds true for personal monitors. And since personal monitors do not contribute to the rest of the listening environment and are free from feedback concerns, giving mix control to the performers is a much more reasonable solution than with loudspeakers on the stage.

43.8.3 Supplementary Equipment

In-ear monitoring is a different auditory experience from traditional stage monitoring. Since your ears are isolated from any ambient sound, the perception of the performance environment changes. There are several other types of audio products that can be added to a personal monitor system to enhance the experience, or try to simulate a more “live” feel.

43.8.3.1 Drum Throne Shakers

Something performers may miss when making the transition to personal monitors are the physical vibrations created by amplified low-frequency sounds. Drummers and bass players are particularly sensitive to this effect. Although using a dual driver earphone usually results in more perceived bass, an earphone cannot replicate the physical sensation of air moving (sound) anywhere but in the ear canal. Drum shakers exist not to provide audible sound reinforcement, but to re-create the vibrations normally produced by subwoofers or other low-frequency transducers. Commonly found in car audio and cinema applications, these devices mechanically vibrate in sympathy with the musical program material, simulating the air disturbances caused by a loud subwoofer, Fig. 43-16. They can be attached to drum thrones or mounted under stage risers.

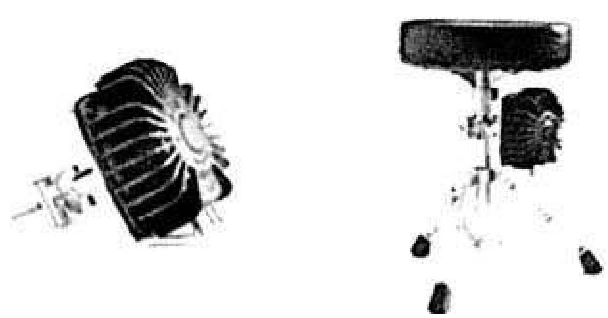


Figure 43-16. Aura Bass Shaker. Courtesy AuraSound, Inc.

43.8.3.2 Ambient Microphones

Ambient microphones are occasionally employed to restore some of the "live" feel that may be lost when using personal monitors. They can be used is several ways. For performers wishing to replicate the sound of the band on stage, a couple of strategically placed condenser microphones can be fed into the monitor mix. Ambient microphones on stage can also be used for performers to communicate with one another, without being heard by the audience. An extreme example (for those whose budget is not a concern) is providing each performer with a wireless lavalier microphone, and feeding the combined signals from these microphones into all the monitor mixes, but not the main PA. Shotgun microphones aimed away from the stage also provide good audience pick-up, but once again, a good condenser could suffice if shotguns are not available.

43.8.3.3 Effects Processing

Reverberant environments can be artificially created with effects processors. Even an inexpensive reverb can add depth to the mix, which can increase the comfort level for the performer. Many singers feel they sound better with effects on their voices, and in-ear monitors allow you to add effects without disturbing the house mix or other performers.

Outboard compressors and limiters can also be used to process the audio. Although many personal monitor systems have a built-in limiter, external limiters will provide additional protection from loud transients. Compression can be used to control the levels of signals with wide dynamic range, such as vocals and acoustic guitar, to keep them from disappearing in the mix. More advanced monitor engineers can take advantage of multiband compression and limiting, which allows dynamics processing to act only on specific frequency bands, rather than the entire audio signal.

In-ear monitor processors combine several of these functions into one piece of hardware. A typical in-ear processor features multiband compression and limiting, parametric equalization, and reverb. Secondary features, such as stereo spatialization algorithms that allow for manipulation of the stereo image, vary from unit to unit. Any of the plethora of digital consoles available now typically include a full complement of effects like the ones discussed above.

43.8.4 Latency and Personal Monitoring

An increasing number of devices used to enhance the personal monitor system are digital instead of analog. While the advantages of digital are numerous, including more flexibility and lower noise, any digital audio device adds a measurable degree of *latency* to the signal path, which should be of interest to personal monitor users. Latency, in digital equipment, is the amount of time is takes for a signal to arrive at the output after entering the input of a digital device. In analog equipment, where audio signals travel at the speed of light, latency is not a factor. In digital equipment, however, the incoming analog audio signal needs to be converted to a digital signal. The signal is then processed, and converted back to analog. For a single device, the entire process is typically not more than a few milliseconds.

Any number of devices in the signal path might be digital, including mixers and signal processors. Additionally, the signal routing system itself may be digital. Personal mixing systems that distribute audio signals to personal mixing stations for each performer using Cat-5 cable (the same cable used for Ethernet computer networking) actually carry digital audio. The audio is digitized by a central unit and converted back to analog at the personal mixer. Digital audio snakes that work in a similar manner are also gaining popularity.

Since the latency caused by digital audio devices is so short, the signal will not be perceived as audible delay (or echo). Generally, latency needs to be more than 35ms to cause a noticeable echo. The brain will integrate two signals that arrive less than 35ms apart. This is known as the Haas Effect, named after Helmut Haas who first described the effect. Latency is cumulative, however, and several digital devices in the same signal path could produce enough total latency to cause the user to perceive echo.

As discussed, isolating earphones are the preferred type for personal monitors, because they provide maximum isolation from loud stage volume. Isolating earphones, however, result in an effect known as the occluded ear. Sound travels by at least two paths to the listener's ear. The first is a direct path to the ear canal via bone conduction. An isolating earphone reinforces this path, creating a build-up of low frequency information that sounds similar to talking while wearing earplugs. Secondly, the "mic'ed" signal travels through the mixer, personal monitor transmitter and receiver, and whatever other processing may be in the signal path. If this path is entirely analog, the signal travels at the speed of light, arriving at virtually the same time as the direct (bone-conducted" sound. Even a small amount of latency introduced by digital devices, though, causes comb filtering.

Before continuing, an explanation of comb filtering is in order. Sound waves can travel via multiple paths to a common receiver (in this case the ear is the receiver). Some of the waves will take a longer path than others to reach the same point. When they are combined at the receiver, these waves may be out of phase. The resultant frequency response of the combined waves, when placed on a graph, resembles a comb, hence the term *comb filtering*, Fig. 43-17.

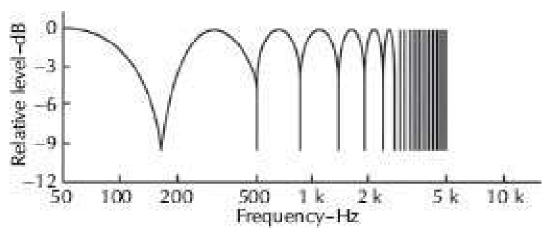


Figure 43-17. Comb filtering.

Hollow is a word often used to describe the sound of comb filtering.

It is generally believed that the shorter the latency, the better. Ultimately, changing the amount of latency shifts the frequency where comb filtering occurs. Even latency as short as 1ms produces comb filtering at some frequencies. What changes is the frequency where the comb filtering occurs. Lower latency creates comb filtering at higher frequencies. For most applications however, up to 5ms of delay is quite acceptable. When using personal monitors, though, total latency may need to be much less. While in reality it may be difficult to achieve latency shorter than a couple of milliseconds, be aware that any digital device will cause some latency. The individual user will have to determine what amount of latency is tolerable. As an alternative, some users report that inverting the polarity of certain input channels, or even the entire mix, improves the sound quality. Keep in mind that comb filtering still occurs, but at frequencies that may be less offensive to the listener.

The degree of latency is generally not more than a few milliseconds, which, as mentioned, will not cause the processed signal to be perceived as an audible delay. The concern for users of in-ear monitors, though, lies primarily with horn players, and occasionally vocalists. When a horn player sounds a note, the vibrations are carried directly to the ear canal via bone conduction. If the microphone signal is subject to digital processing, too much latency can cause comb filtering. The user generally perceives this as a hollow, unnatural sound. Care should be taken to avoid introducing unnecessary processing if comb filtering occurs. Adjusting the delay time in the processor (assuming digital delay is one of the available effects) could also compensate for latency. Alternately, route the effects through an auxiliary bus, rather than right before the monitor system inputs, which will minimize the latency effect by keeping the dry signal routed directly to the monitor system.

43.8.5 Safe Listening with Personal Monitors

No discussion of monitoring systems would be complete without some discussion of human hearing. The brain's ability to interpret the vibrations of air molecules as sound is not entirely understood, but we do know quite a bit about how the ear converts sound waves into neural impulses that are understood by the brain.

The ear is divided into three sections; the outer, middle, and inner ear, Fig. 43-18. The outer ear serves two functions—to collect sound and act as initial frequency response shaping. The outer ear also contains the only visible portion of the hearing system, the pinna. The pinna is crucial to localizing sound. The ear canal is the other component of the outer ear, and provides additional frequency response alteration. The resonance of the ear canal occurs at approximately 3kHz, which, coincidentally, is right where most consonant sounds exist. This resonance increases our ability to recognize speech and communicate more effectively. The middle ear consists of the eardrum and the middle ear bones (ossicles). This section acts as an impedance-matching amplifier for our hearing system, coupling the relatively low impedance of air to the high impedance of the inner ear fluids. The eardrum works in a similar manner to the diaphragm of a microphone, it moves in sympathy to incoming sound waves, and transfers those vibrations to the ossicles. The last of these bones, the stapes, strikes an oval-shaped window that leads to the cochlea, the start of the inner ear. The cochlea contains 15,000 to 25,000 tiny hairs, known as cilia, which bend as vibrations disturb the fluids of the inner ear. This bending of the cilia sends neural impulses to the brain via the auditory nerve, which the brain interprets as sound.

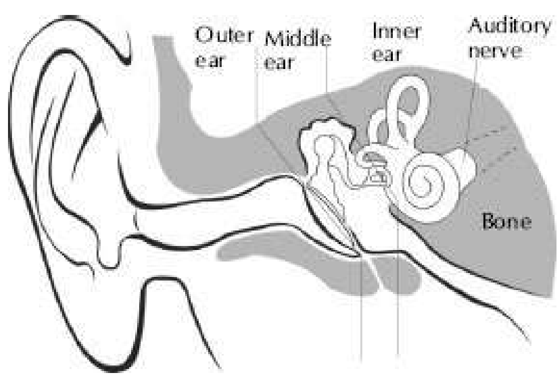


Figure 43-18. Illustration of ear anatomy. Courtesy Shure Incorporated.

Hearing loss occurs as the cilia die. Cilia begin to die from the moment we are born, and they do not regenerate. The cilia that are most sensitive to high frequencies are also the most susceptible to premature damage. Three significant threats to cilia are infection, drugs, and noise. Hearing damage can occur at levels as low as 90db SPL. According to OSHA (Occupational Safety and Health Administration), exposure to levels of 90dB SPL for a period of 8 hours could result in some damage. Of course, higher levels reduce the amount of time before damage occurs.

Hearing conservation is important to everyone in the audio industry. As mentioned before, an in-ear monitor system can assist in helping to prevent hearing damage —but it is not foolproof protection. The responsibility for safe hearing is now in the hands of the performer. At this time, there is no direct correlation between where the volume control is set and the sound pressure level present at the eardrum. Here are a few suggestions, though, that may help users of personal monitors protect their hearing.

43.8.5.1 Use an Isolating Earphone

Without question, the best method of protection from high sound pressure levels is to use a high-quality earplug. The same reasoning applies to an in-ear monitor. When using personal monitors, listening at lower levels requires excellent isolation from ambient sound, similar to what is provided by an earplug. Hearing perception is based largely on signal to noise. To be useful, desired sounds must be at least 6dB louder than any background noise. Average band practice levels typically run 110dB SPL, where hearing damage can occur in as little as 30 minutes. Using a personal monitor system with a nonisolating earphone would require a sound level of 116dB SPL to provide any useful reinforcement, which reduces the exposure time to 15 minutes.

Inexpensive ear buds, like those typically included with portable MP3 players, offer little, if any, isolation. Avoid these types of earphones for personal monitor applications.

Not all types of isolating earphones truly isolate, either. Earphones based on dynamic drivers typically require a ported enclosure to provide adequate low frequency response. This *port*, a small hole or multiple holes in the enclosure, drastically reduces the effectiveness of the isolation. Note that not all dynamic earphones require ports. Some designs use a sealed, resonating chamber to accomplish the proper frequency response, thus negating the need for ports but preserving the true isolating qualities of the earphone. Earphones that employ a balanced armature transducer, similar to those used in hearing aids, are physically smaller and do not require ports or resonating chambers. In fact, balanced armature-type earphones rely on a good seal with the ear canal to obtain proper frequency response. They can be made somewhat smaller, but are typically more expensive, than their dynamic counterparts.

43.8.5.2 Use Both Earphones!

A distressing, yet increasingly common, trend is only using one earphone and leaving the other ear open. Performers have several excuses for leaving one ear open, the most common is a dislike for feeling removed from the audience, but the dangers far outweigh this minor complaint. First, consider the above example of a 110dB SPL band practice. One ear is subjected to the full 110dB, while the other ear needs 116dB to be audible. Using only one earphone is equivalent to using a nonisolating earphone, except one ear will suffer damage twice as fast as the other. Second, a phenomenon known as *binaural summation*, that results from using both earphones, tricks the ear-brain mechanism into perceiving a higher SPL than each ear is actually subjected to. For example, 100dB SPL at the left ear and 100dB SPL in the right ear results in the perception of 106dB SPL. Using only one earphone would require 106dB SPL at that ear. The practical difference is potential hearing damage in one hour instead of two. Using both earphones will usually result in overall lower listening levels.

Table 43-4 shows OSHA recommendations for exposure time versus sound pressure level.

Table 43-4. OSHA Recommended Exposure Time Versus Sound Pressure Level

Sound Pressure Level	Exposure time
90 dB SPL	8 hours
95 dB SPL	4 hours
100 dB SPL	2 hours
105 dB SPL	1 hour
110 dB SPL	30 minutes
115 dB SPL	15 minutes

Ambient microphones are commonly employed to help overcome the closed-off feeling. An ambient microphone can be a lavalier clipped to the performer and routed directly to the in-ear mix, or a stereo microphone pointed at the audience. The common thread is that they allow the user to control the level of the ambience.

43.8.5.3 Keep the Limiter On

Unexpected sounds, such as those caused by someone unplugging a phantom-powered microphone or a blast of RF noise, can cause a personal monitor system to produce instantaneous peaks in excess of 130dB SPL, the equivalent of a gun shot at the eardrum. A brick wall-type limiter can effectively prevent these bursts from reaching damaging levels. *Only use a personal monitor system that has a limiter at the receiver, and do not defeat it for any reason*. A well-designed limiter should not adversely affect the audio quality, as it only works on these unexpected peaks. If the limiter seems to be activating too often, then the receiver volume is probably set too high (read as: unsafe!). Outboard compressors and limiters placed before the inputs of the monitor system are certainly appropriate, but are not a substitute for an onboard limiter, as they cannot protect against RF noise and other artifacts that may occur post-transmitter.

43.8.5.4 Pay Attention to What Your Ears Are Telling You

Temporary threshold shift (TTS) is characterized by a stuffiness, or compressed feeling, like someone stuck cotton in the ears. Ringing (or tinnitus) is another symptom of TTS. Please note, though, that hearing damage may have occurred even if ringing never occurs. In fact, the majority of people who have hearing damage never reported any ringing. After experiencing TTS, hearing may recover. Permanent damage has possibly occurred, though. The effects of TTS are cumulative, so a performer who regularly experiences the above effects is monitoring too loud and hearing damage will occur with repeated exposure to those levels.

43.8.5.5 Have Your Hearing Checked Regularly

The only certain way to know if an individual's listening habits are safe is to get regular hearing exams. The first hearing test establishes a baseline that all future hearing exams are compared against to determine if any loss has occurred. Most audiologists recommend that musicians have their hearing checked at least once a year. If hearing loss is caught early, corrections can be made to prevent further damage.

A frequently asked question about in-ear monitors is: “How do I know how loud it is?” At this time, the only way to develop a useful correlation between the volume knob setting and actual SPL at the eardrum is by measuring sound levels at the eardrum with specially made miniature microphones. A qualified audiologist (not all have the right equipment) can perform the measurements and offer recommendations for appropriate level settings.

Personal monitors can go a long way toward saving your hearing, but only when used properly. Monitoring at lower levels is the key to effective hearing conservation, and this can only be accomplished through adequate isolation. Used correctly, professional isolating earphones, combined with the consultation of an audiologist, offer the best possible solution for musicians interested in protecting their most valuable asset. It cannot be stated strongly enough: a personal monitor system, in and of itself, does not guarantee protection from hearing damage. However, personal monitors not only offer improved sound quality and convenience, but they also provide performers with an unprecedented level of control. Reducing stage volume also improves the listening experience for the audience, by minimizing feedback and interference with the house mix. As with most new technologies, an adjustment period is usually required, but few performers will ever return to floor monitors after using personal monitors.

43.8.5.6 Information on hearing health for musicians

For information contact:

- House Ear Institute, Hotline: (800) 388-8612, Web site: www.hei.org.
- H.E.A.R., Hotline: (415) 409-3277, Web site: www.hearnet.com.
- Sensaphonics Hearing Conservation, 660 N. Milwaukee Avenue, Chicago, IL 60622. Toll Free: (877) 848-1714, Int'l: (312) 432-1714, Fax: (312) 432-1738, Web site: www.sensaphonics.com, E-mail: saveyourears@sensaphonics.com.

Bibliography

Bob Ghent, "Healthy Hearing and Sound Reinforcement," *Mix*, Mar. 1994, pp. 150—162.

G. Gudmundsen, "Occlusion Effect," *Etymotic Research Technical Bulletin*, Oct. 2000.

Susan R. Hubler, "The Only Ears You've Got,"*Mix*, Oct. 1987, pp. 104–113.

Thom Fiegle, "Notch Filter Allows for Best Monitor Mix Ever!" http://shure.custhelp.com/app/answers/detail/a_id/32319, Oct. 2005.

Chapter 44

Message Repeaters, Museum/Tour Group Systems, Voice Evacuation/Mass Notification Systems

by Glen Ballou and Hardy Martin

44.1 Digital Audio Storage

There are numerous applications for digital audio storage, that for day-to-day use, and that which is preserved for future generations. We will first address the applications and workings of message repeaters as they are used on a daily basis. Secondly, we will discuss the specifics of archiving audio information.

44.2 Message Repeaters

The original repeater was a person sitting at a microphone making an announcement over a public address system at or about the right time. Messages could be changed at any time, and if area switching was available, different messages could be sent to different areas in real time. However different messages could not be sent at the same time. Another disadvantage was the requirement for dedicated personnel. In an emergency, an individual was required to stay at the microphone and announce in a calm and persuasive voice—a difficult thing to do at best.

With the introduction of tape recorders, messages could be prerecorded and played back manually or automatically. Unless a multichannel recorder or multiple recorders were used, only one message could be played at one time. To play different recorded messages required recording them in series and locating the desired message by fast forwarding the tape, time consuming in an emergency situation. With the design of lubricated tape, continuous loop tape recorders were used. This eliminated the requirement of rewinding but also meant the message could not be repeated until the tape had taken its course. Tape stretch and breakage was always a possibility. Auto-rewind cassette players were also used but had the same problem of reel-to-reel machines, they had to be rewound.

The late 1970s brought about the introduction of digital storage devices using solid state equipment. Digital message repeaters have numerous advantages over earlier systems including the following.

- Reliability.
- Flexibility.
- Solid state reproduction quality.
- User-recordable audio and text messages.
- User-programmable, locally or remotely.
- Scheduled messages,
- Remotely controllable.

44.3 How Message Repeaters Work

The use of digital signal processors (DSPs) has greatly simplified digital circuitry and made it possible to design digital message repeaters (see Chapter 35, *DSP Technology*).

A digital message repeater in its simplest form is shown in Fig. 44-1. In this system, a permanent message is digitized by the manufacturer and stored in the digitized message storage circuit. Changing the message may require sending the unit back to the manufacturer for reprogramming. New technology now allows messages to be updated without sending the unit back to the manufacturer by using memory which can be electronically erased and re-recorded on site. These systems use memory controlled by a microprocessor to record and play the stored message.

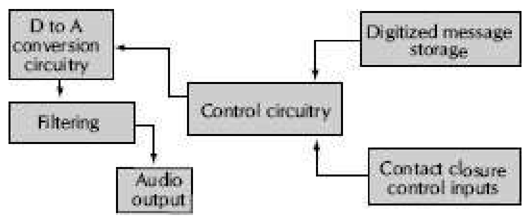


Figure 44-1. Digital message repeater.

Upon contact closure, the control circuitry transfers the digitally stored message to the digital to analog converter (D/A) circuitry where it is changed into an analog circuit. The analog signal is only approximate at this time so a filter is used to smooth the waveform and limit the frequency response. Normally this type of unit has a frequency response from 300Hz–3kHz or the basic telephone response. The filtered output is then directed to the audio output circuitry where it is amplified, made balanced or unbalanced, and matched for the proper output impedance.

An intelligent digital message repeater, such as an Instaplay™ by ALARMCO is shown in Fig. 44-2. In this system, messages can be recorded from a microphone (Mic), Aux. Input (AUX), e.g., a CD or MP3 player, or standard touch tone telephone (control Phone). The analog input is then filtered and converted to a digital signal (A/D conversion circuitry) and stored (digitized message storage). Several thousand messages can be stored and individually replaced at any time by using flash memory. Sound quality is assured by storing audio data using 16 or 24 bit samples in the Instaplay™. Audio and programming data may also be downloaded digitally from a PC. With memory and intelligent firmware, each new recording can be longer or shorter than the original with all unused memory available for other recordings.

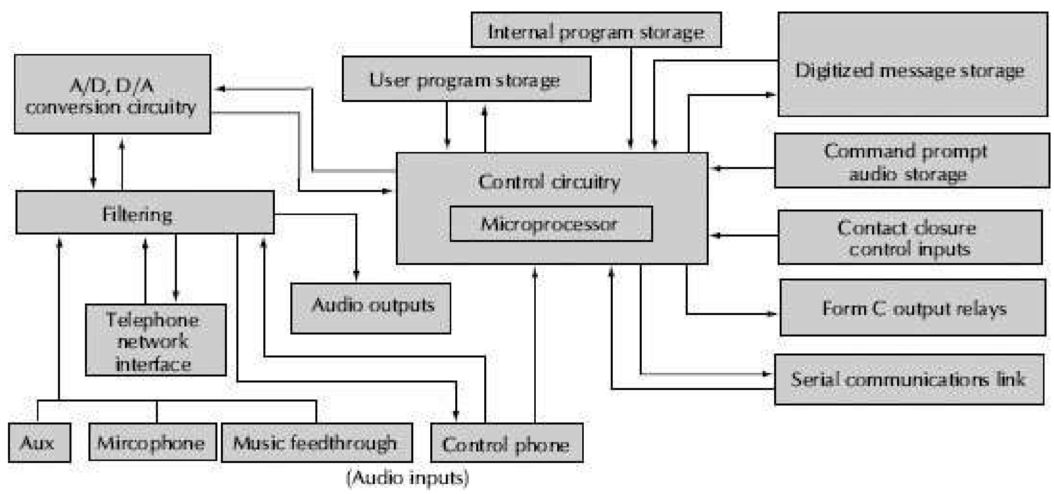


Figure 44-2. An intelligent message repeater. Courtesy ALARMCO.

To simplify recording from a control phone (either locally or remotely), *Control Phone* and *Telephone Network Interface*, prerecorded instructions are stored in the announcer in digitized form, *Command Prompt Audio Storage*. These instructions guide the user through each step. The telephone's keypad is used to respond to the prompts.

Installers can create a schedule for the playback of messages with the scheduler, *Internal Program Storage*, supplied by the manufacturer. Using this scheduler, the announcer can automatically switch from one playlist to another at various times throughout the day or week. The announcer insures not only that announcements are made, but also made on time. A text message can also be sent to a computer to keep a record of the event or a text message can be sent to a sign to display a visual message.

With an intelligent announcer, the installer can specify, *User Program Storage*, where a message should be played, *Audio Outputs*, as well as the time delay between the announcements over each output channel.

Instaplay™ always knows when it's talking. By activating one of the numerous relays, *Output Relays*, the installer is able to use the announcer to direct other activities, such as turning on lights or to trigger other Instaplays. The installer can control these relays by entries on a playlist, *User Program Storage*.

During playback, selected messages can be played in any order (including repeats), as specified by the installer, *User Program Storage*. Background music can be played between messages, *Music Feedthrough*, with many message repeaters.

An intelligent message repeater, such as the Mackenzie Dynavox however, can interrogate the software [*User Program Storage*] and decide whether to duck down or mute the background music during each individual message This system includes a built in timer, which the user can set to automatically play a message with pre-defined intervals.

With intelligent message repeating, messages can be triggered externally in numerous ways—*Contact Closure Control Inputs, Serial Communications Link, Ethernet Link, Control Phone, Remote Phone*—and to play a message sequence or queue new messages. In addition, intelligent repeaters allow customers to record new messages, modify schedules, playlists and other programming parameters either locally or remotely. When accessing the announcer remotely, *Telephone Network Interface* and *Ethernet Interface*, a security code can be employed. A playlist is a command list that runs when a control input is activated or an event is scheduled. Playlists queue the messages to play in a defined sequence and channel and include the ability to display the text messages, perform out-dial sequences, and operate external devices. Playlists can also start and stop messages from an external trigger.

Instaplay™ gets its native intelligence from its embedded firmware, *Internal Program Storage*. Supplied by the manufacturer, these programs not only define the internal operation of the machine, but also define default parameters such as how often to repeat a recorded message or how to act when a request is received.

Instaplay™ gets its application intelligence from the programming entered at the job site and stored in the announcer's random access memory, *User Program Storage*. The installer must easily be able to change these default parameters within the announcer to have it operate as the application dictates. Whereas a casino operation may want to announce hundreds of events throughout a particular day according to a predetermined schedule, the announcement of a train's arrival at the platform needs to correspond with the actual arrival time, rather than the scheduled arrival time.

The microprocessor must always coordinate both the default system values and the user-specified parameters to operate appropriately for different applications.

44.4 Message Repeater Usage

Message repeaters are used in many venues:

- Hospitals.
- Factories.
- Amusement Parks.
- Retail Stores.
- Schools.
- Tourist Attractions.
- Transportation Services.
- Information Providers/Broadcast Services.
- Message on Hold.
- Museums.
- Mass Notification.

44.4.1 Hospital Applications

Message repeaters are popular in hospitals. One common application is to broadcast different messages into separate locations in the hospital. The messages may be divided into all area announcements, or those that are announced only in public areas or patient areas. For example, visiting hour reminders are broadcast into all areas of the hospital, including patient's rooms. No smoking reminders might be announced only in lobbies, cafeterias, and waiting rooms, while doctor calls would only be announced in patient areas.

In addition, with a scheduler, message repeaters can be used to announce when visiting hours are about to end and when visiting hours begin again.

Hospitals around the country are trying unique ways of increasing their patient satisfaction. Many hospitals are now using message repeaters to play “Brahms Lullaby” when a baby is born. It adds a little smile to everyone’s face.

A message repeater can also be connected to controlled doors such as in geriatric and psychiatric wards to tell nurses of unauthorized door openings. Voice messages can announce immediately throughout the building which door is open, so personnel do not have to go to a central alarm panel to determine which door is ajar, a much more effective and faster method than visual indications.

When a Code Blue Dispatch happens, message repeaters can instantaneously repeat the message as often as required without requiring dedicated personnel. As hospitals grow in size, the communication between staff and visitors also increases. With larger message requirements a mass storage system is required. The Mackenzie M3 system can accommodate over 500 stored messages, with RS-232 control capability for fast and reliable message activation.

44.4.2 Factory Floor Applications

Just-in-time manufacturing is credited by some companies for documented savings of over one million dollars annually in reduced inventory expenses. By implementing just-in-time techniques, unused parts don’t require valuable and expensive floor space, nor does the production line ever slow due to lack of components. Many of the largest factories in the country, such as Motorola, General Motors, and Xerox, have been practicing this method of cost reduction for many years.

Message repeaters can be used to broadcast a message when certain parts need to be restocked. Messages can be triggered manually by having the assembly operator push a button, or automatically with a sensor to trip the message when the weight of the bin containing components becomes too light, sending a verbal message to stock.

Repeaters can also be used to announce lunch time, scheduled breaks, safety reminders, company announcements, and so forth. They can be configured to know when someone is entering (versus exiting) a hard hat area. A warning message can be announced on entering. A message reminding visitors to return their hard hats and protective eyewear can also be announced on exiting the area.

Another interesting use of a message repeater is the elimination of acoustical feedback in noisy environments. By recording the message or page on a message repeater and playing it back as soon as the recording is finished, the microphone-amplifier-loudspeaker-room feedback loop is broken and feedback is eliminated and pages can be automatically repeated.

44.4.3 Retail Store Applications

Message repeaters are ideal for in-store assistance requests. In-store assistance is becoming increasingly popular with retailers because it allows them to cut the expense of large staffs while still delivering service to customers who need it. Signs telling customers to "Press the button for assistance in this department" are often connected to a message repeater.

Convenience stores have a captive audience when it comes to customers who are pumping gas. Message repeaters can trigger different advertising messages when someone drives up to the gas pump. Studies have proven that sales increase dramatically with this type of advertising. Repeaters can be configured to play the appropriate message to shoppers as they enter or exit a store.

Message repeaters are used in some stores to do targeted in-store advertising. Special repeaters allow messages to influence the customers when they are near the point-of-purchase. Targeting is done by using sensors to trip the control circuit, thereby giving advertising messages when someone enters a particular aisle or department. Even more specific targeting can be accomplished, for example, when someone reaches for a brand name product, as detected by a motion sensor, a message can be played to try the store brand.

Customized modes of operation form an elegant solution for in-store use. Specific options could include individual timers that would allow each department manager to specify the amount of time before the message repeater transmits another message triggered from the same department. A second audio output channel allows office or security personnel to know when someone needs assistance in sensitive areas, such as someone in the area carrying a handgun. Message repeaters can allow any department zone to be enabled or disabled from a central location, such as the front office. If a child is playing with a button, the store manager can temporarily disable that button until the child moves on.

With a built-in scheduler, message repeaters can announce appropriate closing times each day. Sophisticated software allows a single intelligent message repeater to perform many of these functions simultaneously.

44.4.4 School Applications

School closing or shortened day announcements for snow days etc.

can be recorded and played back over the school PA system and the local town access TV channel. hotlines, movie theater schedules, sports score lines, and

Also class change announcements and activities announcements can be prerecorded and transmitted day or night automatically through the internal scheduler. A message repeater can also be set up as a dial-in line for sports schedules, latest news, and so forth.

Often school gyms or auditoriums are used after hours. Due to fire regulations, many do not have security gates to separate the used area from the rest of the school. A message repeater, triggered by a light beam or motion detector, can energize a message repeater to notify people that they are walking into a closed area, while simultaneously alerting personnel or security staff. In the event of a lock down situation, message repeaters quickly and reliably direct students and staff in safety measures.

44.4.5 Transportation Services Applications

To ease the burden on the driver, bus route stops can be manually selected, controlled, and played on a message repeater or they can be completely computer controlled. Message repeaters on tour buses are often used to eliminate the necessity of a tour director and to give a running dialogue of the tour route. Usually the driver controls the repeater by pushing a switch when he or she is ready for the next message to be announced.

Message repeaters on mass transit loading platforms can announce train arrivals and departures, safety messages, and upcoming schedule changes. The Statue of Liberty Ferry chose Alarmco's Instaplay™ to intelligibly announce tour information and required safety messages, Fig. 44-3.

44.4.6 *Information/Broadcast Services Applications*

USIA Voice of America chose Instaplay message repeaters for a number of European countries to retrieve messages that are broadcast over a worldwide satellite network. The broadcast messages are downloaded in various languages into the appropriate repeater. Local radio stations in the individual countries dial into the repeater and download the messages, which are in turn broadcast over the radio stations. Intelligent message repeaters have the ability to record remotely from a line level input, thereby allowing the remote telephone to control them, while they are recording from the satellite.

Because messages can be easily changed, they are ideal for radio broadcast directions, Traveler Advisory Radio (such as AlertAM from Information Station specialists), news broadcast services, and visitors information services. They are also used for employment weather information lines.

Figure 44-3. Instaplay message repeater. Courtesy ALARMCO.

44.4.7 *Message On-Hold Applications*

Telephone on-hold equipment is operating whenever we hear anything but silence when we are put on hold. The programming and messages for on-hold players can be recorded either locally or remotely. Local recording and programming requires someone on-site to record or load in the messages. Some of the more intelligent systems allow for remote downloading of the programming and messages using satellite systems, FM subcarrier audio channels, a standard telephone line, or the Internet. During playback, these remote download units have all the reliability advantages of other digital on-hold equipment. They have the additional advantage over conventional tape download equipment that they are completely hands-off at the installation site. They require no intervention from on-site store personnel, who may be unwilling or unable to load in the new audio messages.

Digital on-hold players employ memory chips to store the program. To produce a frequency response from 20Hz–20kHz would require a tremendous amount of memory. Telephone lines normally will not transmit a greater frequency response than 300Hz–3.5kHz, therefore it is not practical to increase the response and the equipment cost to cover a much wider range.

As is the case with CDs, digital on-hold units sample the incoming signal many times a second to store the signal into digital memory. Theoretically, the greater the samples per second, i.e., the sampling rate, the better the sound quality. Sampling rate is usually expressed in kilobits per second (Kbps). Toll quality telephone performance (the best performance any telephone network will allow) is 64Kbps so there is no need to produce on-hold units with a sampling rate greater than 64Kbps.

Sampling rate is only one measure of the audio quality of a digital downloadable on-hold unit; a network of filters and frequency compensators also contributes to the sound quality.

In most cases, nontechnical employees connect the on-hold equipment to the line and load new music/messages. Some units employ microprocessors to control all aspects of the download/play process.

The Bogen HSR series unit is an example of a full microprocessor-controlled on-hold system. Various models have a 4, 6, 8, or 12 minutes of memory capacity. The HSR's automatic operation assesses the start and stop point of the audio, sets record levels, downloads, and goes into play mode automatically. The unit also incorporates a one-play trigger mode for making a single message such as store closing.

The Mackenzie Laboratories, Inc. Dynavox series, Fig. 44-4 are on-hold systems that can also can be used as storecasters. One series has a 3.4kHz bandwidth for telephones and a second series has a 6.8kHz bandwidth for storecast and other wide-band requirements. The bit rate increases from 96 to 196Kbps and the sampling frequency increases from 8kHz to 16kHz with the increasing frequency response. Audio storage requires 16MB DRAM (dynamic random-access memory) to record 32min at 96Kbps and 16min at 196 Kbps. These units have a noise floor and dynamic range of greater than 70dB. The DVSD-3000 series uses internal compact SD Memory cards with storage capacity of 30+ minutes. Audio content is loaded via USB memory key directly to the internal SD memory.

The Mackenzie Laboratories, Inc., Dynavox Store-casting systems, Fig. 44-5 are used to play back overhead marketing messages mixed with background music systems. These systems include a built-in user adjustable timer to allow the message interval to be set on site for any given application. In a gas station or convenience store, the interval would be short 3–5min, while the interval in a supermarket could be 15–25min as shoppers are in these locations for longer periods of time. Audio performance is 15kHz using MP3 or WAV audio files. Messages are loaded directly to internal SD Memory via a removable USB key.

Figure 44-4. MacKenzie DSVD-3000 Message-On-Hold series repeater. Courtesy MacKenzie Laboratories, Inc.

Figure 44-5. MacKenzie DSVD-3000 MTS/LR Message-On-Hold series repeater. Courtesy MacKenzie Laboratories, Inc.

Intelligent message repeaters, such as the Instaplay™ series by

ALARMCO, can provide on-hold music and messages, storecasting messages, triggered customer assistance messages, and automatic store closing announcements—all with a single announcer. In addition, this type of repeater can also play existing background music through, thereby eliminating the need for customers to purchase a cassette with recorded music.

The Mackenzie Laboratories, Inc. M3 system, Fig. 44-6, is a high fidelity stereo or mono message repeater which can be used in a single or large capacity application, where a large number of messages or long playback times are required. Audio performance is 15kHz using MP3 or WAV files, and loading is via removable USB key.

Figure 44-6. MacKenzie M3 High Fidelity Multi-Channel Digital Audio Message System. Courtesy MacKenzie Laboratories, Inc.

The M3 module can store hours of audio using large capacity SD memory cards. The system includes an RS-232 command port allowing access to hundreds of messages using a basic ASCII command structure. This system lends itself to work with Programmable Logic Controllers (PLC's) and other logic systems such as Crestron controllers. The M3 can be configured to operate as a single or dual stereo output system which is useful for themed restaurants, amusement parks, and general background and foreground music playback requiring high quality and large storage capacity. Four internal M3 modules with 2 stereo channels each, for a total of 16 audio channels, can be enclosed in a single 1RU rack chassis.

Additional applications for the M3 are Hospital Code Blue, Airport announcements, Voice Warning and Evacuation system.

Process control systems require a large number of indicators to operate efficiently. Using a supervisory system to give commands directly to the M3 system when measurements are out of tolerance, means a message can be played to advise workers when temperatures are not within specification, equipment speed is not correct, or oil, gas and water pressures are wrong. Multiple levels of messages can be given as the systems run, from warnings to complete system shut down. The large number of stored messages allow the M3 to broadcast messages and provide a safe and efficient workplace.

44.5 Museum and Tour Group Systems

Many years ago, I spent rainy Sunday afternoons in museums. My father loved history, so we would stop at every artifact and read those infernal placards word for word. I vowed that I would never subject my family to those static and boring museums.

Fortunately, times have changed. Museums can be exciting with interactive displays and audiovisuals replacing the static placards of the past. Visitors are now accustomed to a multimedia environment that mixes entertainment with an educational message.

The original museum audio systems included narrations and sound effects. Loudspeakers would be mounted in front of the exhibit, often at knee level. The source was often continuous loop tapes which could not be rewound. If they ran continuously, the listener could come into the dialogue at any point and have to wait before he or she could hear the beginning. If the tape was capable of stopping at the end of the narration, the listener could push the start button and hear the message from the beginning. Often multitrack playback tape machines were used so you could pick your language of choice. Although it was not the best system, it was much better than those boring placards.

Around 1957 the visitors carried a reel-to-reel tape recorder over their shoulders. With this system, they not only heard the message but also received their exercise for the day.

44.5.1 Inductive Loop Systems

Another early system, and still used by some museums today, is to transmit the signal on a wire inductive loop antenna that surrounds the audience area. The listeners wear a receiver and earpiece and as long as they are within the boundary created by the loop, they can hear. As they step outside of the loop, the signal disappears. They can then go to the next exhibit, step into its loop, and hear the dialogue. An advantage of the system is it is simple and reliable and works with hearing aids. The drawbacks are:

- Poor frequency response making it useful for voice.
- As the signal is analog and operates much like an AM radio station, the volume and sensitivity vary with the distance of the listener to the loop.

- Affected by external electrical noise such as lightning, electric motors, and SCR lamp dimming circuits.
- Requires a wire loop around the area of interest, sometimes rather difficult to install and hide.

For more information on magnetic induction loop systems, see Chapter 46, *Assistive Listening Systems*.

44.5.2 Infrared Systems

Another type of system by Sennheiser and others uses infrared (IR) transmission. In this system the message is transmitted via wireless infrared using amplitude and frequency modulation processes. They come in either narrow band for multichannel setups or wideband for high quality.

The area of reception is confined to line-of-site or an individual room. Through reflections, however, it can bounce around corners into other unwanted areas. While they can cover large areas effectively, they are limited when it comes to multiple exhibits in a confined area. Another problem with IR is its poor operation in the sun or very bright areas.

Dual channel systems normally operate subcarriers of 95/250kHz or 2.3/2.8MHz. The emitters are placed around the room to give even coverage and they may be daisy-chained for easy installation.

For more information on infrared systems, see Chapters 45, *InterpretationSystems*, and 46, *Assistive Listening Systems*.

44.5.3 RF Systems

Today radio frequency (RF) systems are mostly used. This has made the systems much more versatile and easier to install. Many RF systems are available, some simple and others complex. The following are a few of those systems. This is only an indication of some of their capabilities; to find more about each system and the latest updates, go to the company's Web site.

44.5.3.1 RF Systems with Stationary Transmitters

Message repeaters with multiple inputs are ideal for museum displays that require several messages associated with a single display. For example, messages that are tailored to either adults or children, short messages or detailed messages, or multiple language messages.

The following systems generally use a stationary message repeater and transmitter and multiple worn or handheld receivers.

Acoustiguide has been in business for over 50 years. Its major system is the Acoustiguide 2000 Series and includes three AG 2000 players—the *Wand*, the *Mini*, and the *Maxim*.

All three systems use MP3 and Windows Media Audio 4.0 for full production sound, and their own software called Vocoder for voice only.

The *Wand* is 12.5in × 2.5in × 0.75in (31.8cm × 6.4cm × 1.9cm) and weighs 9.2oz (261g), Fig. 44-4. It can hold up to 500 selectable languages or programs or up to 8000 messages. The controls include Play, Clear, Pause, Fast Forward, Rewind, Volume Up, and Volume Down. It can play for 12h continuous without charging and can accommodate surveys, games, and educational question-and-answer formats. Battery charging can be accomplished in three hours in the charging/programming rack.

Figure 44-7. Typical wands used in museum systems. Courtesy Acoustiguide Inc.

Because of the design of the *Wand*, it is easy to encourage corporate sponsorship including rotating logos on the LCD screen and the flat areas on the casing are good for applying logos and graphics.

The *Mini* has many of the same features as the *Wand*. The *Mini* is ideal for highly produced audio programs that blend narration, archival audio, large interviews, music, and sound effects making exhibits come to life. The *Mini* comes with headsets or single earpieces and is 5.6in × 2.6in × 0.75in (14.2cm × 6.6cm × 1.9cm) and weighs 5oz (148g). The controls are the same as with the *Wand*. It will also play for 12h without charging and can be fully recharged in 3h.

The *Maxim* can hold 200h of stereo sound or over 2000h of voice in either linear, random access, or combination tours. It can hold 500 different programs and over 12,000 messages on each unit to provide tours on different subjects or foreign languages. The unit is

7in × 3.9in × 1.5in (17.8cm × 9.9cm × 3.8cm) and weighs 15oz (444g). It has the same controls as the *Wand* and the *Mini*.

The Acoustiguide storage racks recharge the batteries and include a programming card that is about the size of a credit card. The programs can be either written by the client or by Acoustiguide which can provide creative and production services. The programs are downloaded from the Internet or from CDs onto a laptop computer. As new material is written, recorded, and digitized, it is put on the program card, which automatically updates the players as they are being charged.

To operate the system the visitor is given a player. A staff member sets up the player for the language and the complexity of the tour. The tour could be long or abbreviated to control traffic when the museum is crowded or can be set up for adults or children. The visitor can adjust volume at each area to compensate for noise level. When the visitor is at an exhibit, he or she punches in the number corresponding to the exhibit as shown on a placard. The visitor can then pause the program, rewind it, or fast forward it.

Acoustiguide's newest unit is a compact screen-based player, developed and designed specifically for on-site interpretation of museums and visitor venues. The Opus series allows institutions to provide visitors access to various digital resources—video, images, and animation, plus the traditional audio, Fig. 44-5.

The high-performance computing capabilities of Opus include sophisticated graphic images and digital movies as its processing speed and memory capacity enable delivery of high-resolution video files and CD-quality sound.

The administrative user interface allows for simple additions and deletions of content, as well as more complex functions, such as integration of timed audio with video and images.

Opus comes in two formats: Opus Click™ and Opus Touch™. The click format utilizes a keypad while the touch format utilizes a touchscreen.

Figure 44-8. Acoustiguide Opus Touch™ screen-based player. Courtesy Acoustiguide Inc.

The systems incorporate the following:

- Remote triggering and synchronization.
- Remote content activation via IR and RF technologies.
- Synchronization to external multimedia or show control systems.
- Data collection and visitor surveys.
- Software tracks user click stream.
- Customized surveys can be incorporated into the guide's audio/visual content.
- Produces easy-to-read reports.
- Visitors can bookmark items of interest either for on-demand printing via MyCollection™ or other postvisit services such as e-mailing information home.

- Map-driven or object-driven modes.
- Full color TFT LCD screen.
- Large, expandable memory.
- Compatible with a complete range of audio, video, image and animation multimedia formats.
- Dual-listening mode, via internal loudspeaker and/or through headset/earpiece.
- Up to 12h of usage between charges.
- Remote activation via IR and RF.
- Opus Content Management System for easy setup and install, and can be used by client.
- Dual-listening mode.
- A fold-out loudspeaker, each player can be used either as a true wand or as a headset unit.
- Headphone integrated into strap.
- MP3 stereo sound quality.
- Range of 16 stepped volume levels.
- 500h of multilingual audio content.
- MP4 and JPEG visual quality.
- QVGA resolution (320 × 240 pixels) and 65,536 color depth.
- 28h of video, or 10,000 images.
- 2GB memory; expandable.
- Holds multiple languages and tours.
- Guides can contain any combination of audio, images, animation, or video clips.
- Graphically rich user interface with menu selection and navigation functions.

The Tour-Mate SC500 Listening Wand is 13in × 1.8in × 1in (33cm × 4.6cm × 2.5cm) and weighs 8oz (227g). A carrying strap is attached inside the wand for added strength and so it that it cannot be unclipped by the user. The system is powered by a rechargeable nickel metal hydride battery pack which will deliver 10h of continuous play from a full charge and can be recharged in 3–4h, Fig. 44-6.

Figure 44-9. Typical charger/programmer.

The maximum capacity of the system is 24h of mono sound or 12h of stereo sound. The message can be expanded on-site. The wand can store several tours and/or versions of the tour. The software permits a staff member to type in a code that locks out all tours but the desired one. A keystroke permits one to see what version of the tour has been selected.

The Tour-Mate editing capability software is windows compatible. The editing software permits the user to input tour messages or message segments and to perform such functions as: cut, paste, parametric equalization, normalization, variable gain, variable compression, insert message queues, and program message sequences.

MyGuide by Espro is a system much like the previous two. It uses a wand that has the tour narration downloaded from the storage-rack/power supply recorded through a flash memory card. This system runs for 10h between charging and can have up to 4h of audio capacity. The bandwidth is between 300Hz and 4kHz so it is particularly useful for voice only.

ExSite MP3 system by Espro can have up to 72h of multilingual content, uses a wide alphanumeric and graphic LCD screen, and can be synchronized to external multimedia presentations such as DVD and video. It can also collect and analyze visitor usage data. The system can be used with the unit's built-in speaker or with plug-in earphones.

GroupGuide by Espro is a portable system for group tours where the visitor wears a personal receiver with headphones, and the guide wears a transmitter with microphone.

GuidePORT™ by Sennheiser is a completely different type of system. To operate GuidePORT, the museum is set up into zones or cells, Fig. 44-10. These zones may be separate rooms or floors or a section of a large room. Audio files associated with the exhibit in a cell and their corresponding identifier unit are created and/or stored on a standard PC. The files are uploaded by GuidePORT software to multichannel RF (radio frequency) wireless cell transmitters located in each individual cell. Each cell transmitter stores the audio for its particular zone. The audio (prerecorded and/or live stream) for that particular cell is downloaded into the visitors' receivers when they enter the cell.

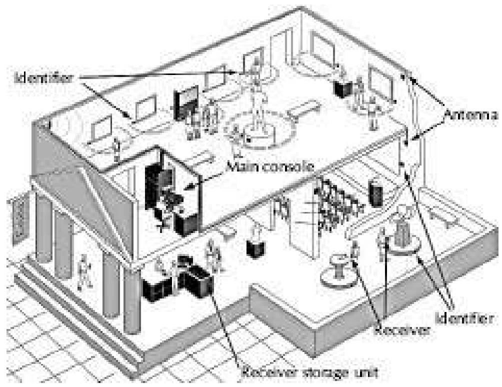


Figure 44-10. GuidePORT system layout for a typical museum. Courtesy Sennheiser Electronic Corporation.

GuidePORT's charger system can store and charge ten wireless receivers. Chargers can be linked to accommodate 5,000+ receivers. Receivers can operate for up to 8h between charges. The charger system is linked to the control unit (PC) to allow programming the receivers for language and/or level.

To enable management of a frequently changing exhibit environment, Sennheiser has engineered a list-based audio configuration software so the museum management can control the audio tours by simply updating the master audio list as exhibit items and corresponding identifiers are moved.

Discreet wireless antennas are strategically placed throughout the exhibit to allow receivers and cell transmitters to interoperate. The system is designed to operate on license-free dedicated radio frequencies in the 2.4GHz ISM band that are ideal for digital audio and resistant to outside radio interference.

Battery-operated or externally powered wireless identifier units, Fig. 44-11, are hidden near or behind each exhibit. The wireless architecture behind the GuidePORT system allows for quick and easy setup of the museum because, as an exhibit is moved, the associated identifier is moved along with it making it easy to rearrange an exhibit space.

Figure 44-11. GuidePORT identifier. Courtesy Sennheiser Electronic Corporation.

The visitors are given a lightweight receiver that fits in the palm of their hand or can be hung around their neck, and a headset. The receiver is programmed by a staff member to the language and level the visitor desires. The system is hands free so the visitor is not required to press buttons to match exhibit placards. The visitors proceed into the exhibit at their own pace and as they move from exhibit to exhibit, the system automatically dissolves the audio from the previous message to the new one. Visitors can adjust the volume and pause or repeat information they would like to hear again, Fig. 44-12. The headphones fit all age groups and can include a sponsor logo.

When the visitor enters a zone, audio files for all of the exhibits within that zone are downloaded into the visitor's receiver. The identifiers automatically trigger the receiver when a visitor is within a specified range of the displayed item to play the corresponding audio file. The trigger range, along with other parameters of the identifier, can be programmed via an infrared enabled Palm™ compatible PDA.

Figure 44-12. GuidePORT receiver. Courtesy Sennheiser Electronic Corporation.

GuidePORT can integrate live audio into the presentation. The visitor can listen to live demonstrations, concerts, movies, and video presentations with synchronized sound just by walking into the area. The visitor can leave the area and walk to a new area and the audio program will automatically change.

All stationary components of GuidePORT are located in a central location. Cell transmitters, Fig. 44-13, interface with their base station PC via USB ports. A larger facility can network multiple base station PCs, including through an existing network. Antennas are connected using standard shielded Cat5 cable. Audio files may be created anywhere in any standard format, which are then converted to .WAV files before they are imported into the GuidePORT system. The base station PC and/or central control unit is only needed when configuring or reconfiguring the system, and could be substituted with a temporary PC or notebook.

Today RF FM systems are most often used, making the systems much more versatile and simpler to install. These systems range from simple to quite complex. The following systems are only a smattering of what is available but give an indication of the features available.

44.5.3.2 Portable Tour Group Systems

A portable tour group FM system consists of a portable FM transmitter and portable FM receivers. A microphone, often worn over the head, is connected to the transmitter and broadcasts the presenter's voice to everyone in the audience or group. The portable transmitter allows the audio to be delivered without having to carry a microphone or be plugged into the wall. Participants wear a portable FM receiver with an earphone to hear the presentation. An unlimited number of receivers can be used with one transmitter as long as the participants are within the broadcast range, typically up to 150ft (45.7m).

The transmitter and the receivers are tuned to the same channel and depending on the frequency, three–eight channels can be used simultaneously. This allows for multiple tours to be conducted at a time and/or provide language interpretation. Currently tour group products are available in 72MHz, 150MHz, 216MHz, 863MHz, and 926MHz.

Transmitters and receivers offer a mix of features and functionality for channel selection, programming, power, signal strength, and use. Portable transmitters and receivers are typically battery powered with standard alkaline or NiMH batteries.

Tour group systems are excellent for factory tours, museums, outdoor events, wireless microphone applications, classroom or training, or personal use. Anywhere you need to amplify sound but don't have (or want) an installed sound system.

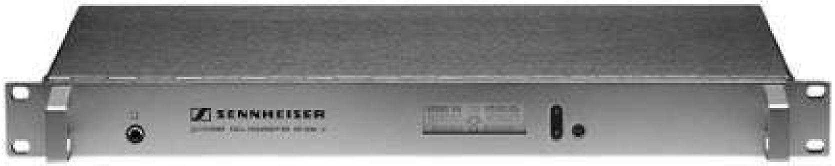


Figure 44-13. GuidePORT SR 3200-2 Twin Cell Transmitter. Courtesy Sennheiser Electronic Corporation.

44.5.3.3 FM Transmitter

A typical portable FM transmitter is shown in Fig. 44-14.

Figure 44-14. Typical high-quality FM transmitter, the Listen LT-700-072. Courtesy Listen Technologies Corporation.

The specifications for this transmitter are:

General

- Number of channels: 17 wide-band, 40 narrow-band.
- Channel tuning capable of being locked.
- *SNR*: 80dB or greater.
- Output power: adjustable to quarter, half, or full.
- Audio frequency response: 50Hz-15kHz ±3dB.
- Includes: a microphone sensitivity switch.

- Includes: a mute switch.
- Operates on two AA batteries.
- Includes: an LCD display that indicates battery level, channel, channel lock, low battery, battery charging, programming, and RF signal strength.
- Includes: automatic battery charging circuitry for recharging NiMH batteries.

Radio Frequency

- RF frequency range: 72.025–75.950MHz.
- Frequency accuracy: ±0.005% stability from 32 to 122°F (0–50°C).
- Transmitter stability: 50PPM
- Transmitter range: 0 to 150ft (45.7m).
- Output power: less than 100mW (72MHz).
- Antenna: the microphone cable.
- Antenna connector: 3.5mm connector.
- Compliance: FCC Part 15, Industry Canada.

Audio

- System frequency response: 50Hz–10kHz ±3dB 72MHz.
- *SNR*: SQ enabled; 80dB; SQ disabled 60dB.
- System distortion: <2% harmonic distortion (THD) at 80% deviation.
- Microphone input: unbalanced, +4dBu maximum, −10dB nominal input level adjustable, impedance 10kΩ).
- Microphone sensitivity: three position switch: high, middle, and low in 6dB increments.
- Line input: unbalanced, −10dBu nominal input level, −3dBu

maximum, impedance 10kΩ
- Microphone power: 3Vdc bias.

Controls

- User controls: power, mute, channel up/down.
- Setup controls: in the battery compartment, microphone sensitivity, NiMH/alkaline battery, SQ enable/disable.
- Programming: channel lockout, channel lock.

Indicators

- LED red: illuminated when unit is on, flashes when batteries are low, or to indicate charging. Flashes two times when muted.
- Display: channel designation, lock status, signal strength indication, battery life, RF power.

Power

- Battery type: 2 AA batteries, Alkaline or NiMH.
- Battery life: alkaline –10h, NiMH rechargeable.
- Battery charging: (NiMH only), fully automatic, 13 h.
- Power supply compliance: RoHS, WEEE, UL, PSE, CE, CUL, TUV, CB compliant.

Physical

- Dimensions: 5.0in × 3.0in × 1.0in (13.0cm × 7.6cm × 2.5cm).
- Color: dark gray with white silk screening.
- Unit weight: 3.9oz (111g).
- Unit weight with batteries: 5.8oz (164g).

Environmental

- Temperature – operation: 14° to 104 °F (–10° to 40°C).
- Temperature – storage: –4° to 122°F (–20 to 50°C).
- Humidity: 0 to 95% relative humidity, noncondensing.

44.5.3.4 FM Receiver

A typical FM receiver is shown in Fig. 44-15.

Figure 44-15. Typical high-quality FM receiver, the Listen LR-500-863. Courtesy Listen Technologies Corporation.

The specifications for this FM receiver are:

General

- Number of channels: 17 wide band, 40 narrow band.
- *SNR*: 80dB or greater.
- Programmable to electronically lock out unneeded channels.
- Can seek channels and can be locked on a single channel.
- Adjustable squelch.
- Audio frequency response: 50Hz–15KHz ±3dB.

- Includes a stereo headset jack for either a mono or stereo headset.
- Includes an LCD display that indicates channel, battery level, low battery, battery charging, and RF signal strength.
- Functions in both DX and Local mode.
- Operates on two AA batteries.
- Includes an automatic battery charging circuitry for recharging of NiMH batteries.

Radio Frequency

- RF frequency range: 72.025–75.950MHz.
- Number of channels: 17 wide band, 40 narrow band.
- Sensitivity: 0.6µV typical, 1µV maximum for 12dB sinad.
- Frequency accuracy: ±0.005 stability, 32 to 122°F (0–50°C).
- Antenna uses the earphone cable.
- Squelch is programmable in twenty steps, automatic on loss of RF signal.
- Compliance: FCC Part 90, Industry Canada.

Audio

- Frequency response: 50Hz–15kHz ±3dB 72MHz.
- *SNR* (A-weighted): SQ enabled: 80dB; SQ disabled 60dB.
- System distortion: <2% total harmonic distortion (*THD*) at 80% deviation.
- Output: unbalanced, 0dBu nominal output level, 16mW maximum, impedance 32Ω

Controls

- User controls: channel up/down, seek, volume.
- Set up controls (battery compartment): alkaline/NiMH batteries,

SQ enable/disable.
- Programming: channel lock, squelch, channel lockout.

Indicators

- Red LED: illuminated when unit is on. Flashes when batteries are low, or to indicate charging. Flashes when locked and seek is pushed.
- Display: channel designation, lock status, signal strength indication, programming.

Power

- Battery type: two AA batteries, alkaline or NiMH.
- Battery life: alkaline –15h NiMH rechargeable.
- Battery charging (NiMH only): fully automatic, 14h.
- Power supply: Input: 120Vac, Output: 7.5Vdc 250mA.

Physical

- Dimensions: 13.0cm × 7.6cm × 2.5cm (5.0in × 3.0in × 1.0in).
- Color: dark gray with white silk screening.
- Unit weight: 3.9oz (111g).
- Unit weight with batteries: 5.8oz (164g).

The above specifications are fairly common for high-quality receivers.

CD display shows channel, lock, and battery level.

44.6 Narrow Beam Loudspeaker System for Exhibits and Tour Groups

The directivity (narrowness) of any wave producing source depends on the size of the source compared to the wavelengths it generates.* Audible sound has wavelengths ranging from a few inches to several feet, and because these wavelengths are comparable to the size of most loudspeakers, low- to medium-frequency sound (20Hz to 10kHz) generally propagates omnidirectionally. Only by creating a sound source much larger than the wavelengths it's producing can a narrow beam be created. To accomplish this with standard loudspeakers would require loudspeakers 50ft (15m) in diameter. A narrow beam of sound from a small acoustic source is accomplished by generating a beam of ultrasound, which becomes audible as it travels.

Ultrasound, whose wavelengths are only a few millimeters long, are much smaller than the source, and consequently travel in an extremely narrow beam.

Ultrasound contains frequencies far outside of our range of hearing, and is completely inaudible, but as the ultrasonic beam travels through the air, the inherent properties of the air cause the ultrasound to distort (change shape) in a predictable way. This distortion gives rise to frequency components in the audible bandwidth, which can be accurately predicted, and therefore precisely controlled. By generating the correct ultrasonic signal, we can create, within the air itself, essentially any sound desired.

Note that the source of sound is not the physical device you see, but the invisible beam of ultrasound, which can be many meters long. This new sound source, while invisible, is very large compared to the audio wavelengths it's generating, so the resulting audio is extremely directional, just like a beam of light.

Often incorrectly attributed to so-called Tartini tones, the technique of using high-frequency waves to generate low-frequency signals was in fact pioneered by physicists and mathematicians developing techniques for underwater sonar over 40 years ago.

Dr. F. Joseph Pompei, then a researcher at MIT, solved the problems of using ultrasound as an audible source that plagued earlier researchers. His design of the Audio Spotlight®* sound system has become the very first, and still the only, directional loudspeaker system which generates low-distortion, high-quality sound in a reliable, professional package, Fig. 44-16. Fig. 44-17 shows the sound field distribution with equal-loudness contours for a standard 1kHz tone. The center area is loudest at 100% amplitude, while the sound level just outside the illustrated beam area is less than 10%.

Figure 44-16. Audio Spotlight AS-16B system. Courtesy Holosonic Research Labs, Inc.

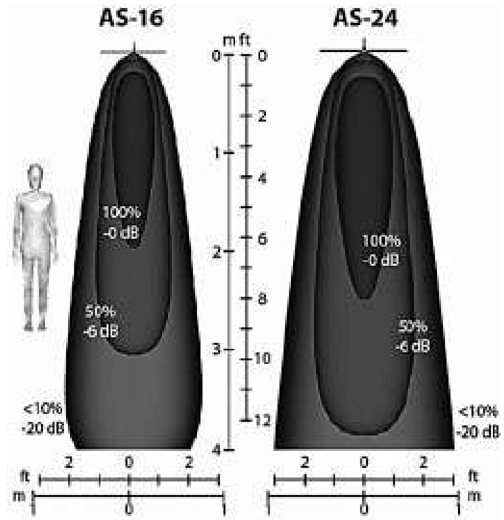


Figure 44-17. Sound field distribution of the Audio Spotlight AS-16 and AS-24 systems. Courtesy Holosonic Research Labs, Inc.

Audio Spotlight systems are much less sensitive to listener distance than traditional loudspeakers, but maximum performance is attained at roughly 1–2m (3–6ft) from the loudspeaker.

Typical levels are 80dB SPL at 1kHz for the AS-16, and 85dB SPL for the AS-24 models. The larger AS-24 can output about twice the power and has twice the low-frequency range of the AS-16.

The most common use of the Audio Spotlight system is to deliver sound to a specific, isolated area. Just as with lighting, the Audio Spotlight system is best mounted directly above the listener, aimed downward which provides maximum localization, Fig. 44-18. The speaker panel can also be mounted on a wall, and angled downward, to reach the listener.

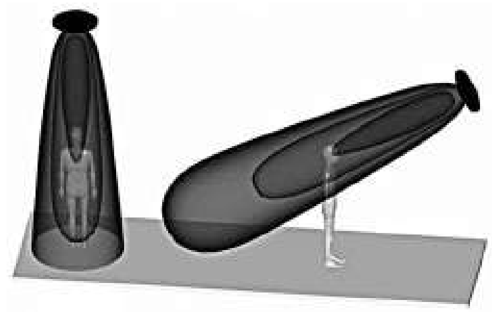


Figure 44-18. Mounting angles for Audio Spotlight systems. Courtesy Holosonic Research Labs, Inc.

Multiple Audio Spotlight systems can be used to create a larger field of sound, or to increase the sound intensity in a given region. Just like visual spotlights, beams of sound can be aimed next to each other to shape the sound field, Fig. 44-19A, or multiple loudspeaker panels can be aimed to one position to increase the output level substantially, Fig. 44-19B.

While the beam generated by the Audio Spotlight system is very narrow, the beam will reflect from surfaces (and listeners) in your environment. To sound waves, solid surfaces are much like mirrors are to light. Therefore, to reduce reflections, an acoustically absorbing surface (such as carpet, padding, or curtains) should be used to catch the beam and reduce the reflection. Generally, this is most important only in very quiet spaces, where there is little background noise to mask minor scattered energy. Also, like light, reflections can be used as projection of audible sound. By directing the beam against a surface, one can create very interesting virtual loudspeaker effects. The beam will generally maintain its directivity after projection, so it is best to insure that the listener is in the path of the reflected beam.

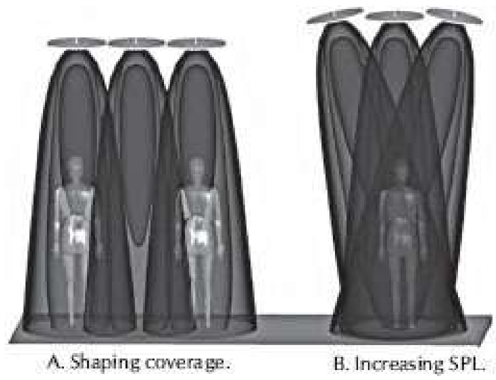


Figure 44-19. Multiple Audio Spotlight systems. Courtesy of Holosonic Research Labs, Inc.

The loudest sound area is directly in front of the speaker panel at a distance of 1–2m (3–6ft). Reasonable listening areas are within the darker zones. Sound levels outside the beam are down by over 90%. In all sound systems, audibility is determined by sound level received versus background noise levels. Therefore, the beam will be perceived as more narrow in the presence of background noise, as any scatter from a listener or floor will be inaudible. This is much like the difference in shining a flashlight in a completely dark room versus one with background lighting.

44.7 Voice Evacuation/Mass Notification Systems (MNS)

by Hardy Martin, Innovative Electronic Designs, LLC

44.7.1 UL2572 Standard for Safety/Mass Notification Systems

Information in this chapter is either directly or indirectly derived from NFPA 72, Edition 2013.

The UL2572 Standard for Safety specifies a configuration of components and interfaces that are used to communicate information to the occupants in a building, area site, or other spaces about emergency conditions. These systems may consist of equipment that can reproduce live and/or prerecorded voice messages, tones, and visual indicators such as strobe lights and visual displays.

Knowledge of issues such as risk tolerance for survivability, intelligibility, Acoustically Distinguishable Space (ADS), sound pressure levels and supervision, fault detection, and fault reporting are all needed to have an effective MNS.

Using an MNS on a daily basis is important to ensure that personnel will hear normal live and/or prerecorded announcements and become accustomed to receiving these non-emergency announcements.

When the MNS is integrated with an in-building fire emergency voice/alarm communications system, the NFPA 72 requirements will take precedence.

The MNS system in control shall be capable of establishing and maintaining the following priorities.

Live Emergency Announcements. Will override all other announcements.

Prerecorded Emergency/Emergency Text-to-Speech (TTS) Announcements. Will override all other announcements except Live Emergency Announcements.

System Mute. Mutes entire system including any background music (BGM) or other programs. Will be overridden by a Live Emergency Announcement or Prerecorded Emergency Announcement. System will stay muted for all lower priority announcements including BGM or other programs.

Non-Emergency Announcements. A risk analysis should be conducted to determine what priorities are required.

The time period for processing and activation of signals in worst case loaded systems shall not be greater than 10 seconds from the initiation of an alarm condition, or operation of a manually activated switch.

Visual display devices, electronic advertising signs, etc. are considered supplemental MNS devices. These devices can be used to display emergency messages in place of normal display information, Figs. 44-20 and 44-21.

Visual display devices can further provide directional information to the occupants for proper safe locations, Fig. 44-22.

Synchronized visual paging allows further support to all prerecorded and Text to Speech (TTS) announcements, both Emergency and Non-Emergency. The text on the displays must be synchronized with the announcements, Fig. 44-23.

Secondary Power Operations. The secondary power supply for in-building Mass Notification Systems (MNS) shall be capable of operating the system under quiescent load for a minimum of 24 hours and then shall be capable of operating the system during emergency load for a period of 15 minutes. This applies to both engine driven generators and/or uninterruptible power supplies (UPS).

Figure 44-20. Fire warning message. Courtesy Innovative Electronic Designs, LLC.

Figure 44-21. Tornado warning message. Courtesy Innovative Electronic Designs, LLC.

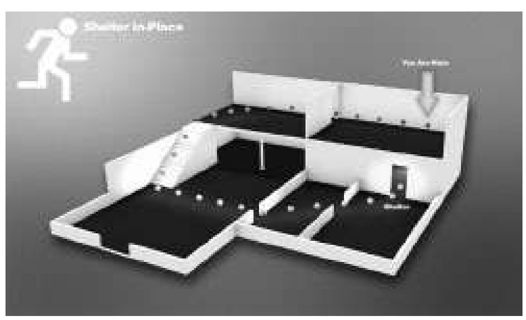


Figure 44-22. Emergency wayfinding message. Courtesy Innovative Electronic Designs, LLC.

Authority Having Jurisdiction (AHJ). AHJ is an organization, office, or individual responsible for enforcing the requirements of a code or standard, or for approving materials, an installation, or a procedure.

Figure 44-23. Scrolling text message. Courtesy Innovative Electronic Designs, LLC.

Supervision. An initial system test should be performed with logging of all system components, periodic test as per AHJ, fault detection, and fault reporting.

Emergency Communications Systems (ECS). Whether it consist of stand-alone in-building EVACS, MNS, or some other form of integrated ECS, it provides a critical emergency capability that can save lives and property. Designers must keep in mind that for any design of an ECS, fundamental sound and communication concepts, intelligibility, security, survivability, and the proper use of messaging all represent important aspects of a complete design.

44.7.2 NFPA 72 2013 Edition, Chapter 24 Emergency Communication Systems (ECS)

Risk Analysis. Risk Analysis is a process used to characterize the likelihood and magnitude of and vulnerability to, incidents associated with natural, technological, and man-made disasters and other emergencies. Risk analysis should be integrated into the

Emergency Response Plan (ERP). Risk analysis encompasses three interrelated elements.

Risk Assessment. Defines what might occur and establishes the probability of occurrence.

Risk Perception. Represents the psychological and emotional factors that affect behavior. How we perceive a risk will determine how we react to it. By providing clear, concise directions and information, an MNS can help to calm the fears of individuals based on their perceptions of the incident.

Risk Management. Provides strategies for reducing future losses based on risk assessment and the factors that influence risk perception.

Pathway survivability levels for MNS shall be determined by the risk analysis. Each application of a MNS shall be specific to the nature and anticipated risk of the facility. This is a brief description of potential content with the mass notification event. The actual questions must be tailored to the area or areas of the facility.

What is the type of emergency. Is it a fire, meteorological, environmental, safety, or another type of event?

What is the urgency of the emergency event. Does it represent immediate danger, has it already occurred, is it expected to occur soon or is it expected to occur in the future?

What is the anticipated or expected severity of the emergency event. How will it impact the facility and its functions, is it expected to be moderate, extreme, severe, etc.?

What is the certainty of the emergency event. Is the event happening now, is it likely or very likely to occur, will it occur in the future, or is the occurrence unknown?

What zone or areas should receive the emergency announcements and messages. Is it an area or areas of a building, the entire building, is it more than one building if so what buildings, or is it the entire facility?

What is the validity of the emergency event. Has the emergency event been investigated and/or confirmed?

What is the recovery plan. What actions will be taken to restore operations to the pre-emergency condition?

It is important to remember that when an emergency event occurs, the response must be immediate and deliberate. There is no time for indecision. The risk analysis shall be used as the basis for development of the ECS provisions of the facilities Emergency Response Plan (ECR).

44.7.3 NFPA 72 2013 Edition 3.3.135 Intelligibility

The quality or condition of being intelligible. The term used in intelligibility relates to voice communications used in Emergency Communication Systems (ECS). If voice messages to occupants in buildings and other locations cannot be understood, the message system will have little, if any benefit. Requirements for voice messages to be intelligible are not new to the code, however, with the expansion of requirements for Emergency Communication Systems (ECS), intelligible voice communications have become more important. Requirements for voice intelligibility are referenced in Chapter 24, Annex D – Speech Intelligibility which provides guidance on system design with emphasis on testing.

Acoustically Distinguishable Space (ADS). ADS can be an emergency communication zone, or subdivision thereof, that can be an enclosed or otherwise physically defined space, or that can be distinguished from other spaces because of different acoustical, environmental, or have different characteristics such as reverberation time and/or ambient sound levels. Each ADS might require different components and design features to achieve intelligible voice communications. For example, two ADSs with similar acoustical treatments and noise levels might have different ceiling heights. The ADS with the lower ceiling height might require more ceiling mounted loudspeakers to insure that all listeners are in a direct sound field, Fig. 44-24. Other ADS might benefit from the use of alternate loudspeaker technologies such as line arrays to achieve intelligibility. In areas of 85dBA or greater ambient sound pressure levels, meeting the pass/fail criteria for intelligibility might not be possible and other means of communications might be necessary.

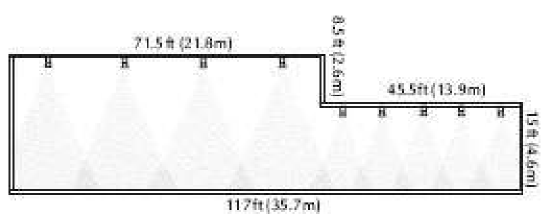


Figure 44-24. The effect of ceiling height on loudspeaker coverage. Source R. P. Schifiliti Associates, Inc.

44.7.3.1 Speech Transmission Index (STI)

STI measured in public address systems can be very time consuming as a complete set of 98 measurements of *modulation transfer functions* (MTF) has to be obtained and summed. The basic principle of STI measurements consists of synthesized test signals instead of a human speaker's voice. The speech intelligibility measurement acquires this signal and evaluates it, as it would be perceived by the listener's ear.

Speech Transmission Index Public Address (STIPA). STIPA simplifies the full STI measurements by calculating the SNR by using only two modulation frequencies in each of the seven octave bands for a total of 14 MTF. Now portable STIPA analyzers are able to evaluate speech intelligibility within 15 seconds per room position. These findings are incorporated into the speech intelligibility meter that is able to display the intelligibility results as a single number between 0 (unintelligible) and 1 (excellent intelligibility), Fig. 44-25.

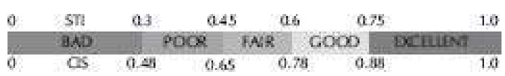


Figure 44-25. STIPA criteria for speech intelligibility. Source NTi Audio.

44.7.3.2 STIPA Test Signal

The STIPA test signal is a special audio signal that is played over the Emergency Communication System (ECS) being tested. It consists of:

Talkbox. An instrument usually consisting of a high quality audio loudspeaker used to play an STIPA test signal. The input signal of the talkbox should be configured to produce the proper level for testing a handheld microphone, a gooseneck microphone, a telephone handset, or other devices that would receive signals from a person's voice. The talkbox may also have the STIPA Test Signal loaded into the talkbox. Set the volume of the test signal to match that of speech level under normal conditions. The microphone feeding the ECS should be from 1 to 2 inches from the talkbox loudspeaker. When using the NTi Audio TalkBox select Track 1 for the STIPA test signal.

Testing the Emergency Communication System (ECS), with the STIPA signal, from the microphone input through the system to the output of the power amplifier, should produce a STI intelligibility result of 0.9 or better. See Fig. 44-26 for details of the signal path. The purpose of this test is to set a baseline of the ECS from the input of the microphone to the output of the amplifier. After this baseline is established you can record the STIPA test signal into the system or feed the STIPA test signal directly into the system for all future tests.

Measuring and matching speech levels. Set the analyzer (meter) to measure sound pressure level, A-weighted, fast. At a typical ADS location in the facility, position the analyzer so its microphone is approximately 5ft (1.5m) above the finished floor. The test should be performed during a period of time when the ADS area is unoccupied and the ambient sound pressure level is low. Activate the prerecorded voice message from the ECS.

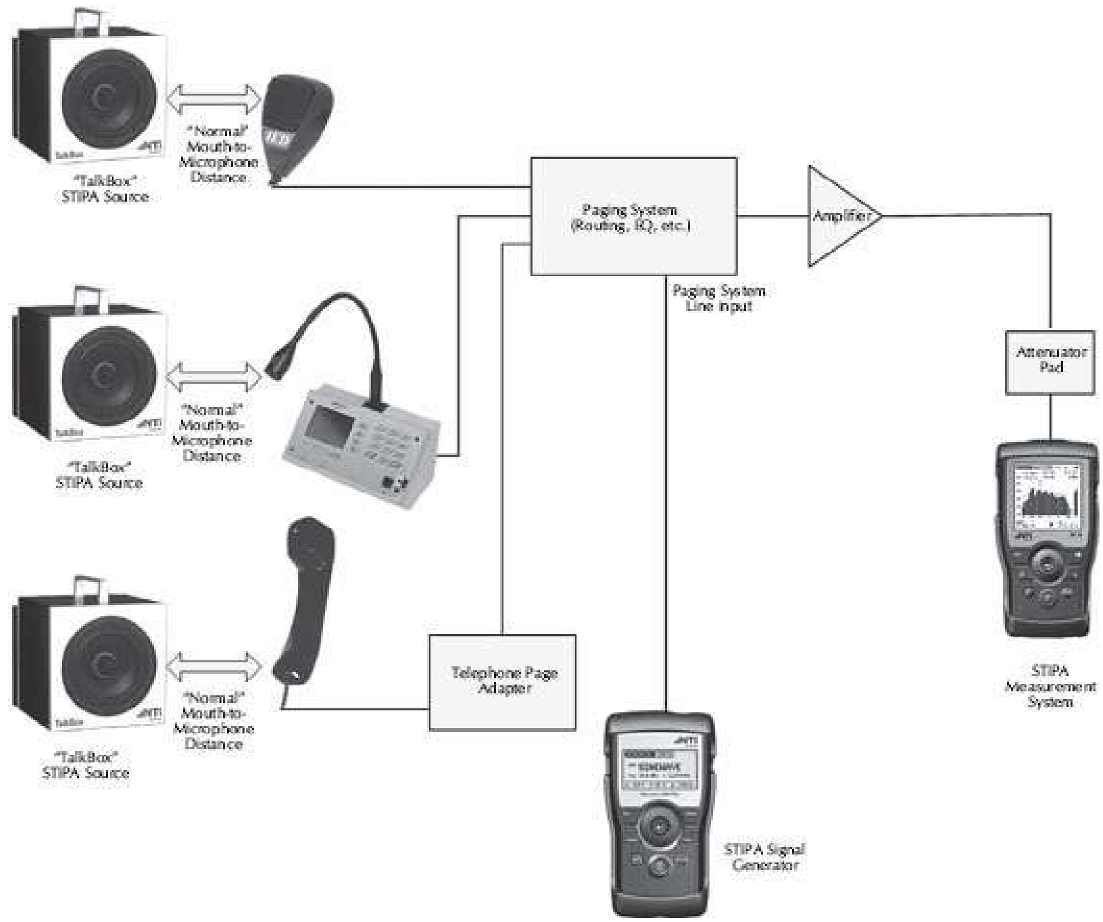


Figure 44-26. Measuring STIPA with NTi Audio instruments. Courtesy Innovative Electronic Designs, LLC.

The sound pressure level produced by the STIPA test signal should be matched with the sound pressure level of the prerecorded voice messages.

1. Play a prerecorded message.
2. Measure the sound pressure level of the prerecorded message. Do not move the analyzer from the test location.
3. Turn off the prerecorded message and turn on the STIPA test signal.
4. Adjust the sound pressure level of the STIPA test signal to match the sound pressure level of the prerecorded message. Use

the adjusted level of the STIPA test signal for all other tests.

44.7.3.3 Testing all the Acoustically Distinguishable Space (ADS)

All parts of a building or area intended to have occupant notification are subdivided into ADSs. In smaller areas, such as those under 400ft^2, walls alone will define the ADS. Physical characteristics such as a change in ceiling height of more than 20% or change in acoustical finish, such as carpet in an area and tile in another would require those areas to be treated as separate ADSs. In areas of 85dBA or greater ambient sound pressure level, meeting the pass/fail criteria for intelligibility might not be possible and other means of communications might be necessary.

There are many factors that affect the intelligibility of the announcements and prerecorded messages played through the ECS, into the ADS areas. Major intelligibility factors include:

1. Background noise.
2. The configuration of the space being addressed.
3. The acoustical properties of the materials on the walls, floors, and ceilings.
4. The distortion and bandwidth of the sound system equipment.
5. Poor loudspeaker directivity or coverage.

44.7.3.4 Procedure for Measuring the Intelligibility in Each of the ADS Areas

First measure, record and store with an analyzer such as the NTi XL2, a realistic ambient noise level in the ADS area during the time of day when the area is heavily occupied. Then using the ambient noise stored in the analyzer, measure the speech intelligibility using the STIPA Test Signal during a time of day when the ADS area is unoccupied. The corrected STI is arrived by post-processing of the occupied ambient sound pressure level measurements, the unoccupied sound pressure level measurement, and the unoccupied STI measurement. In effect, the measured STI (uncorrected) is being corrected by adding in the effects of the actual expected (occupied) ambient sound pressure level.

If an ADS is small enough to only require one measurement location, the result should be 0.50 STI (0.70 CIS) or greater for the ADS to pass the requirement for speech intelligibility. Some ADS might require multiple measurement points due to their large size. Where an ADS has multiple measurement locations, the requirement is that at least 90% of the measurement locations have values not less than 0.45 STI, (0.65 CIS) and that all measurement points average to 0.50 STI, (0.70 CIS) or greater.

The test results should be fully documented and provided to the building owner, the Emergency Communication System (ECS) contractor, the system designer, the Authority Having Jurisdiction (AHJ), and any other individual or organization deemed appropriate.

44.7.4 Acronyms

ADA – Americans with Disabilities Act.

ADS – Acoustically Distinguishable Space.

AHJ – Authority Having Jurisdiction.

CIS – Common Intelligibility Scale.

dBA – Decibel, A Weighted.

ECS – Emergency Communications Systems.

ERP – Emergency Response Plan.

EVACS – Emergency Voice/Alarm Communication System.
MTF – Modulation Transfer Functions.
MSN – Mass Notification Systems.
NFPA – National Fire Protection Association.
STI – Speech Transmission Index.
STIPA – Speech Transmission Index Public Address.
TTS – Text to Speech.
UPS – Uninterruptible Power Supply.

44.8 Audio Archival

In addition to message repeating applications, there is a need for long term storage—i.e., to preserve audio and video data for future generations. It has been only the last quarter century that we came out of the dark ages and into the domain of digital storage.

As the world is changing from analog to the digital domain, the media for archiving must be improved. After all, the medium used 20,000 years ago for written information can still be read today, while today's media for audio storage lasts only a few years and playback equipment becomes obsolete.

Today, most memory for audio storage is accomplished through solid state technologies. Solid state technologies fall into four broad categories:

1. Electrical memories based on semiconductor IC technology.
2. Magnetic memories based on magnetic materials.
3. Optical memories based on the interaction of light with matter.
4. Molecular, chemical, or biological memory based on changes in the atomic, molecular, or biological level.

Electrical memory is the most used technology for digital audio storage. Information is stored in digital form in various types of memory units. Common circuits are DRAMs, PROMs, EPROMs, flash EEPROMs, and ROMs.

Dynamic random access memory devices (DRAMs) store information dynamically, that is, as a charge on a capacitor. These designs feature one field-effect transistor (FET) to assess information for both reading and writing and a thin-film capacitor for information storage. Most nonvolatile cells rely on trapped charge stored on the floating gate of the FET. These units can be rewritten many times, the limit being determined by programming stress-induced degradation of the dielectric. Erasure of the charge from the floating grid is accomplished by tunneling or by exposure to ultraviolet light.

DRAMs are volatile, the average memory is about 10 years. Programmable memories can be programmed at least once and some can be programmed a million times. A few nonvolatile memories are programmable just once. These have an array of diodes or transistors with fuses or antifuses in series with each semiconductor cross point.

Electrically programmable read only memory devices (EPROMs) are usually used to describe cells that are electronically written and UV erased. EEPROM is probably the most common technology used. Static random access memory devices (SRAMs) are sometimes connected to EEPROM for storage when power is removed. Flash EPROMs require bulk erasure and therefore cannot be written over by the consumer.

Read only memory (ROM) is the only form of semiconductor storage that is permanently nonvolatile. Even with no power source present, information is retained in a ROM without any information loss.

Optical storage devices, the CD and DVD, are popular for long term archiving for the following reasons:

- Disk medium is highly standardized.
- Disk medium is multimedia (sound, data, still images, moving images).
- Disk medium format has a commercial life expectancy of many decades.
- Disk medium is an efficient and evolving medium.
- Disk medium has good chemical and mechanical resistance.
- Disk medium has good resistance to harsh environmental conditions.
- Disk medium has contactless reading—i.e., nondestructive.
- Disk medium is cost effective.
- Disk medium is an unrecordable system in the ROM version that prevents erasing or overwriting.

Archived CDs must be chemically stable, have good resistance against scratching, breaking, etc., and must be tolerant to extreme conditions of temperature, humidity, and electromagnetic fields. Some companies, such as DIGIPRESS, produce a stable CD. Rather than using a polycarbonate substrate, the CENTURY-DISC ARK from DIGIPRESS has a desalcanized etched tempered glass substrat, which is covered with titanium nitride—a very resistant metal. They can reach a lifetime of over 200 years—not forever maybe, but a great deal better than our present mediums can.

* Much of this section was copied with permission from copyrighted text by Holosonic

Research Labs, Inc.

* Audio Spotlight is a registered trademark of Holosonic Research Labs, Inc.

Chapter 45

Interpretation Systems

by Glen Ballou

45.1 Interpretation Systems

As the world gets smaller and smaller, communications become increasingly important. Countries must talk to countries, businesses to businesses, and people to people. Only a few years ago, simultaneous interpretation systems were only found in places such as the United Nations and NATO. Today businesses are doing business with partners around the world, religious organizations have international meetings, schools are multilingual, and video and audio conferencing is common place.

Designing and building a simultaneous interpretation system is not just adding a set of earphones and another microphone to a sound system. A simultaneous interpretation system requires sound equipment and an acoustically correct room for the interpreter. The output from the various interpreters is transmitted to the various listeners in their language. This can be done via hardwire and earphones, AM or FM transmission, induction loop, or with infrared transmission systems.

Simultaneous interpretation systems allow a presentation by a talker to be heard and understood in or close to real time by all people in the audience. To accomplish this, the voice of the talker is directed to interpreters in soundproof booths or areas. The interpreters hear the original or floor language on headphones and instantly or simultaneously interpret it into the language they are assigned. The translated signal is then transmitted back into the audience area through the interpreters, microphones and transmission medium to the listeners through their control panel and headsets.

There are two basic types of simultaneous interpretation systems: bilingual and multilingual. Bilingual systems are designed for places where two and only two languages are used, such as in eastern Canada where French and English are used. Bilingual systems are the least expensive and the simplest to set up and use. These systems usually use only one interpreter's booth with either one or two interpreters.

Multilingual systems are used in the United Nations, large church conferences, boardrooms and schools, just to name a few. These systems are much more complicated and harder to install and use as they require individual interpreter rooms and a means for the listener to switch between languages.

45.1.1 Central Control Unit

The central control unit is the hub of the system. Most systems are microprocessor controlled and/or operated through an IBM compatible PC, Fig. 45-1. The floor language enters the unit at line level and is routed to the interpreters' booths and a tape recorder if required. The interpreted languages are returned to the central control from the interpreters' booths where they are prepared for transmitting to the listeners. This could be on hardwire, induction loop, infrared, or any combination of the three. Provision is also made for taping the interpreted language. The unit incorporates various operating modes and interlocks and a means for the interpreter and the operator to communicate with each other.

45.1.2 Interpreter's Booth

In the multilingual system, each booth normally has two or more interpreters that work as a team to interpret the floor language into the designated language of the booth. If many floor languages are allowed, each booth could require as many as four interpreters. The ISO standard for fixed interpreters' booths in systems with six or more languages recommends three interpreters per booth. Systems can have from two to 64 languages; at such high numbers of different languages, the relay and auto relay functions become very important, Fig. 45-2. Today most systems are digital, which can reduce background noise, distortion, and crosstalk. AGC assures equal listening level on all input channels, and the systems can be chained together with shielded FTP or STP Cat-5e cables, Fig. 45-3.

Booth size is specified by international standards. Permanent interpreters' booths and equipment are specified under ISO 2603 (1983), which specifies the minimum dimensions of 2.5m wide × 2.3m high × 2.4m deep (8.2ft × 7.75ft × 7.87ft). In booths with three interpreters, the width shall be 3.3m (10.5ft). An 80cm (31.5in) high window should extend the full width of the booth with the bottom of the window flush with the console. The room construction should attenuate the live sound so that if the nonreinforced sound does not exceed 80dB, the inside signal will not exceed 35dB.

Portable interpreters' booths are specified by ISO 4043 1998, and sound transmission using infrared is specified by IEC 764. The HCS-851A/02 interpreter booth by TAIDEN Industrial Co is a portable solution with four window panels, three acoustically treated wall panels, the door panel and two roof panels with ceiling mounted ventilation fans; ISO 4043 compliant, Fig. 45-4. A permanent interpreters' booth is shown in Fig. 45-5.

Simultaneous interpretation systems are not just a simple input to an interpreter booth. The system must enable the interpreter to hear the Floor content (whether that be a delegate speaking, program material from a presentation or content from the far end of a teleconference) and to distribute the interpreted language back to the language distribution system where it is sent to the various listeners. Beyond this the system must accommodate the following operations:

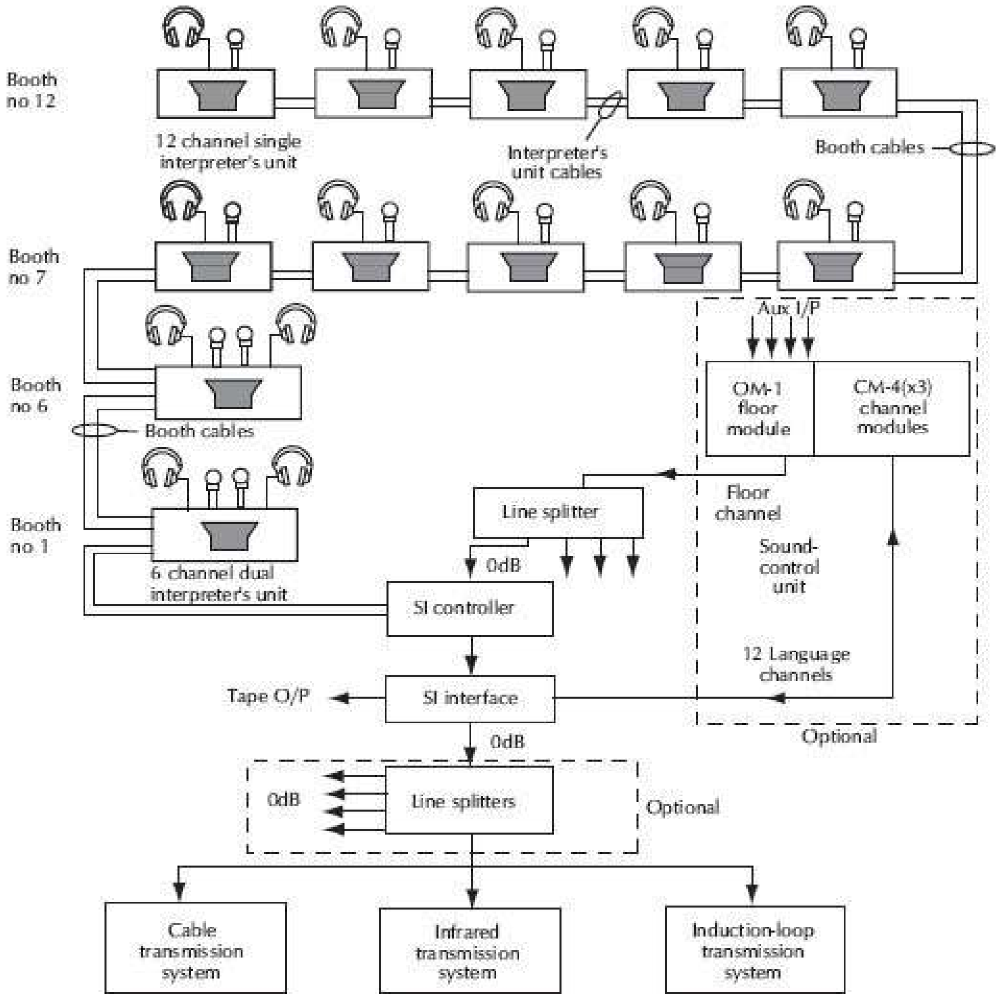


Figure 45-1. Interpretation system. Courtesy Auditel Systems Limited.

Figure 45-2. Twelve language interpreter's unit. Courtesy Auditel Systems Limited.

- Should Interpretation be unnecessary, the floor channel shall pass through the booth back to the Language Distribution system where participants and observers can hear the original speaker in their native tongue.

- If there are many languages at the meeting, the interpreter may have to rely on the interpretation of another booth to do their own interpretation. This is called relay interpretation and essentially allows for the Interpreter to select their incoming channel. If the language requirements for interpreters are known, these relay switches can be made to happen automatically; this is called Auto Relay Interpretation and the selection of the B or C channel is used as the trigger for the switch in assigned booths.

Figure 45-3. A fully digital two outgoing A/B channels and five input relay buttons. Courtesy Listen Technologies Corporation.

Figure 45-4. A portable or fixed installation interpreter's booth. Courtesy TAIDEN Industrial Co.

> Example: The Floor is speaking Chinese; the Chinese booth will switch to the B (or C) output to interpret into English; the other booths select the French channel instead of the Floor channel to begin their interpretation.

- It is advisable to have some degree of flexibility and control over the output channel of the booth to make fullest use of the capabilities of the team of interpreters in the booth. There must also be a means to prevent inadvertent interference with an engaged channel by another booth. A versatile system will allow for the selection of Interlock, Override or B/C Override for operation between booths as different organizations will handle this differently.

Figure 45-5. A fixed installation interpreter's booth showing three interpreter stations. Courtesy Listen Technologies Corporation.

- When necessary, the output of any or all channels must be available for recording, and/or connecting to other feeds for transmission elsewhere.
- Occasionally, the interpreter will need assistance and the equipment within the booth must make this communication easy for the interpreter. Common methods are the "Help" button which sends a text message to the Control Booth engineer or a "Call Booth" button which opens a one way audio call to the Control Booth.
- Should the interpreter need a second chance at hearing a complex number or formula, the ability to "Repeat" the Floor channel becomes a very important tool within the Interpretation System.

Fig. 45-6 is a modern interpreter console compliant to the ISO 2603. The unit is compact but with ergonomics for efficient use of various controls including Microphone On/Off, Mute, headphone volume, relay select channels as well as output channel selection. The unit allows for an interpreter to carry out the standard functions of their position and also to receive information from the Conference Operator via text message; to repeat the floor channel and get a second chance at hearing a complex statement; to contact both the Conference Officer directly or the Conference Engineer in the Control Booth.

Fig. 45-7 is an interpreter terminal for two interpreters according to ISO 4043. The unit is a double interpreter terminal for alternating operation and includes two microphone/headphone combinations. The two output channels are directly selectable by the interpreter and relay translations are possible from all languages. The listening area contains volume, bass, treble controls, an incoming channel selector, and an original/relay lever switch. It also has extra communication channels to and from the system operator and status information lights.

Figure 45-6. Interpreter console. Courtesy TAIDEN Industrial Co.

45.2 Language Distribution

Once the language has been interpreted and sent to the master station, it must be routed to the listeners. There are two basic systems of transmitting the signal; the hardwired system and the infrared transmission system.

45.2.1 Hard Wired Systems

Hardwired systems are primarily used to transmit interpreted language channels to delegate stations on the conference hall floor. They are most useful in areas such as the United Nations building where the listeners are always seated in the same place and can tolerate the cable to the earphones. A hardwired system is the most reliable, has the best security against eavesdropping and has the best audio performance. As a rule, hardwired systems are cheaper in hardware costs but more expensive in installation costs. In multiconductor cable systems, each channel is amplified and transmitted on a pair of conductors. Each listener usually has a panel located at his or her seat that includes a language selecting switch, a volume control, and an earphone jack. If the conductors are a twisted pair, there is little crosstalk in lines in excess of 1000m (3280ft), and farther with shielded cable. Hardwired systems are not particularly good for portable systems as it is not easy or physically safe to lay out cables on the floor to the various listeners. It is important that the user does not place the earphones next to a microphone, unless it is shutoff, as they may be on the same channel and cause feedback. Acoustical crosstalk can occur if open back earphones are used because an adjacent live microphone can sometimes pick up the interpreted language.

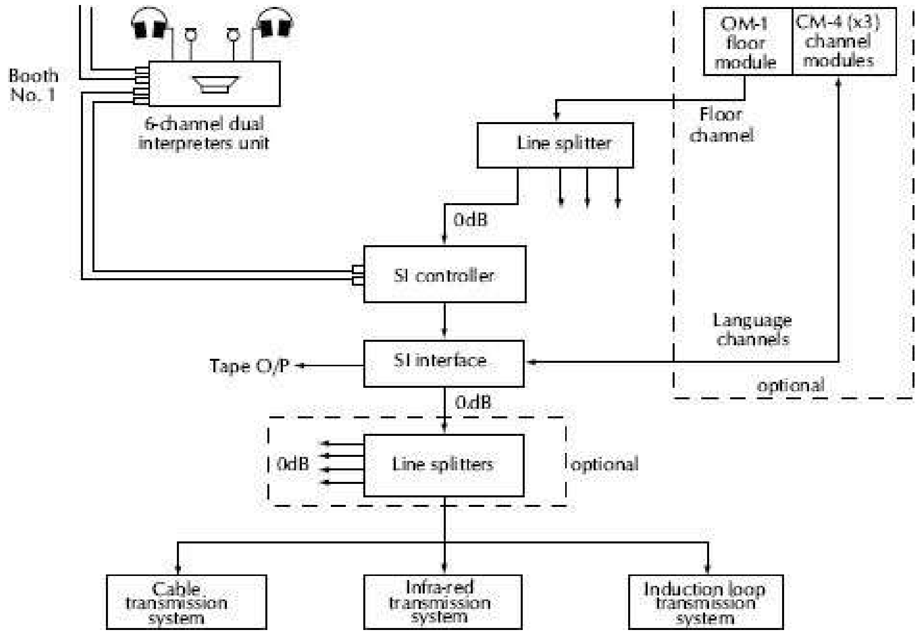


Figure 45-7. Two person interpreter's terminal.

If many languages are used, it may be better to multiplex the signal rather than use a multiconductor cable. This system would consist of a central modulator with up to twelve channels driving a network of active channel selector units using coaxial cable in a loop-through configuration. Power for the channel selectors is provided by power supplies injecting dc into the network.

Most single cable conference systems with delegate microphone units incorporate a built-in loudspeaker. The loudspeaker signal is derived directly from a common audio line. This simplifies cabling by avoiding the necessity for a second audio line to drive the loudspeakers. This does mean that the input and output signal are on the same line, and would create a closed loop and feedback unless some means of isolating the two signals is employed. This problem is overcome by Auditel with the application of a common mode reverse audio feed (CMRAF), Fig. 45-8. The technique is based on selective rejection of large common mode signals. The output from the microphone preamplifier is a balanced signal and is extracted in the central unit via transformer. After signal processing, the loudspeaker drive signal is injected into the audio pair in common mode form. Since the loudspeaker drive amplifiers and delegate units reject balanced signals the two signals can be carried over the same conductors without interaction or without compromising the signal quality.

45.2.2 FM Interpretation Systems

FM products can be used for language interpretation by connecting a stationary FM transmitter to an audio system transmitting an FM signal to a portable receiver for assistive listening and a language interpreter. In addition to the portable receiver, the interpreters use a portable transmitter and an over-the-head microphone and earphone unit. This combination allows them to hear the audio clearly in an adjoining area while speaking their translations in a normal tone of voice. Their translations are sent via FM back to participants' receivers. It is important to have transmitters and receivers with multiple channels allowing users to find clear channels even in a crowded venue with extensive FM use.

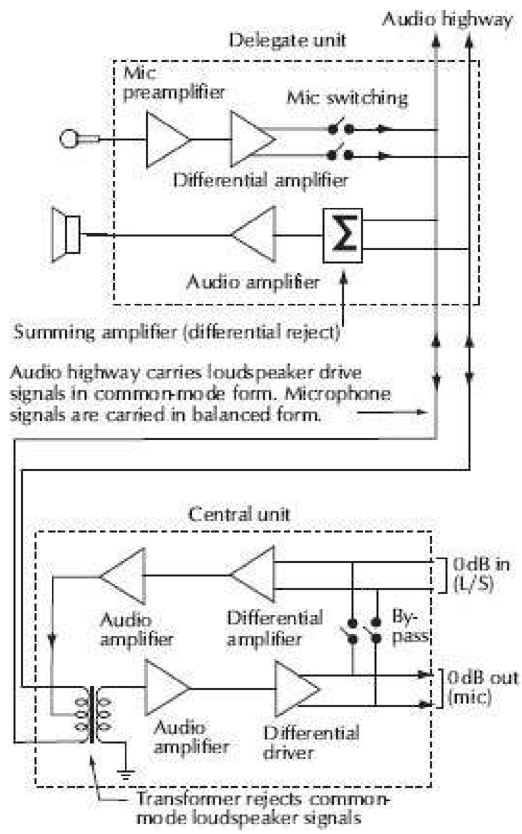


Figure 45-8. Common mode audio. Courtesy Auditel Systems Limited.

45.2.3 Digital Infrared Systems

Originally designed in the 1970s, infrared was initially used in home theater systems as a way to control various pieces of equipment and then later adapted to work with assistive listening equipment. Rather than using radio frequencies, the infrared medium by passes all of the interference issues that RF devices incur from crowded airspace. Since the 1970s, however, infrared technology has made a jump into what is casually referred to as Digital Infrared; the narrow band modulation techniques are still in the 875nm wavelength band, but the significant difference is that the frequency of IR light has moved up into the 2–8MHz range; conforming to the new standards set by the IEC 61603 and 61603-7. In this standard the system can transmit up to eight different carrier signals; each carrier can carry up to four audio channels. The exact number of channels per carrier depends on the selected quality mode. Stereo signals use twice as much bandwidth as mono signals; perfect quality (20Hz–20kHz frequency range) uses twice as much bandwidth as standard quality. A mix of channels with different quality modes can be chosen for each carrier.

Many Infrared systems audio signal is based on transmission by modulated infrared radiation (IR). Infrared radiation is part of the electro-magnetic spectrum, which is composed of visible light, radio waves and other types of radiation. The IR wavelength is larger than the wavelength of visible light.

Infrared is unable to pass through opaque objects such as walls so the signal cannot be overheard in other rooms. Moreover, infrared does not emit radio radiation so it does not require a radio frequency license worldwide. The walls of the building acts as a barrier to infrared signals escaping. Just like visible light, infrared radiation is reflected from hard surfaces and refracted by hyaloid (glassy or transparent appearance) objects. Both objects in the room and the structure of the walls and ceilings will influence the distribution of infrared light. Infrared radiation is reflected from most hard surfaces including smooth, bright or shiny surfaces. Dark or rough surfaces absorb a large part of the infrared energy. Normally surfaces opaque to visible light are also opaque to infrared radiation.

Shadows from walls and furniture will influence the transmission of infrared light which is solved by using a sufficient quantity of radiators positioned in a manner to provide an infrared field strong enough to cover the entire conference area. Care should be taken to not direct radiators towards uncovered windows as most of the radiation will be lost.

Most systems used today are based on modulated carrier techniques using FM or phase modulation. The operating frequencies for wide-band two-channel infrared systems are 95kHz and 250kHz with peak deviation of ±50kHz. Narrow-band systems operate on twelve or more channels between 2.3MHz and 2.8MHz with peak deviation of ±56kHz. These standards are specified in the IEC 76 international standard and insure compatibility between manufacturers. General specifications for infrared systems are given in Table 45-1. The system comprises three sections: the transmitter, the emitter (sometimes both are combined in one unit), and the receiver. The transmitter imparts the audio signal onto a subcarrier that the emitter converts into infrared light. The receiver decodes the infrared signal to retrieve the original audio, Fig. 45-9.

The analog audio is first converted into a digital signal and compressed. It then modulates the carrier, is filtered, magnified and sent to the radiator. The radiator transmits the IR signal to the receiver which magnifies, filters, applies agc, demodulates and turns it back into an analog signal to the earphones. Fig. 45-10 is a block diagram of the TAIDEN HCS-1500 Digital Infra-red Language Distribution System.

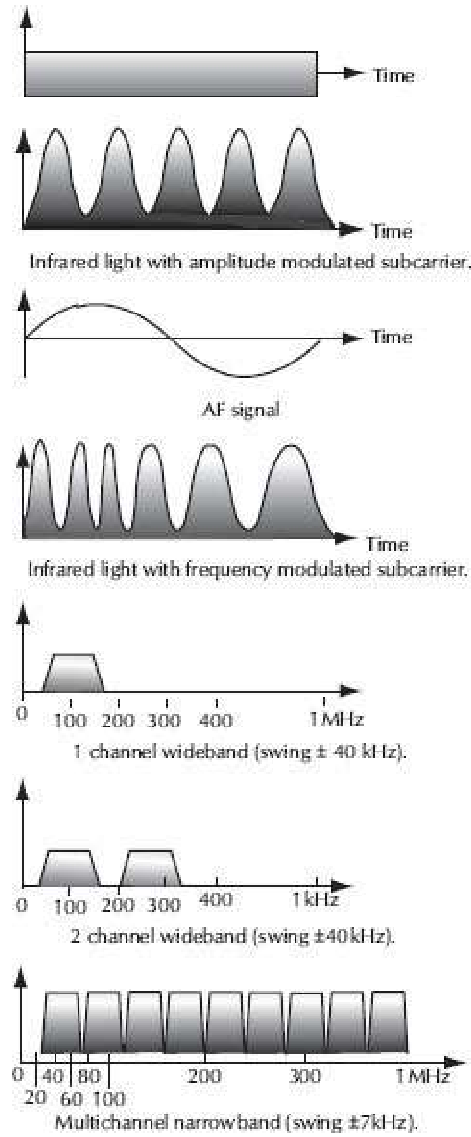


Figure 45-9. Infrared transmission characteristics.

The signal enters the listening area via infrared radiators. The IR light emitting diode incorporates a collecting lens which bends the infrared light to hit the diode, Fig. 45-11. Increasing the number of diodes increases the IR intensity and the coverage area multiplies by the number of diodes in the panel. The diodes have a continuous operating life of 100,000h before the light output diminishes to 70% of its original value.

Table 45-1. Technical Specifications for Infrared Systems

Characteristics	Narrow Band System	Wide Band System
No. of channels:	32	2
Carrier frequencies:	55kHz–1335kHz (excl. 455kHz)	2.3MHz and 2.8MHz
Channel spacing:	40kHz	155kHz
Modulation:	FM	FM
Pre-emphasis:	100µs	50µs
Normal deviation:	±6kHz	±35kHz
Peak deviation:	±7kHz	±56kHz
Transmitters:		
Frequency response (–3dB):	50Hz–8kHz	50Hz–13kHz
Max. distortion at 1kHz:	<1.0%	<1.0%
Signal/noise (A weighting):	>55dB	>70dB
Receivers:		
Frequency response (–3dB):	50Hz–8kHz	100Hz–9kHz
Max. distortion (at 1kHz):	<2.5%	<1.0%
SNR (A weighting):	>55dB	>63dB
Emitter Panels:		
Frequency response (–3dB):	30Hz–710kHz	

Sennheiser narrow-band multichannel operation channels are shown in Table 45-2:

Table 45-2. Sennheiser Narrow-Band Multichannel Operation Channels

Channel 0 55 kHz	Channel 16 735 kHz
Channel 1 95 kHz	Channel 17 775 kHz
Channel 2 135 kHz	Channel 18 815 kHz
Channel 3 175 kHz	Channel 19 855 kHz
Channel 4 215 kHz	Channel 20 895 kHz
Channel 5 255 kHz	Channel 21 935 kHz
Channel 6 295 kHz	Channel 22 975 kHz
Channel 7 335 kHz	Channel 23 1015 kHz
Channel 8 375 kHz	Channel 24 1055 kHz
Channel 9 415 kHz	Channel 25 1095 kHz
Channel 10 495 kHz	Channel 26 1135 kHz
Channel 11 535 kHz	Channel 27 1175 kHz
Channel 12 575 kHz	Channel 28 1215 kHz
Channel 13 615 kHz	Channel 29 1255 kHz
Channel 14 655 kHz	Channel 30 1295 kHz
Channel 15 695 kHz	Channel 31 1335 kHz

Depending on the size of the room, its shape, and surface characteristics, a single small radiator or multiple large radiators may be required. Auditel Systems Limited states that the number of radiators required for a *SNR* of >40dB can be calculated using the following equation:

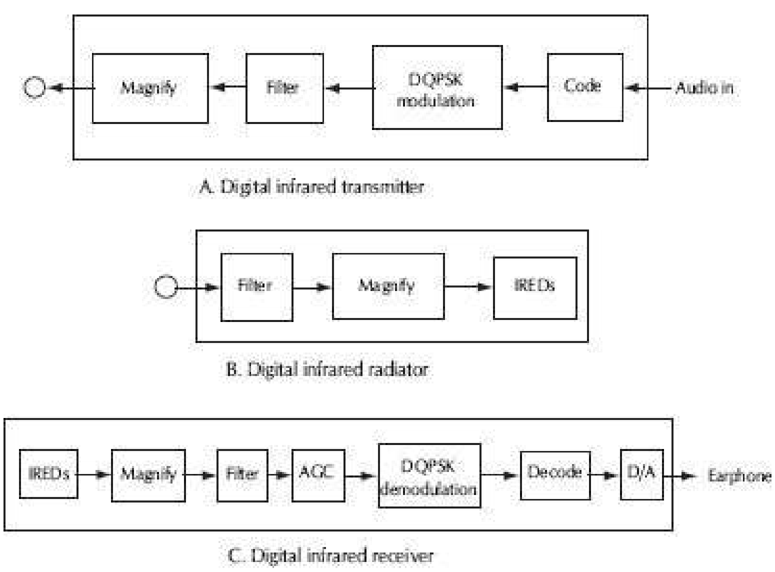


Figure 45-10. Block diagram of the TAIDEN HCS-1500 system.

$$N = \frac{area\ (m^2) \times number\ of\ channels}{D} \qquad (45\text{-}1)$$

where,

$$D = \frac{total\ emitted\ power\ (\mathrm{mW})}{receiver\ sensitivity\ (\mathrm{mW/m^2/channel})}.$$

This does not take into account the wall surfaces, niches, and obstructions and it assumes that at least 95% of the radiation is usable. Sennheiser states its large radiator can cover $11{,}000\mathrm{ft}^2$/number of channels in the system. TAIDEN Industrial Co states its large radiator can cover $1274\mathrm{m}^2$.

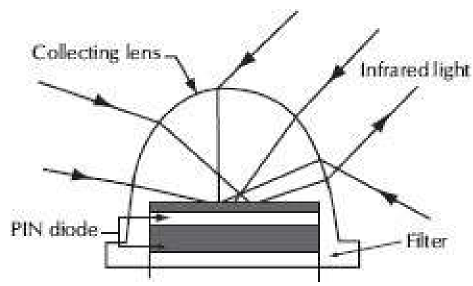


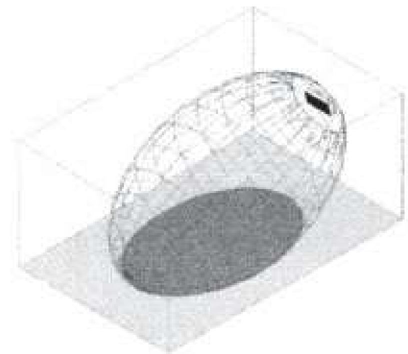
Figure 45-11. Light emitting diode (LED). Courtesy Sennheiser Electronics.

Layout of the panels is also important. Every seat must have a view of a panel. The range and coverage of a radiator are influenced by its orientation to the surface to be illuminated. A panel that is located so its pattern is parallel to the floor will have a long footprint with decreasing signal with distance. A panel that is aimed straight down will have a circular pattern that has about the same signal everywhere, Fig. 45-12.

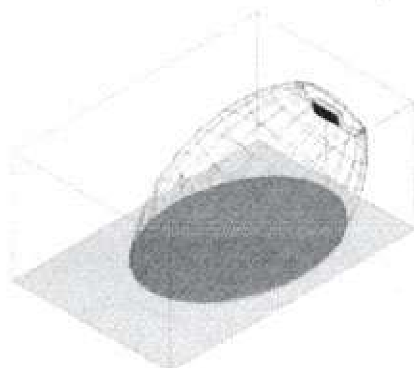
Total infrared power is proportional to the number of diodes in the panel. Power can be doubled by using two panels at the same location or using a panel with twice as many diodes. It is often best to use more than one radiator to eliminate dead spots in the area. If the overlapping areas have reduced signal, the two signals can add to bring the *SNR* ratio up to an acceptable level.

The infrared signal is received by the listener through a small receiver that can be worn around the neck on a lanyard or placed into an outside pocket. The receiver is about 155mm × 46mm × 24mm (6.1in × 1.8in × 0.9in) and weighs 135g (4.7oz) with the battery pack and incorporates a channel selector switch, volume control, and earphone jack., Figs. 45-13 and 45-14.

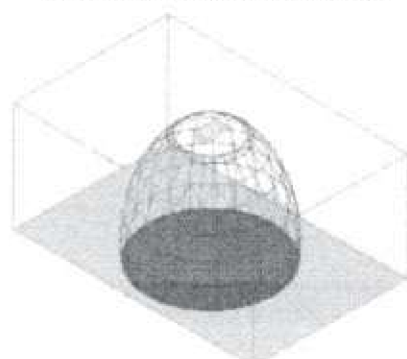
Security from eavesdropping is good since an unwanted receiver cannot see the transmitter or its reflections. Interference between rooms is also good since infrared cannot go through solid walls. In the past, infrared systems were affected by other infrared sources such as sunlight and incandescent and fluorescent lights. But with the advances in technology and the shift in frequencies used, Infrared is usable even in direct sunlight. Because it is line-of-sight, objects, including people between the transmitter and the receiver, can cause dropout unless at least two transmitters are covering the same area.



A. Radiator mounted 15° to the ceiling.



B. Radiator mounted 30° to the ceiling.



C. Radiator mounted 90° to the ceiling.

Figure 45-12. Coverage pattern at various angles of the radiator.

Figure 45-13. Infrared receiver with easy to read display screen. Courtesy TAIDEN Industrial Co.

Infrared systems are used for portable systems, where a large room can be subdivided, and for fixed installations. This system is more expensive than the induction loop but has much better audio quality, and is not as susceptible to electrical interference. Infrared is the system of choice today.

Figure 45-14. Infrared headset with built-in receiver. Courtesy

Sennheiser Electronics.

Chapter 46

Assistive Listening Systems

by Glen Ballou

46.1 Nature of the Problem

There are millions of people in the world (40 million in America alone) with hearing impairments for whom the acoustical and electronic systems described elsewhere in this book are inadequate. It is surprising that the special hearing needs of so large a group have been largely ignored for so long, especially when each of us faces the very real probability of joining that group through disease, trauma, or just by growing old.

According to the National Association of the Deaf, NAD, assistive listening systems (ALSs), sometimes called assistive listening devices (ALDs), are amplifiers that bring sound directly into the ear. They improve the speech-to-noise ratio by separating the sounds, particularly speech, that a person wants to hear from background noise.

Research indicates that people who are hard of hearing require a signal-to-noise ratio increase of about 15–25dB in order to achieve the same level of understanding as people with normal hearing. An ALS allows them to achieve this gain for themselves without making it too loud for everyone else.

ALSs are used by people with various degrees of hearing loss, from mild to profound, including hearing aid users and those with cochlear implants, as well as those who use neither. ALSs are sometimes described as “binoculars for the ears” because they stretch hearing aids and cochlear implants, thus extending their reach and increasing their effectiveness.

ALSs address listening challenges by minimizing background noise, reducing the effect of distance between the sound source and person with hearing loss, and overriding poor acoustics such as echo. ALSs are used in places of entertainment, employment, education, and home/personal use.

The hearing impaired are not just the people who wear aids. In fact, only about 20% of the hearing impaired wear aids. Many people with hearing losses are able to function in close or face-to-face situations but are lost in noisy or reverberant settings. Hearing aids work well with a single conversation and when the two people are only 4–6ft apart with little or no background noise. Even people who wear hearing aids have problems in reverberant rooms or where there is a high background noise level. Our standards for speech intelligibility are based on listening tests with normal-hearing subjects and are not directly applicable to the hearing impaired, see Chapter 3, *Psychoacoustics* and Chapter 40, *Designing for Speech Intelligibility*. Noise and reverberation degrade intelligibility far more rapidly for the hearing-impaired individuals whether they are fitted with hearing aids or not. Often the very highly prized acoustical qualities of theaters and concert halls operate against the needs of the hearing impaired, and the acoustical design of most classrooms and lecture halls are inadequate for the hearing-impaired student.

For years the only special assistance offered to the hearing impaired was headphones in a couple of pews in the front of the church sanctuary. In recent years new wireless technologies have been developed or adapted to meet the special needs of the hearing impaired in public assembly spaces. No longer is the user restricted to a specially wired seat. Depending on the technology used, now every seat is available; no special ticketing is required. The user is free to sit with family and friends.

Wide area ALSs are covered under Title III of the ADA (Americans with Disabilities Act of 1990). This title stipulates that ALSs be provided in public places unless a provider can prove that it is an undue burden. Examples of such venues include movie cinemas, live performance theaters, and public classes. The ADA specifies that ALS receivers be provided at no cost and specifies the number of receivers that must be provided depending on the number of seats (4% rule). Significant changes were made to the ADA in 2010 that affect the Standards for State and Local Government Facilities: Title II, and Standards for Public Accommodations and Commercial Facilities: Title III.

The adoption of the ADA 2010 changed “Capacity of Seating In Assembly Area” from the previous requirement of 50 seats or more to now read 50 seats or less.” They also changed the percentage of the Minimum Number of Receivers (which used to be 4%) to a sliding scale for larger venues. They also added a requirement of the “Minimum Number of Receivers that are Hearing Aid Compatible.” These changes were in agreement with the requirements of the (IBC) “International Building Code.” Please note: In the State of California, they have adopted their own (CBC) California Building Code which as the (AHJ) Authority Having Jurisdiction supersedes requirements of the ADA.

ALSs may also be indicated under ADA Title I (employment accommodations) as well as Title II (accommodations provided by state and local governments). Other public policies that may require use of ALSs include Section 504 of the Rehabilitation Act (affecting federally funded agencies) and Individuals with Disabilities Education Act.

46.2 Types of Assistive Listening Systems

There are three basic types of wireless systems: magnetic induction or also known as Hearing Loop, FM broadcast, and infrared (IR) light. Each type has its own set of advantages, problems, and limitations. There is no single best system for every application; each system is simple to operate and to install.

The system, no matter what type, must pick up the program sound. In a fully mic'ed event, this pickup could be a feed from the reinforcement control console. Where the event is not mic'ed, there must be a special microphone or microphones to feed the hearing-impaired system. It is very important that the feed to the system be of the highest quality possible with a minimum of reverberation pickup and extraneous noises. A pressure zone-type microphone, see Chapter 20, *Microphones*, on the forestage floor or mounted on an acoustical reflector panel over the fore-stage would be good for many shows. An even better system, which would reduce room effects, would be to individually or close- mic the actors, talkers, or singers. A sound pickup that does not reject the reverberant field and extraneous noise, and/or has distorted sound will not be useful for a system for the hearing impaired.

46.2.1 Magnetic Induction Loops

Magnetic induction, or sometimes called a *hearing loop system*, is one of the oldest but still useful systems. The principal advantage of this system is that it can operate directly into the user's cochlear implant or hearing aid that is T-Coil or Telecoil equipped without the need for a portable receiver as required by all other systems. A loop of wire or a specifically designed pattern of two loops—known as phase arrays—are wrapped around the seating area, usually under the carpet, and connected to an amplifier. The electrical current flowing through the loop will create a magnetic field (as the primary of a transformer) that can be picked up by a cochlear implant or hearing aid equipped with a *T-coil* (T as in telephone). About 60% and that number is rising of the hearing aids in the United States have T-coils for magnetic coupling with the earpiece of a telephone. Portable receivers are available for use by patrons who do not have an aid with a T-coil.

There are, however, several problems with the loop system. Most buildings have other magnetic fields that will be picked up by the induction loop system. Ordinary electrical wiring will radiate a large 60Hz field throughout the room, so the T-coil and the portable receivers are designed to have a rolled off bass response in the low-bass region in order to avoid the 60Hz hum. There are other power-line-related noises that cannot be filtered out—motors, dimmers, and fluorescent lamps being the most common. The size and shape of the loop and the amount of nearby steel in the building or in the seats will affect the strength and uniformity of the magnetic field. Simultaneous use of loops in adjoining rooms is often a problem because of crosstalk between the systems. However, this can be overcome by using a “low spill phased array” design.

The limited statistics available indicate that of the people needing hearing assistance, only about 20% are actually wearing aids, and only 60% of those aids are equipped with T-coils, which suggests that only 12% of those needing help are able to make use of a magnetic loop through their hearing aids. It has been argued that the majority of those without T-coils actually are young children and the very old; active adults are most likely to have T-coils. Despite the comparatively limited availability of T-coils and the several substantial limitations of the magnetic induction loop, or hearing loop system, it remains popular and enjoys the vocal support of many of those 2.5 million and growing number of people who have T-coils.

The cost of a magnetic induction loop system is largely the cost of the amplifier and the labor of installing the wire loop(s). The receivers are inexpensive. However, advances in solid state electronics have made the AM and FM broadcast systems very competitive in price. Where large areas are to be covered, the magnetic loop is probably not as cost effective as a broadcast system.

Loop Design Criteria

The international standard for the magnetic field strength of a loop system with an input signal of normal speech level is 0.1 A/m. Magnetic field strength H = 0.1A/m in SI units or 0.125Oe in cgs units. This field strength produces an audio voltage in the T-coil about equal to the output of the hearing aid's microphone at normal speech levels, Fig. 46-1. This eliminates the user having to make volume control adjustments when switching between microphone and T-coil. Also, this field is strong enough that noise and interference problems are minimized, yet it is not so strong as to overload the hearing aid amplifier. The proper received magnetic field is measured with a field strength meter at 400mA at a 1kHz sine wave. There are several measurements that are taken with a field strength meter.

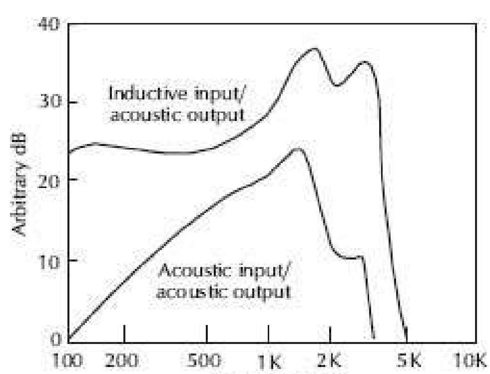


Figure 46-1. Typical hearing aid response. (From Reference 1.)

The field strength should be as uniform as possible over the coverage area. An achievable criterion for uniformity is a maximum variation of ±3dB in the audio output signal.

System design is based on the vertical component of the magnetic field, ignoring the horizontal field components for three reasons:

- The vertical field strength predominates over most of the loop area, Fig. 46-2.
- The T-coils in hearing aids are typically positioned to be most sensitive to the vertical field.
- Rotating the hearing aid about the vertical axis (as in turning the head) results in no change in the pickup of the vertical component, whereas the pickup of the horizontal component changes from zero to maximum to zero with such rotation.

46.2.2 Loop Location and Size

The field strength produced by the loop will vary in intensity from the edge of the loop to the center, Fig. 46-3. The range of variation is dependent on the area and shape of the loop and the listening height, which is the vertical distance between the plane of the loop and the receiver. This interrelationship is expressed as the relative listening height and is determined by

$$h_r = \frac{h}{\sqrt{0.5A}} \qquad (46\text{-}1)$$

where,

h_r is the relative listening height,

h is the listening height,

A is the area covered.

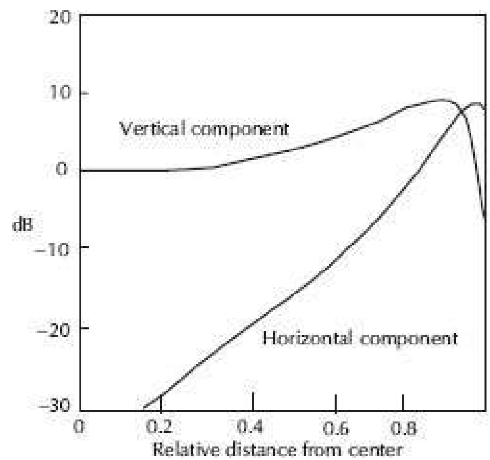


Figure 46-2. Field strength along the diagonal of a square loop. (From Reference 2.)

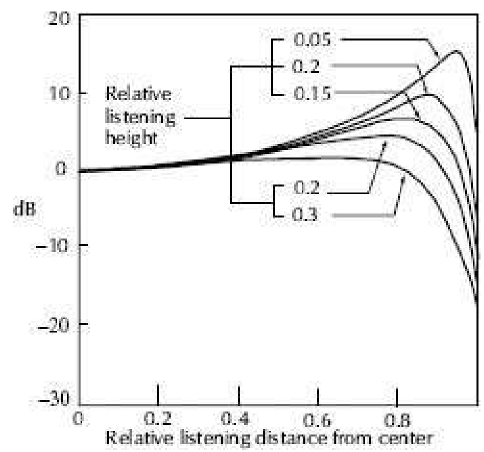


Figure 46-3. Vertical field strength along the diagonal of a square loop. (From Reference 2.)

The normalized field strength along the diagonal of loops of various shapes and the corresponding range of acceptable values for h_r are shown in Fig. 46-4.

By application of Eq. 46-1 and the h_r values found in Fig. 46-4, it is possible to design a loop of acceptable shape, area, and listening height. The penalty for an inequality in Eq. 46-1 is degraded uniformity of field strength, as can be seen in Fig. 46-4.

If the loop is to be placed at floor level (h = 48in for seated listeners), square loops falling within the acceptable h_r range will vary from 28ft × 28ft to 38ft × 38ft. A rectangular 1:4 loop may range in size from 24 ft × 96 ft up to 32 ft × 126 ft. Smaller loop dimensions will require a smaller h; larger areas need a larger h.

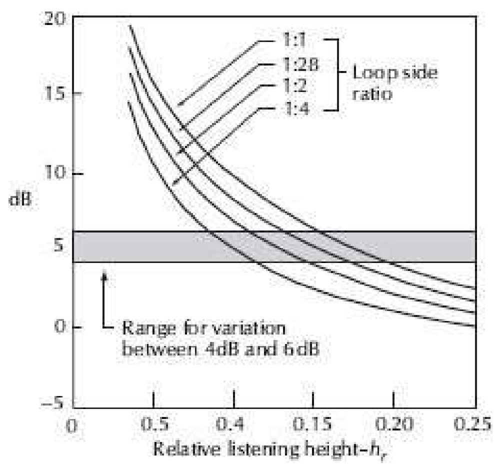


Figure 46-4. Vertical field strength along the diagonal of rectangular loops. (From Reference 2.)

As h grows larger, the field-distorting effect of steel in the building structure becomes more pronounced. This field distortion is manifest by dead spots within the loop. At its worst, the entire system may be rendered useless.

The great listening height required of large loops also presents architectural problems. Often the only practical place to locate a loop is at floor level, either below the floor or under the carpet. Where it is not feasible to locate the loop far above (or far below) floor level, the single, large loop can be broken up into a number of smaller loops that can be sized to locate at floor level. Because the vertical field strength rapidly falls to a minimum above the conductors, it is important to locate the loop wires in aisles or other areas that do not require coverage. For multiple loops, the current in parallel conductors of adjacent loops must flow in the same direction, Fig. 46-5.

Unfortunately, multiple loops will almost always have poorer uniformity than a single loop of the same size as one of the multiples. There is a special design technique for achieving a more nearly constant vertical field strength when using multiple loops or a phased array design. It involves the use of two sets of overlapping loops that are driven with electrical signals 90° out of phase. This complex procedure is described by Bosman and Joosten.[3]

46.2.2.1 Loop Current

Once the size and location of the loop are fixed, the required current in the loop can be calculated. The strength of the magnetic field is directly dependent on the current in the loop. The required current, *I*, in a single-turn loop is

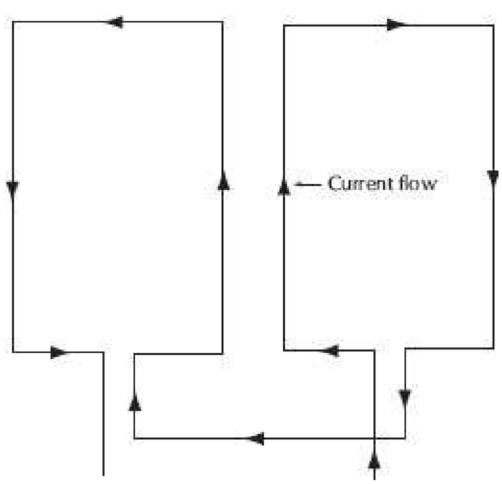


Figure 46-5. Multiple-loop current flow diagram.

$$I = \frac{0.1\,\text{A/m}^{*}\pi A}{2D} \times (1 + 2h_r^{\,2}) \times \sqrt{1 + h_r^{\,2}} \qquad (46\text{-}2)$$

* 0.03055 A/ft in English units

where,

0.1 A/m is the field strength criterion,

A is the loop area in m^2 or ft^2,

D is the loop diagonal in m or ft.

The terms containing h_r are a correction for the distance of the listener from the plane of the loop and are obtained from Fig. 46-6 by going vertically from *h* to the line and horizontally to the correction distance.

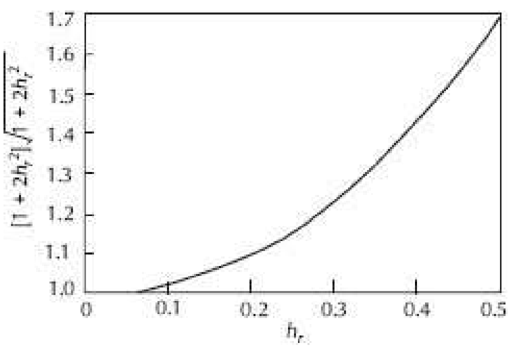


Figure 46-6. Graph for obtaining distance correction.

If a multiturn loop is used, the required current in the loop is

$$I_M = \frac{I}{n} \qquad (46\text{-}3)$$

where,

I is the current from Eq. 46-2,

n is the number of turns.

46.2.2.2 Loop Impedance

Wire size and number of turns in the loop must be selected to handle the required current safely and to control the range of variation of impedance across the audio band. A loop can be designed to provide the required magnetic field strength by using a relatively small wire with one turn or by using a larger wire with several turns. In the first case the loop impedance would be mainly resistive; in the second, it would be heavily inductive.

The impedance increases with frequency because of the inductive reactance of the loop. This increase is limited by adjusting wire size and the number of turns so the impedance at 1000Hz is no more than three times the impedance at 100Hz. This moderately rising impedance characteristic and falling loop current will complement the rising sensitivity characteristic of the T-coil, Fig. 46-7. Too high an impedance at high frequencies will result in too low current, producing poor response and degraded *SNR*.

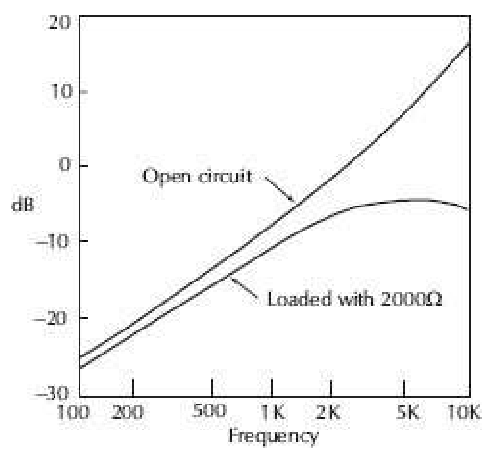


Figure 46-7. Sensitivity of typical inductive coil. (From Reference 1.)

A wire size that can handle the required current with an acceptable heat rise is selected from Table 46-1. The impedance is then calculated at several frequencies such as 100Hz, 1kHz, and 10kHz. The most common wire gauge used is a 14 or 16 AWG; however using flat copper tape is a very efficient medium and installs with virtually no visibility under carpet and other flooring materials. The following equations are useful:

$$L = \frac{rn^2}{13.5}\log\frac{2.8r}{d} \qquad (46\text{-}4)$$

where,

L is the inductance in H,

r is the radius of the loop in in,

n is the number of turns,

d is the diameter of the conductor in in,

(this is a simplification of Wheeler's equation).

$$Z = \sqrt{R^2 + (2\pi f L \times 10^{-6})^2} \qquad (46\text{-}5)$$

where,

Z is the loop impedance,

R is the dc resistance of the length of the coil,

f is the frequency of interest,

L is the inductance of the loop from Eq. 46-4.

An example of the calculations required to determine the impedance of a one turn 20 ft × 20 ft loop of 20 AWG wire at 1kHz is

$$L = \frac{(10\,\text{ft} \times 12\,\text{ft}) \times 1^2}{13.5} \log\left[\frac{2.8(10\,\text{ft} \times 12\,\text{in})}{0.03196}\right]$$

$$= 143\text{ H}$$

$$Z = \sqrt{\left(80 \times \frac{10.15}{1000}\right)^2 + (2\pi \times 10^3 \times 143 \times 10^{-6})^2}$$

$$= 1.21\,\Omega$$

If this is connected to a 1:4 autotransformer, the amplifier will see 4.84Ω.

Throughout this design procedure there is a certain amount of approximation involved; for instance, Eq. 46-4 applies to round loops. Error is introduced in calculating the inductance of a square or rectangular loop, however, this error is not great enough to seriously affect the results.

46.2.2.3 Electronic System

The power amplifier is selected that can supply the required current to the loop. Make sure you have selected an amplifier (also known as a hearing loop driver) that is designed to drive the low impedance and current required for these systems. Standard audio amplifiers that drive voltage to regular 4 or 8Ω loudspeakers should not be used. The power required is determined with the basic equation

$$P = I^2 Z \qquad (42\text{-}6)$$

The adjustment of the output current is determined by the equation

$$L = \frac{rn^2}{13.5}\log\frac{2.8r}{d} \qquad (46\text{-}4)$$

A typical loop system diagram is shown in Fig. 46-8. In very large halls, a delay unit may be required in the more distant loops in order to avoid excessive time delays between the loop signal and the acoustic signal. Equalization is desirable to compensate for any frequency response irregularities. The equalizer is adjusted to provide a natural sound quality with a typical receiver and to insure that power is not transmitted outside the power bandwidth of the receiver.

Table 46-1. Copper Wire Data

AWG #	Ohms/ 1000ft	Dia in In	Current for Heat Rise* 5°C	10°C	20°C	Melting Current
18	6.38	0.04030	5.5A	7.8A	10.0A	82.4A
16	4.02	0.05082	7.5A	10.0A	15.0A	117.0A

* Heat rise based on an insulation thickness of 10 mils.
Heavier insulation allows more current for the same heat rise.

A compressor is needed to insure that the system does not produce excessive distortion at high signal levels, either from clipping the amplifier or from overloading the hearing aid T-coil. The compressor should be adjusted according to the nature of the principal program material. Some manufacturers of hearing loop drivers have included either a compressor or AGC into their designs and external processing may not be necessary. If the system is used mostly for music, a compression ratio of about 4:1 will result in minimal harm to the music. If speech is the principal program, compression ratios up to 20:1 can be used to improve both intelligibility and *SNR*.

46.2.2.4 Installation

If the loop is installed in conduit, it must be nonmetallic conduit such as PVC and should be placed so that there is little (or no) steel between the loop and the listener. Often the conduit is run in the top of a concrete slab or below a wood-framed floor; but it can also be run in walls or even the ceiling of a room. When installing a loop in an existing room, it is often easiest to run the loop wire under the carpet, using conduit only for the run to the amplifier.

46.2.3 FM Broadcast

FM broadcast systems have replaced many magnetic loops in classrooms where hearing-impaired children are taught because the FM signal is normally free from noise and provides a more uniform and reliable signal. Several channels are available so systems can be used in adjacent rooms. The sound quality is excellent. The useful receiving range will vary from 30–914m (100–3000ft) depending on the amount of steel in the building. Transmitters are available for operation from the powerline for permanent installations or by battery for portable applications.

The Federal Communications Commission (FCC) has set aside a band of frequencies, 72.025–75.975MHz, for FM broadcasting to the hearing impaired under FCC Rules Part 15. These frequencies cannot be used for any other purpose, such as language translation systems or communications applications. No license is required, although the manufacturer of the transmitter is required to have FCC approval of the transmitter design. The FCC restricts radiation to a maximum field strength of 8000μV/m at 30m. The FCC rules require a special antenna connector on the hearing assistance transmitters to prevent the use of illegal gain antennas that could result in a higher transmitted field strength than dictated by the FCC. The system requires no special knowledge to install; sufficient instructions are provided by the manufacturers.

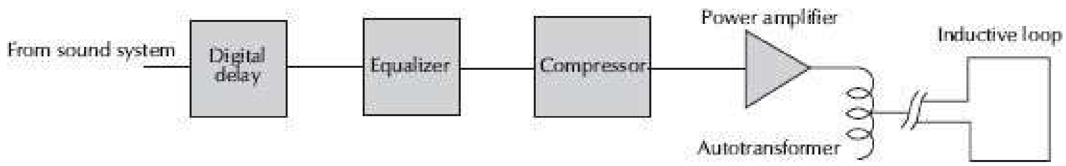


Figure 46-8. Induction loop system block diagram.

The FCC has opened the 216–217MHz band for assistive listening devices. This band falls under the *low power radio services* (LPRS) of the FCC, which limits the power output to 100mW. While the 72MHz band could transmit 500–1500ft, the 216MHz band can transmit 1000–3000ft.

The systems can be either wide-band or narrow-band. Wide-band systems have the following characteristics:

- High fidelity for all applications.
- Low cost.
- Good rejection of unwanted or external radio signals.
- Limited to six simultaneous channels.

Narrow-band systems have:

- High immunity to unwanted or external radio signals.
- More than eight simultaneous channels.
- Good fidelity for voice applications.

FM broadcast is much like wireless microphones; more information can be found in Chapter 20, *Microphones*.

An FM assistive listening system by Listen Technologies LS-16

ADA RF System includes programmable receivers, neck loops, antenna kit, and signage, Fig. 46-9. The LR-400 receiver can be programmed to receive only the channels available at the venue. This system ensures venues meet the current Americans with Disabilities Act (ADA) standards for accessible design.

The system has an *SNR* of 80dB and is available for 72MHz, 150MHz, 216MHz, or 863MHz band. The system includes one LT-800 stationary transmitter, an antenna kit, rack mount kit, four LR-400 programmable display receivers with ear speakers, two neck loops, and an assistive listening notification signage kit.

The Personal System by Listen Technologies includes a portable transmitter and receiver in a soft-sided carrying case. The Personal System's LT-700 portable transmitter, lapel mic, LR-400 display receiver, and ear speaker all fit in a soft case so they can be taken to school, house of worship, or theater. The listener gives the transmitter with its microphone to the presenter and the listener uses the receiver with an LA-166 neckloop or LA-164 earphone. The neckloop generates a magnetic field that is picked up by cochlear implants or hearing aids that are equipped with a T-coil, Fig. 46-10.

The Sennheiser Mikroport 2015 is suited for classroom use and allows a hearing impaired student to have an improved learning experience via wireless audio connection to the teacher. The system includes a wireless transmitter with a lavalier microphone worn by the teacher and a body worn wireless receiver for the student. Direct audio input cables are available for use with cochlear implants and hearing aids and induction neck loops for use with T-coil hearing aids. The systems can also be used with standard headphones or ear buds. Multiple receivers can be used with a single transmitter and, because there are hundreds of discrete frequencies available, systems can be used in adjoining classrooms without crosstalk or interference, Fig. 46-11.

Figure 46-9. An installed FM assistive listening system. Courtesy Listen Technologies Corporation.

Figure 46-10. Personal system including FM transmitter, receiver, and battery charger. Courtesy Sennheiser Electronic Corporation.

Figure 46-11. Neckloop and earphone. Courtesy Listen Technologies Corporation.

46.2.4 Infrared

Infrared light can be used to broadcast a very high-quality signal. Presently available systems can broadcast up to twelve different programs on the same emitter, making infrared very useful for large-scale language translation systems. Infrared systems are also used in museums and for lecturing and teaching on auscultation, the listening of heart beats. Systems are also produced for home listening, for both stereo and for video (TV). It is the only system for the hearing impaired that can transmit in stereo. Also unlike the other systems, infrared broadcasts are completely contained within the room because infrared behaves like visible light; it cannot go through a wall; even a heavy cloth is an opaque barrier. This control of the broadcast range is a significant factor where confidentiality is important, as in corporate meeting rooms. For this reason and because of its outstanding sound quality, infrared is the system of choice for professional theaters and concert halls.

46.2.4.1 Components of an Infrared System

The basic components of an infrared system are:

- The control transmitter (which is connected to the audio source)
- Slave emitters, daisy-chained together (if needed).
- The receivers.

46.2.4.2 Categories

Any installation generally falls into four categories:

1. Personal listening systems (PLS) for living rooms, bedrooms, offices, etc.
2. Medium area systems (MAS) for lounges, meeting rooms, courtrooms, classrooms, small theater, churches, etc.
3. Large area systems (LAS) for auditoriums, large theaters, churches, arenas, etc.
4. Large area multichannel Systems (up to 32 channels) for simultaneous interpretation and other applications.

46.2.4.3 Coverage

Transmitter and emitter panels with present state-of-the-art design and components allow over 70ft^{2}of coverage per IR light emitting diode in single channel IR systems. The shape of the polar pattern of a panel is nearly identical to the pattern of a single diode. The half-power angle of luminosity for the presently used LEDs is approximately ±25°. More LEDs increase the IR intensity, and the area of coverage multiplies by the number of diodes in the panel. The radiation pattern can be considered the same for horizontal and vertical orientations, and is scarcely influenced by the arrangement of diodes or the housing of the array.

Since there is a physical limit to the light output power of any LED, the total output has to be shared between channels in a multichannel system and the available coverage area per radiator has to be divided by the number of channels in the system. Conversely, for the same required coverage, the number of radiators should be multiplied by the number of channels in the system, or twice the amount of emitters in a stereo system as would be required in a single channel installation in the same venue.

Reflection and scatter off walls, ceilings, floors and furnishings broaden the coverage and make it largely nondirectional. Infrared light behaves a lot like visible light as it reflects best off bright and smooth surfaces like white walls, and is absorbed by dark and rough materials like black velvet curtains.

Emitters should be placed in a manner to provide even illumination throughout the room. They are usually mounted 10–40ft above the floor and pointed down toward the audience. When emitters are placed on both sides of the stage, they should be cross-fired into the audience. Any number of receivers can be used in the system as they will not affect the signal source.

LEDs have degradation of light output over time, however, by using good electronic circuit design, a projected continuous operating life under standard conditions of more than 100,000h, before the light output diminishes to 70% of its original value, can be obtained.

46.2.4.4 Ambient Light

Infrared systems work in virtually any environment except for direct sunlight. Systems can even be installed in shaded outdoor areas. Rooms with very high ambient light levels or poorly filtered fluorescent ballasts may require additional emitters for a sufficient *SNR*.

46.2.4.5 The Infrared Link

Infrared systems can be either narrow band or wide band, depending on your requirements. Table 46-2 gives the technical specifications for infrared systems.

The infrared link uses a specially doped gallium arsenide light emitting diode (LED) to transmit the signal. Each diode emits about 10mW total radiant power, requiring up to 143 LEDs in each array to produce adequate power. The wavelength of the emitted light is 930nm and is neither monochromatic nor coherent, so any number of diodes can be used together without interference between them. The useful coverage pattern of the emitter varies with distance and the number of channels being transmitted, Fig. 46-12. The number of emitters required depends upon the size and shape of the area to be covered and the number of channels in use. Emitters are usually employed in pairs, located at each front side of the audience and cross-fired across the seating area so that each person receives an infrared beam from each side. This cross-firing helps to eliminate shadowing from other people in the audience.

Table 46-2. Technical Specifications for Infrared Systems

Characteristic	Narrow Band	Wide Band
No. of channels	12	2
Carrier frequencies	55–535 kHz, excluding 455kHz	95kHz and 250kHz
Channel spacing	40kHz	155kHz
Modulation	FM	FM
Pre-emphasis	100µs	50µs

Normal deviation	±6kHz	±35kHz
Peak deviation	±7kHz	±50kHz
Transmitters		
Frequency response (–3dB)	50Hz–8kHz	50Hz–13kHz
Max. distortion (1kHz)	<1.0%	<1.0%
SNR (A weighted)	>55dB	>70dB
Receivers		
Frequency response (–3dB)	50Hz–8kHz	100Hz–9kHz
Max. distortion 1kHz)	<2.5%	<1%
SNR (A-weighted)	>55dB	>63dB
Emitter Panels		
Frequency response (–3dB)	30Hz–710kHz	30Hz–710kHz

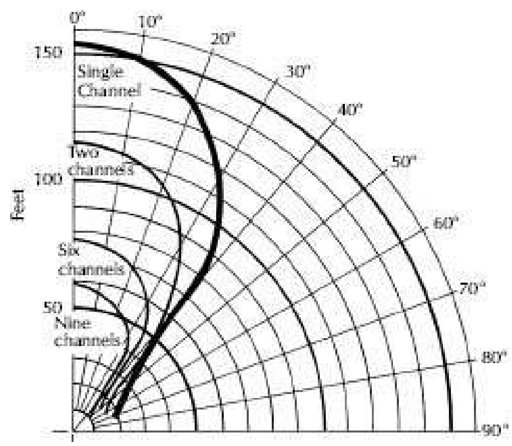


Figure 46-12. Infrared emitter panel coverage patterns.

The receiving end of the infrared link is a silicon photo diode that is reversed biased and produces current when struck by photons. The light gathering area is small, 7mm^2, but is effectively increased by mounting it in a collecting lens, Fig. 46-13.

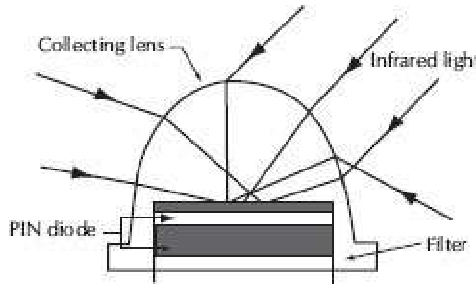


Figure 46-13. Infrared receiving diode.

The silicon PIN receiving diode has a maximum sensitivity at a wavelength of 850nm. Fig. 46-14 shows the spectral sensitivity of the eye, IR LED transmitting diode, IR filter, and receiving diode.

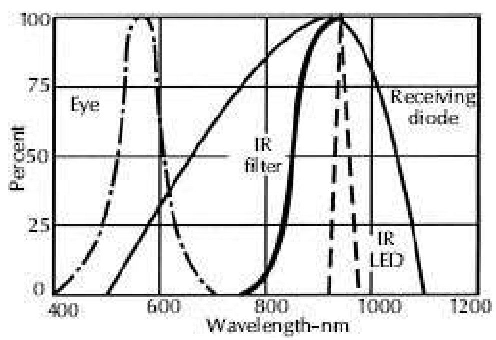


Figure 46-14. Sensitivity of the eye, IR LED transmitting diode, IR filter, and receiving diode.

Infrared light behaves much like visible light; it reflects off of light-colored walls, ceilings, and other surfaces so a receiver can "see" the signal even without direct line-of-sight to the emitter. Also, the receiver's ultrawide-angle fisheye lens captures direct or reflected signals from almost any direction.

There is a unique limitation to the use of infrared: it cannot be used in bright daylight. The infrared light occurring as a natural part of daylight will override the lower power-modulated light from the system. The system can also receive interference from very highlevel incandescent lamps. Partially dimmed incandescent lamps can also be a problem in some situations because the reduced voltage to the lamps causes a shift toward red that greatly increases the infrared output of the lamp. This increase in infrared interference can, on rare occasions, be a problem where audience down lights are left at a dimmed setting and the infrared beam from the system is weak. Deep under a balcony is a likely trouble spot. When this problem occurs, it is necessary to dim the lights more or add more emitters to the infrared system to cover under the balcony.

An infrared system comprises three sections: the transmitter, the emitter (sometimes both combined in one unit), and the receiver. The transmitter imparts the audio signal onto a subcarrier signal which the emitter converts into infrared light. The receiver decodes the infrared signal to retrieve the original audio.

To achieve a usable radiated power level, the IR LEDs are used in multiple arrays. Their light output is amplitude-modulated by one or more frequency-modulated subcarriers (typically 95kHz for single-channel wideband systems; 95kHz and 250kHz for two-channel systems). Each channel's audio signal frequency-modulates its particular subcarrier, Fig. 46-15.

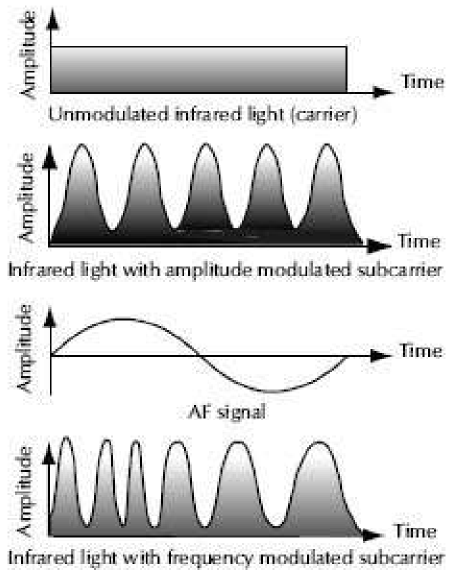


Figure 46-15. Infrared system's modulation technique.

Two transmission modes are available: wideband, for one or two channels of high-fidelity audio; or narrowband, for up to twelve channels with a 70–7000Hz response suitable for communications, Fig. 46-16.

Since the transmission medium is a modulated carrier of harmless invisible light, Fig. 46-17, instead of radio or audio signals, it is immune to outside interference and also causes none itself. No operator licensing is required for use of infrared systems.

Manufacturers provide detailed instructions for planning and installing the system. Typical installations are shown in Fig. 46-18. The rear emitters must cover both under the balcony plus the balcony. For this reason, separate emitters may be required to cover both areas. The advantages of infrared systems are fully realized in applications where different audio programs are required in adjacent rooms, such as a multi-cinema complex. Each room can be equipped with the same system without interference among them. No frequency coordination is required as with radio frequency systems. The same receiver can be used in any theater.

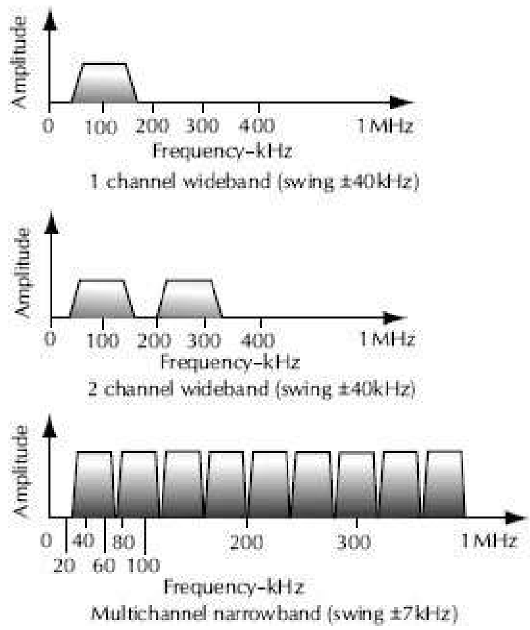


Figure 46-16. Infrared systems channel allocation.

Figure 46-17. Light spectrum.

A useful tool in aiming the emitters is a low-cost, black-and-white television camera and monitor. Most monochrome television cameras have useful sensitivity in the infrared region. With the room lights off, observe on the television monitor the part of the room illuminated by the infrared beam. The well-illuminated area will be the area of good reception. A corollary to this procedure is that the infrared television viewing system can be used to view a darkened stage—for instance, for coordination of rigging and prop moves in a fast, complicated change in the dark.

46.3 Receivers

Receivers are required with all systems, though fewer are needed with an induction loop because many patrons will have aids equipped with T-coils. Most manufacturers of systems for the hearing impaired offer several types of earphones with their receivers. Typically these include a single earpiece, a stethoscope-type dual earpiece, Fig. 46-19, and an induction loop for use with the patron's T-coil. Most users report a strong preference for the dual earpiece instruments and also report a strong dislike for an over-the-head-type headphone because they are uncomfortable for long periods of use and destructive to hair styles. Two types of induction loops are available. One is a small coil that hooks over the ear close by the hearing aid; this type is often hard for elderly people to use properly. The more popular loop is a lanyard type that hangs around the neck and may be used to support the receiver.

Many theaters and churches that have installed a system for the hearing impaired have found normal-hearing patrons using the system to enhance their listening comfort, which is especially true in larger houses with seats in areas of poor natural acoustics. These normal-hearing patrons universally prefer the dual earpiece both because it sounds more natural and because a single earpiece leaves one ear open to receive the live sound from the stage with a signal-delay annoyance. Depending upon the distance from the stage, this delay can be very distracting to those with less than profound hearing impairments.

Battery replacement and earpiece sanitizing are the principal maintenance problems with all systems for the assisted listening devices. Batteries may last up to one year, although that seems an uncommonly long life if the receivers are being used often. The infrared receivers have rechargeable batteries, which should be recharged after each use. Earpieces are most commonly sanitized by replacing the plastic ear tips or by using replaceable foam balls.

Most theaters and concert halls provide receivers to their patrons for no charge or for a small fee to cover the cost of handling, batteries, and sanitizing. Some organizations have been successful in selling receivers to regular patrons, especially in communities where several theaters and churches use the same technology.

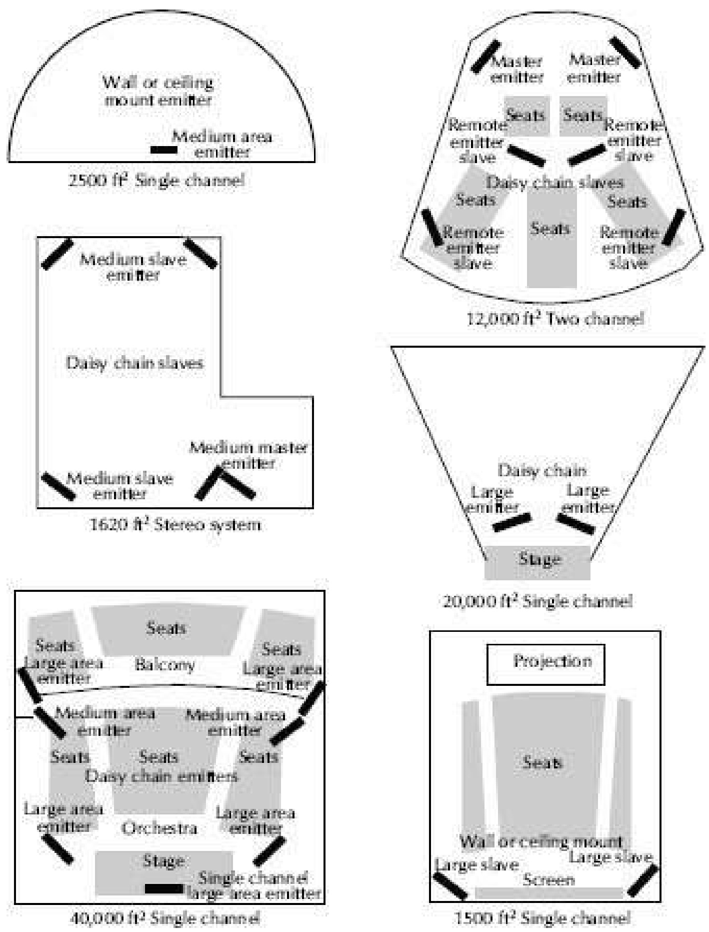


Figure 46-18. Typical infrared system theater installation.

Figure 46-19. A wireless IR system receiver under-the-head multichannel headset. Courtesy Sennheiser Electronic Corporation.

References

1. M. C. Martin, "Power Requirements for Inductive Loops," *Royal National Institute for the Deaf*, London, England.
2. S. C. Dalsgaard, "Field Distribution Inside Rectangular Induction Loops," State Hearing Centers, Research Laboratory for Technical Audiology, University Hospital, Odente, Denmark.
3. D. Bosman and L. J. M. Joosten, "A New Approach to a Space-Confined Magnetic Loop Induction System." *IEEE Transactions on Audio*, May/June 1965.

Bibliography

K. Borrild, "The Induction Loop and Its Possibilities," The State Boarding School for Hard of Hearing and Deaf, Frederica, Denmark.

E. deBoer, C. Kruidenier, and K. W. Kortschot, "On Communication Systems in Schools for the Hard-of-Hearing," *Acustica*, Vol. 19, no. 3, 1967–1968.

RNID Scientific and Technical Department, Installing a Loop System. "Hearing," *Royal National Institute for the Deaf*, October 1977.

Chapter 47

Intercoms

by Glen Ballou

47.1 Introduction

Intercoms, short for intercommunication systems, are often more sophisticated than sound reinforcement systems because sound systems are unidirectional while intercoms are bidirectional. Sophistication varies from the simple home intercom to hotel and apartment intercoms, to hospital nurse call and emergency systems, to partyline systems, to the multiprocessor controlled, multimemory, analog and digital intercommunication systems.

47.2 General Purpose Intercoms

Intercom systems fall into three categories: point-to-point, partyline, and matrix, Fig. 47-1. The *point-to-point* or *dedicated line* is a private line between two stations and is not often used today. Normally, no other stations can hear your conversation.

The *partyline, conference line*, or *distributed line* is a shared line where a number of people, usually sharing a common task, are all talking to each other. This is like being at a conference table where everyone can talk and listen to one another at the same time and there is no privacy capability. By having multiple circuits and multiple wires, this system can communicate to different parties at the same time. Most often this type is found in broadcast intercoms.

Most major intercom manufacturing companies are offering digital *matrix* systems. These systems are multiprocessor controlled, multimemory, and have analog and digital audio and can combine the point-to-point and partyline communications together. All have advantages and disadvantages and they can be either hard-wired to local stations and interfaces or connected over an IP network or a combination of the two.

Intercom systems can be simple, a two unit system with only a call button and a microphone/loudspeaker combination in each unit, to a complicated multichannel multistation system with a separate microphone and loudspeaker, display window, keypad, TV entrance monitor, auxiliary inputs, can be programmable, plus have a multitude of special features.

47.2.1 Point-to-Point Systems

Point-to-point systems are the simplest type of intercom and are mostly used in residential and office applications, schools, apartments, and nurse call and emergency call systems. With point-to-point, a caller or originator makes contact with the desired receiver or receivers and communicates only with them, Fig. 47-1.

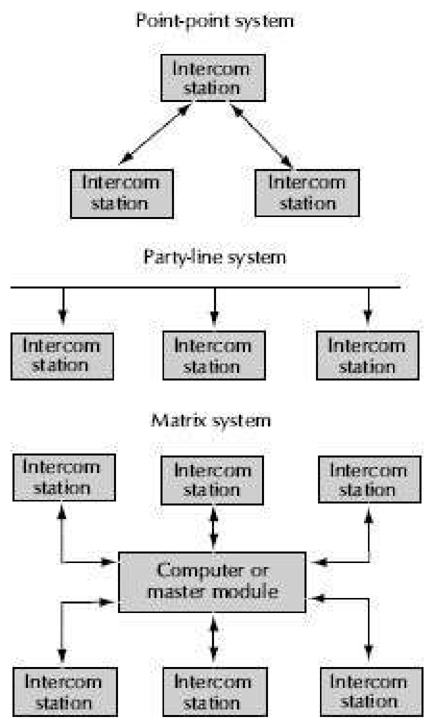


Figure 47-1. The three basic types of intercoms.

All other stations are isolated so they cannot be part of the one- or two-way conversation. This is like the telephone system, you dial a party or parties through a conference call, and the conversation can only be heard by those parties that were called up even though the telephone system has hundreds of thousands of other phones on the system.

Systems are normally made up of master stations, slave stations, door monitoring and opening stations, and input devices such as AM/FM receivers.

These systems have a variety of features and are used for intercommunications between rooms or desks, intercommunications between an indoor station and an outside door including door release, and provide some form of playback signal at all or selected stations in the system. The system can be one master to one or more slaves, all master, or a combination of the two. Masters have the ability to originate calls to any or all stations while slaves can only call the master station they are connected to, Fig. 47-2. Stations can be in-the-wall units or self-contained desktop units, hardwired or wireless. Normally hardwired systems are the best for point-to-point systems as they are less expensive and less likely to have interference. The disadvantage is they are more difficult to move or reposition.

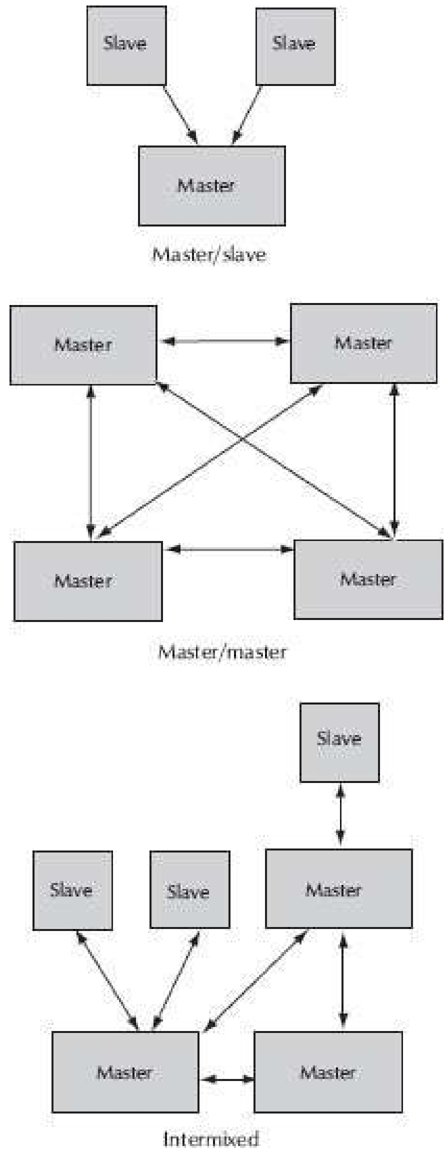


Figure 47-2. Master/slave, master/master, and intermixed intercoms.

The Bogen Model PI35A, Fig. 47-3 is a 25-station intercom. It provides facilities with two-way communication, emergency paging, time tones, and background music or other audio program material to speaker-equipped locations. The Bogen PI35A is designed to meet high power paging and intercom requirements with features suitable for applications with mixed noise environments (construction, retail stores, small factories, parking garages, etc.). A

20W intercom amplifier features a voice-shaped frequency response for intelligibility. A 35W program amplifier is used for program material and/or emergency announcements.

Figure 47-3. Bogen P135A 25-station intercom. Courtesy Bogen Communications, Inc.

Program material from microphones, a CD player, or other background music source can be used. Distribution is accomplished by simple push-button selection. Emergency announcements take precedence over program distribution and are accomplished with a single push-button selection. A time signal can also be sent to all stations. Optional paging facilities permit emergency all-call paging from a remote telephone, or interphone or microphone. Telephone paging captures system priority and overrides all system functions except the emergency page feature.

A room selector panel is provided to select intercom and program functions for each station. Calls from stations are initiated through call-in switches in the various rooms and are announced at the control center by light and tone annunciation. The system provides a 25V balanced output and operates from a 120Vac, 60Hz source. The system consists of a master control panel and a twenty five-station room selector pane. A number of options are available including room call-in switches, call-in adapter modules for existing older call-in switches, two-wire call-in adapters, and various styles of transformer-coupled loudspeakers.

47.2.2 Matrix Systems

Not too many years ago, all intercoms used a mechanical switching matrix to call and to route the voice signal. Systems were simple, wires came from a central system and went out to the individual masters and submasters. Switches were multipole and carried signal and voice. Often the voice line was shielded to eliminate hum and stray noise pickup, however, on inexpensive systems unshielded lines were used and the frequency response of the receiver was limited to 200Hz–4kHz to eliminate noise and hum.

While these systems are still used today, many have been replaced with digital matrixes and electronic switching. These systems use low voltage/low power and because the signal is digital, seldom require shielded lines.

Matrix systems can control upwards of 256 station units and can have over 10,000 stations by tying many matrices together.*

Because they are digital, it is easy to set them up with quick dial single or double digit dialing of often used numbers, connect them with outside C/O or PBX lines, and record a customized outgoing message for callers and record their response, retrievable at any master. Because they are telephone system compatible, they are capable of supporting a DTMF generating, single line telephone instrument. Interconnections are through twisted pairs of copper wire or through fiber-optic cable.

Many of these systems are programmed with a computer. Through simple programming, only selected stations can be given access to special functions, i.e., paging, telephone calls, priority calls, external sources, etc. Compression circuits and/or automatic leveling circuits can be adjusted for individual units as well as their overall level.

47.2.3 Apartment Security/Intercom Systems

Apartment systems are usually point-to-point and are used for the visitor to contact the apartment owner for access into the building and then into the apartment. This system consists of a master panel outside the main access door to the building and slaves in each apartment, Fig. 47-4.

The outside master panel must be waterproof and vandal proof so the outside unit panels and buttons must be made of strong material such as aluminum, stainless steel, or LEXAN®. The microphone/loudspeaker grill must be indestructible and shed water, preferably a continuous part of the panel and the loudspeaker must be waterproof. The panel must be fastened to the wall with vandal-free mounting hardware. The outside master panel consists of a microphone/loudspeaker and a means of contacting the person of choice. This can be accomplished by a series of push buttons that connect the visitor to the desired apartment or it may be a twelve button telephone panel that connects the caller to the suite through digital circuitry.

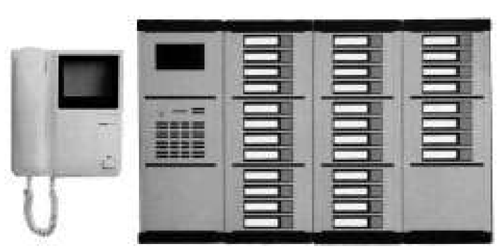


Figure 47-4. Audio/video entry security system. Courtesy Aiphone Communications.

A power supply and amplifier is mounted inside where it is out of the elements. Only one amplifier and power supply is normally used for a system because only one conversation goes on at the same time. The door opener usually operates on 3–6Vdc or 8–16Vac.

The apartment loudspeaker station consists of a microphone/loudspeaker unit, a talk button, a listen button, and a door release button.

The more sophisticated security systems include a video monitor for improved safety. The outside door panel includes a camera and the indoor units include a 4inch flat screen monitor. A pan and tilt capability improves security as it can scan a large area. Because the camera is behind a glass, the visitor cannot see where it is aimed so the visitor cannot hide from it. With today's technology, wide angle cameras are possible, eliminating the need for the more expensive and complicated pan and tilt. The cameras require very good low light level operation as most access is in the evening or at night. One lux sensitivity is normal. Most systems do not require coaxial cable between the cameras and the monitors.

Self-contained video door answering systems are rapidly becoming today's replacement for the common doorbell. They are a simple-to-install answer to the growing need for entry security in both small businesses and homes. Some use the same two wires as a doorbell, often using existing wiring in older buildings, and simplifying installation in new construction.

An example is a system where one pair carries an FM modulated signal with both audio and video, and a second pair operates the door releases shown in Fig. 47-4. In this system the door unit incorporates a high resolution infrared CCD (charge-coupled device) camera with 250,000 pixels, providing a clear, sharp picture down to one lux of light.

By using a wide angle lens, an overall viewing area of 39in × 27in at 20in away can be attained, Fig. 47-5. For more coverage, a PanTilt door station can be used, Fig. 47-5. This allows for a coverage of 72in × 36in at 20in. Like any transmission system, cabling is important. The system was designed to operate with two wires such as typical bell wire. Coaxial cable or two separate wires or multicable will affect picture quality. If the cable has more than one pair, the other pairs should be terminated at each end with a 120Ω resistor. This keeps a proper impedance of the line.

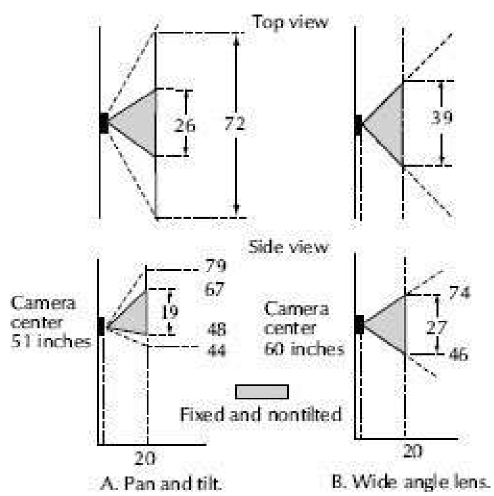


Figure 47-5. Video coverage door stations. Courtesy Aiphone Communications.

Any door entry system camera should be mounted so the unit is not exposed to environmental extremes. Sudden temperature drops, for instance, can cause the camera to fog up and raindrops can distort the picture. Lighting should be on the front of the person. Strong backlight, i.e., sunlight or street lights, can cause a silhouette image on the monitor, and fluorescent lights can cause flickering. Taking these into account will give a good picture under most conditions.

It is important that the door release button has normally open contacts to activate an optional electric door strike. The normally open contact assures that the door will not open during a power failure.

47.2.4 Residential Intercom Systems

Residential intercom systems are used to talk between rooms, or between a master unit and all other rooms, as a door security system and for programming external sources such as AM/FM or CDs to any or all rooms.

Because of the limitations of intercom systems, the music is usually not in stereo and not of the quality of a dedicated stereo system, but is quite adequate for background music. They also have the advantage that the same source is heard as a person walks between rooms. Normally residential intercoms are wall mounted, incorporate hands-free answering and include a privacy switch. The systems can be master/slave, master/master/slave, or all master.

47.2.5 Commercial Security Systems

A building security system assists security personnel in the protection of the lives and property of all tenants, employees, and visitors. People in parking areas, ramps, tunnels, stairwells, and elevators should have access to conveniently located, easily operated hands-free emergency call-in stations.

47.2.5.1 Zone Paging

High-rise buildings and multibuilding complexes have special needs for zoned public address announcements. Buildings with controlled access need audio voice confirmation and integration with CCTV cameras. Elevators require intercommunication to security and the lobby and communication to the elevator machine room for maintenance, Fig. 47-6.

Variations of this type of intercom can be used for campus security where there are multiple buildings, parking lots, dorms, and walkways.

47.2.5.2 Security Audio Monitoring

Unfortunately, all too often crisis situations do develop in the shadows of darkened areas or dimly lit passageways. These are the areas where criminals tend to stalk. These are also the places where a building's highly volatile power transformers, generators, and steam lines are neatly tucked away. Security and maintenance personnel can't be everywhere. Even with the assistance of video surveillance there are limits to the number of cameras used and manpower to monitor them. Even in the best of situations video cameras can't see around corners, through closed doors, or behind parked cars or trucks.

To eliminate problems with video only monitoring, security system manufacturers have developed listening devices and loud-noise triggering alarms. Although these products seem to be moving security systems in the right direction, they still have substantial limitations.

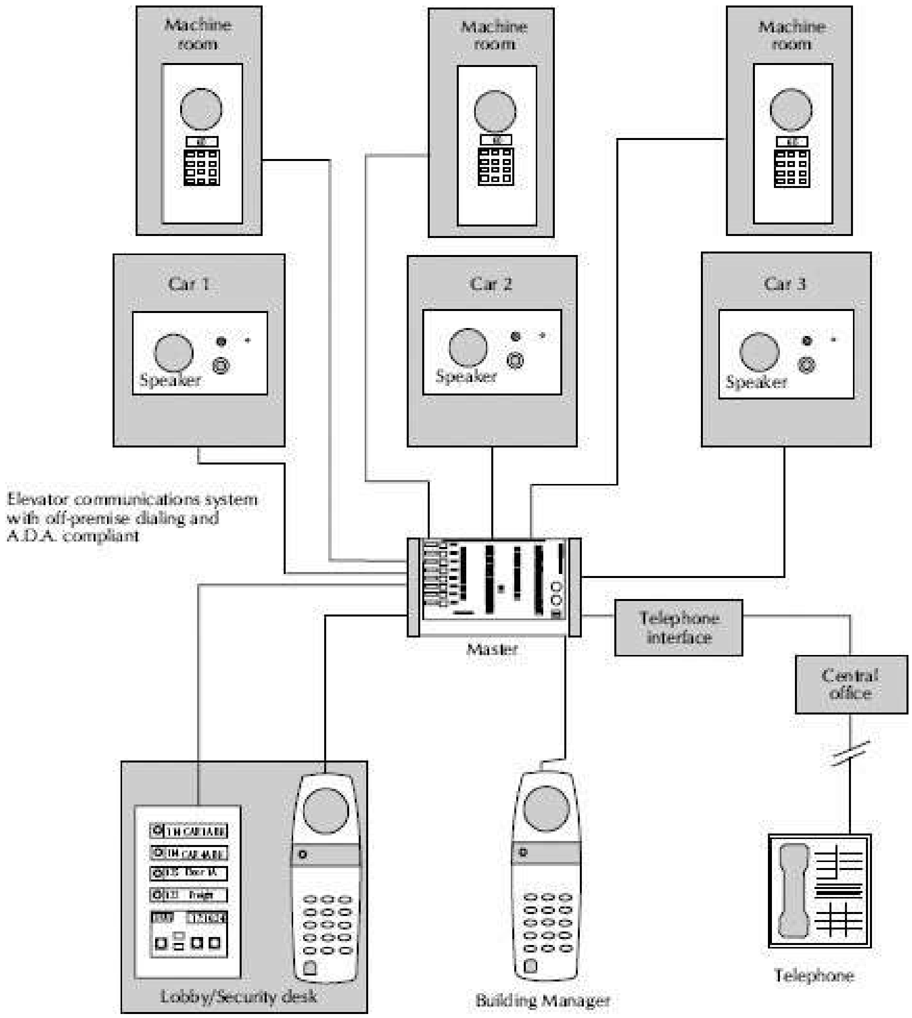


Figure 47-6. Basic building communication system. Courtesy Ring Communications, Inc.

Listening devices are a great idea because they allow security personnel to interpret and discriminate the sounds they hear. Unfortunately, like video cameras, they must continually be monitored to be truly effective. There is also the question of open microphones being construed as invading individual privacy.

Loud-noise triggering alarms have the problem of discerning sounds, for instance, screams from laughter, or the loud noise of a car engine starting up. Without the ability to discern specific sounds and discard normal background sounds, false alarms would render the system useless.

Today there are *smart* security devices that incorporate the best features of both a listening device and a noise trigger without many of the drawbacks.

One effective method is a sound sensing device that continually adjusts itself to background noise levels, while detecting and discerning the unique sound characteristics of specific sound patterns usually associated with crisis and emergency situations. When the system picks up one of the crisis sound patterns it fires and alerts security personnel of a probable emergency situation in a particular area. The system can also automatically turn on video cameras, sound alarms, or activate two-way intercoms, so that security personnel can instantaneously communicate with individuals involved at the scene.

Because this type of system continually adjusted itself to the continuous changes in levels and frequency response of background sounds, and has the ability to insert a time gate to further assist in doing away with false alarms, it is able to isolate the sound of a scream or the smashing of a pane of glass, Fig. 47-7.

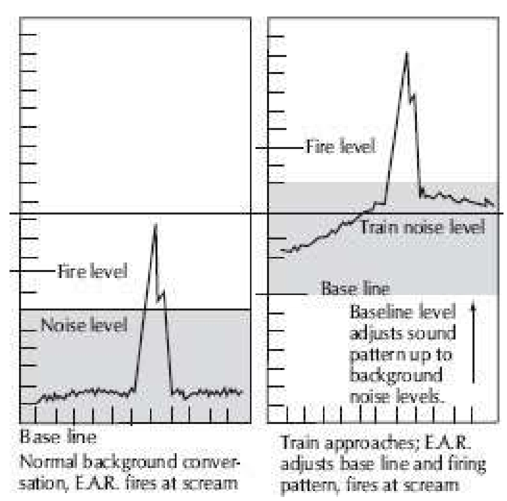


Figure 47-7. Simulated firing of E.A.R. in an active subway

station. Courtesy Ring Communications, Inc.

This type of monitoring is useful in security applications including correctional facilities and mass transit subway authorities, parking lots, schools, large warehouses, and empty office building hallways after regular business hours.

47.2.5.3 Emergency Crisis Communications

Methods of communications change, methods of transmission change, and methods of installation change. Manufacturers continue to upgrade their equipment to handle these changes.

At one time the basic accepted method of signal transmission was via copper conductors only, now there are multiple methods to accomplish this.

Voice communications is often overlooked in some areas, but an alert that is only a bell or horn or chime sounding only alerts people that there is a problem, it does not tell them what the problem is, where it is, or what response they should take to protect themselves. By adding voice communications to panic alarms, fire alarms, or crash alarms, the general public can be advised of what the nature of the alarm is, where it is, and what to do to avoid it.

Often huge geographical areas must be covered. To do this with copper alone is almost impossible. The Ring Master Intercommunications Crisis Alert System has the ability to tie equipment together to operate as one complete system, providing for line supervision over multiple methods of transmission, simultaneously solving any such problem.

Fig. 47-8 shows such a system that is in use today to cover the many and varied needs of an airport site.

The system provides communications for all aspects of airport communications requirements over one integrated supervised system. The system includes:

- Air traffic control.
- Security.
- Baggage handling.
- Staff location.
- Access to general over-head paging.
- De-icing.
- Flight operations and planning.
- ADA elevator communications.
- Door access and control.
- Emergency call boxes for the parking facilities.

When distances between buildings and operational sites are long, copper is not a viable solution. Fiber is better utilized to tie geographically distant areas together. In other areas the newest method of VOIP, over radio links, can be utilized to provide communications, for instance, to the people mover train system.

With this type of system, any station in the system, no matter where its geographical location on the airport site, can be programmed remotely, to call or be called by any other station in the system or to access any of the multiple functions and features of the communications system at any time.

When these connection capabilities are added to the commercial security intercom system there are virtually no restriction to fulfill the requirements for safe, secure communications.

47.2.6 Commercial/Industrial Systems

Commercial/industrial intercom systems are used in airports, car dealers, dormitories, factories, hospitals, nursing homes, department stores, and schools. Most of these systems are master/slave systems or subsystems.

The health care and hospital facilities systems must be designed for emergency hands-free operation from operating rooms, trauma rooms, delivery rooms, etc. The systems must also be capable of sterilization after each use. This can be accomplished with a Mylar covered face plate that is easy to sterilize because of lack of crevices, etc. The system can include a paging system for doctor, nurse, or security call, and for calling patients in waiting rooms.

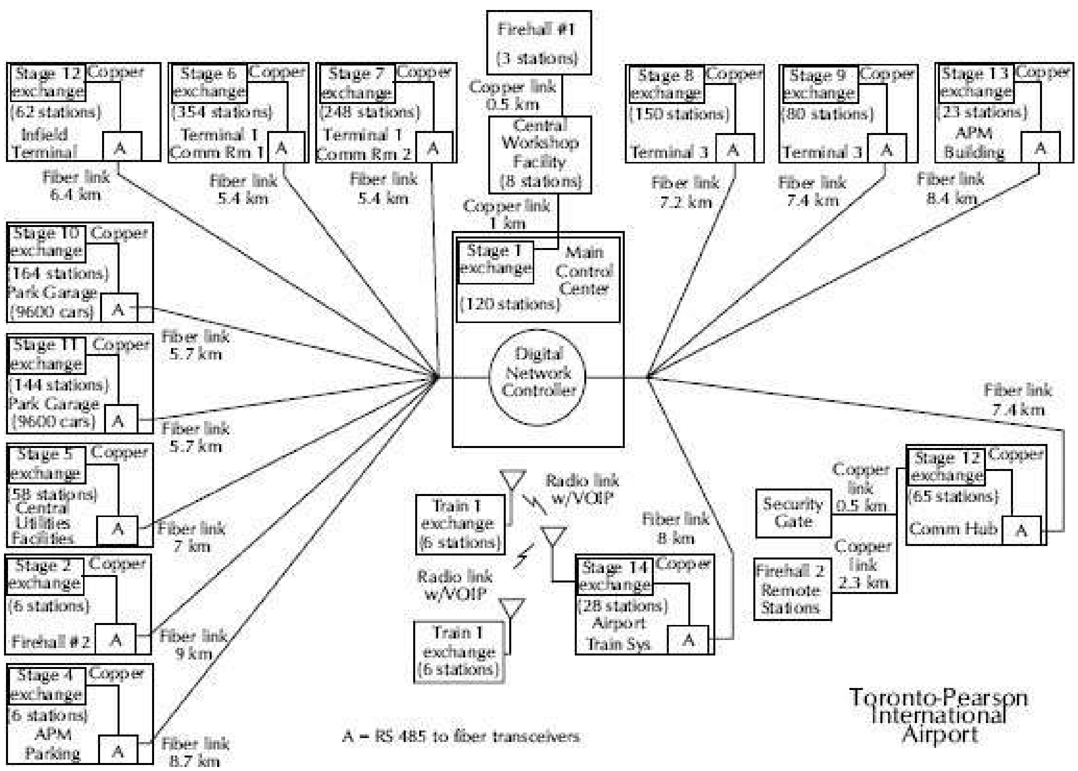


Figure 47-8. Ring Master Intercommunications Crisis Alert System at Toronto-Pearson International Airport. Courtesy of Ring Communications, Inc.

Operations of general commercial/industrial systems are very much like the door access or residential system but include many of the following special features:

Absence Registration. A unit can be programmed to display the station is unattended.

All Call. All call allows a master station to connect to every other station for an all-call message.

Call Back. Call back is initiated by the caller when the receiving line is busy. The caller can put the system on automatic call back which notifies him as soon as the intended receiver is free so he can reinstate the call or the system can automatically recall the receiver.

Call Forwarding. Call forwarding allows all calls directed to a station to be forwarded to another station.

Call Reply. When the called person is absent, the caller may dial a call-back signal which indicates that he has been called.

Call Transfer. Particularly useful for transferring a call to or from a secretary.

Camp On. Camp-on is a feature which allows the caller to camp-on to the called person when that person is in communication with someone else. Once the called person hangs up, the caller is automatically connected to him. Normally, camp-on only holds on for 10–20 s before dropping out, or may remain in service until the next call is initiated.

Central Answering Service. Units can be set up to go through a central answering person with automatic queuing of incoming calls.

Conference. Conference calls can be held with more than one party. Normally three to four parties is the maximum limit, however some systems may be expanded to a 30 party conference.

Display Features. Today's intercoms usually use an LCD type display system which gives the called or calling station number, in addition may display a series of preprogrammed messages, whether or not the called station is in privacy mode and time/date.

Group Callback. Many systems are setup so that various groups can be connected with one button so a person can call an entire group at one time.

Hands Free. In this condition, conversations are hands free from anywhere in the room as long as it is a duplex system. This can be changed to confidential at any time by a person picking up the handset and operating as a telephone system.

Hurry Up. If the conversation channel is busy, and the caller needs to make an emergency or an important call, she can transmit a special signal on the conversation channel which tells them there is an important message to be received.

Last Number Redial. Last number redial, as on telephone systems, allows the caller to redial his last called number again and again and again.

Microphone Cutout. Microphone cutout is used to disconnect the microphone during a conversation so that the receiver cannot hear communications between the talker and another person in the same room.

Paging. Master stations can page to other master stations, substations, and to remote loudspeakers through an external amplifier.

PBX Station Intercommunications. Many intercoms can be connected to the PBX or an outside C/O line for communications to the outside world.

Pocket Page Access. Any master station can place calls to the desired pocket pager receiver, up to 10,000 units.

Priority. If a priority station dials a number while the conversation channel is busy, she has the option to override the busy link and be connected directly through to the called party. This is normally used for emergency systems.

Privacy. Privacy is initiated by a person so that a caller, when calling that person, hears a signal which tells him the receiving person does not want to be disturbed.

Program Distribution Channels. Music or special messages can be setup to be used as background music.

Scan Monitoring. Scan monitoring allows a control station to arbitrarily scan a group of slaves or substations for auditory monitoring. Scanning can be performed either manually or automatically.

Time/Date. Some intercoms have the capability of displaying the time and the date.

Nurse Call Systems. Nurse call systems must be fast, reliable, and most of all easy to use. Patients must not have to figure out how to use a system to call for help. The system must not interfere with the complicated patient monitoring systems and must not produce ground loops between it and other equipment.

The simplest and probably best method for a patient to call a nurse is with a pull chain. This chain, actually cord, electrically isolates the system from the patient and only requires a gentle pull by the patient to operate. The cord can be draped on the rail of the bed for easy access by the patient. A monitoring system is located at the nurse's station with lights corresponding to room numbers or it could be a CRT with various messages. An indicator light outside the patient's room is also energized, allowing a nurse in the hallway to answer a call without going back to the nurse's station. The call remains energized until the nurse turns it off from the patient's station. Because the nurse call systems only tell the nurse which room is calling, dual-bed rooms have one call station with two call cords.

Nurse call systems can also include master/slave two-way communications, eliminating a nurse's extra steps. It also allows a nurse station to monitor a room. Some areas require an emergency button. This could be a pull chain, a large red mushroom button, or in the case of a psychiatric station, it could be key operated.

47.2.7 Wireless Intercoms

When areas are constantly being reorganized, or a user is required to be mobile over several stations, a wireless intercom may be useful. A wireless intercom uses RF and two frequencies are required per unit, one for talking and one for listening unless the system operates like two-way walkie talkies where simultaneous two-way communication is not possible. If each unit must have its own private communication, then each unit would require a different transmit and receive frequency or use a system of subaudio modulation to key the appropriate unit. The subaudio modulation system can have only one conversation at a time. The biggest drawback of wireless intercoms is their ability to pick up noise and stray signals and fewer bells and whistles are available. However, if the area is confined and relatively free from electrical noise, wireless intercoms can greatly reduce installation costs.

47.3 Broadcast Intercoms

The need for rapid, reliable, and flexible communications is required in broadcast intercoms. Fortunately, matrix or Partyline intercom equipment is capable of meeting almost any need that might arise.

Telephones are usually used for less than 20 minutes per call. Intercoms are often used for many hours at a time. The people involved in teleproduction work often can't take a break or remove their headsets. If the system has limited frequency response, then the system's filter effects create distortion. This unnatural sound can cause fatigue, which can be eliminated with a full-frequency intercom. Telephones are also "blocking" disallowing multiple users from communicating together and users are blocked from accepting new calls.

Broadcast intercoms can be partyline, or matrix. The audio line can be balanced or unbalanced. Balanced line operation provides maximum protection from electromagnetic interference generated from sources such as fluorescent lights, patch panels, or light dimmers. A balanced system can run as much as 5000ft with standard two-conductor shielded microphone cable.

An unbalanced line, sometimes called a three-wire system, is easier to switch and operate special circuits as the audio and signal are on different lines, only using the ground (shield) as a common line. A second method provides two channels of unbalanced audio, one channel between each conductor and shield, and combining the dc operating power with one of the audio channels.

Many intercoms are engineered to receive phantom power from the power supply. Stationary or permanent user stations usually operate in a dry mode as they are supplied with power from a local power line. The term *dry* refers to an intercom channel that has audio but not the usual phantom power on the channel. Dry operation have several advantages over wet operation as it is generally quieter, reduces the need and cost of large system power supplies, and takes up less master rack space. System configurations can include a mix of wet and dry channels, depending on the station equipment assigned to the particular channel. Generally, most wired belt-pack and loudspeaker stations require a wet channel and thus need a system power supply.

47.3.1 Broadcast Partyline System

In the partyline system, each station is equipped with all of the required electronics for receiving and transmitting audio and for signal routing. Partyline systems require minimal centralized rack equipment, which usually consists of the system power supplies and passive assignment switching in multichannel systems.

Partyline systems can use analog audio with control signaling or be fully digital.*

Partyline systems allow groups of stations to communicate in real-time, full duplex fashion. In fact as it is a partyline, all units on the line hear and are part of all conversations. Multichannel partyline systems allow users access to several different channels, allowing them to determine which line they talk and listen to. Normally there are no private communications such as point-to-point matrix systems provide.

Two-channel, dual-listen, with monaural output intercoms with programmable switching let the user listen to both channels simultaneously and select which channel to talk on. They include an individual volume control for each channel, microphone on/off control, a call signal button and indicators, and sidetone. These stations are ideal for ENG and EFP mobile production vans, production studio consoles, and smaller TV facilities.

Straight two-channel units allows simultaneous listening and talking on two intercom channels. The headphone output operates in a split-feed stereo mode, feeding each channel into a separate ear of a double-muff headset with an individual volume control for each channel. The operator can talk or listen on either channel, combine them, or access them separately without tying both channels together. Often a stage announce output can be supplied with relay control for external paging. Microphone or line-level program inputs may be assigned to either or both channels. The systems capabilities can include a selectable IFB *program interrupt* function, remote m*ic-kill* function (RMK), dual-action, electronic momentary/latching *talk* buttons, and a *no-fail* power supply with automatic reset and short-circuit protection.

A typical two-channel analog party-line system by Clear-Com consists of a main station, which includes a power supply. Each channel is full duplex allowing everyone on the channel to choose to both listen and talk to everyone else. Although the min station provides two channels, single- and two-channel beltpacks can be mixed within the same system. This reduces cost and allows one to control exactly who is assigned to what channel. All cabling is standard low capacitance shielded microphone audio wiring, Fig. 47-9. A partyline belt pack is shown in Fig. 47-10.

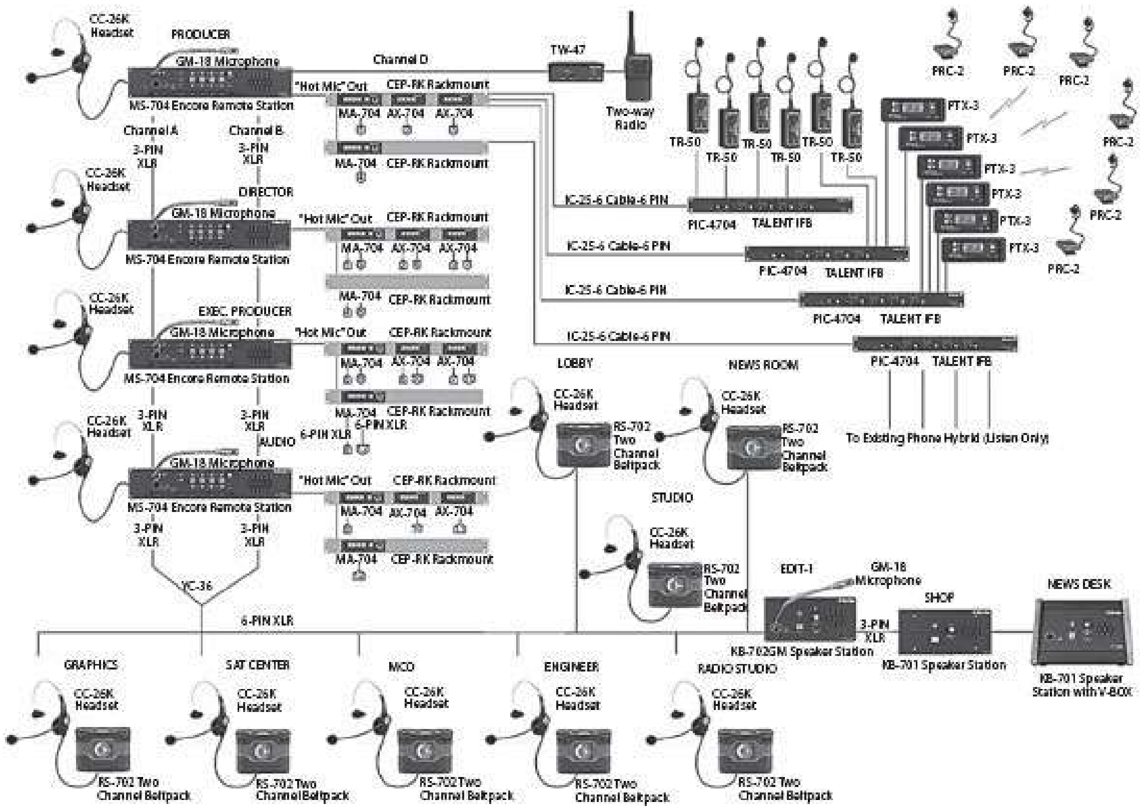


Figure 47-9. Two-channel party line system. Courtesy Clear-Com Communication Systems.

Some multichannel systems have extensive programming capabilities, which allows individual stations to be customized by storing the *button setups* in nonvolatile memory. Many programmed setups can be stored, thus allowing quick and easy switching between setups for rehearsal and performance or shows and events. Individual button assignments can be stored in *presets* for instantaneous recall. When programming this equipment, messages prompt the operator through the programming sequence, simplifying station setup.

A four-channel, two-wire partyline system is similar to the two-channel partyline except with two additional channels. In Fig. 47-11, Channel A is for the production crew, Channel B production and all remote graphics and engineering rooms, Channel C external remotes only including dressing rooms and lobby wall box temporary beltpack positions with Channel D calling the 2-way wireless users such as runners and scenery. Note how the hot mic outputs from the Producer and Director stations are used to feed the IFB interfaces to the studio talent wireless IFB receivers. All cabling is standard low capacitance shielded microphone audio wiring.

Figure 47-10. A partyline belt pack. Courtesy Clear-Com Communication Systems.

Figure 47-11. Clear-Com HelixNet digital Partyline system.

Courtesy Clear-Com Communication Systems.

Digital Partyline systems operate in much the same way as analog systems but have the huge advantage that multiple digital channels can use a single mic line or CAT5 cable, Figs. 47-12 and 47-13. This means that existing cable can be used but give more capacity as system requirements grow. Also because the audio is digitized the user does not have to “null” the line every time it is extended or re-routed to tune out the changes in impedance and capacitance. Digital audio has a superior audio performance and can utilize networks for extending over greater distances. Channels can be labelled so that callers can be identified and set-ups are easier to manage.

Figure 47-12. Delta-HX digital matrix system. Courtesy Clear-Com Communication Systems.

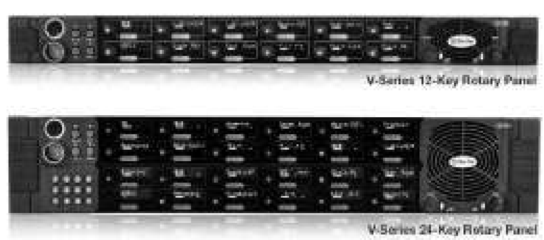


Figure 47-13. Clear-Com Matrix Stations, V-Series. Courtesy Clear-Com Communication Systems.

47.3.2 Broadcast Matrix Systems

Matrix systems are characterized by having multi-key user stations connected back via individual home runs to a central switching matrix which is designed such that station users can make point-point, one to many, conference or partyline and control operations in any combination at the same time as set out in the matrix database. In the past the connections between user stations and matrix frames was all analog but today most are digitally connected, such as IP or AES3 audio, or use a combination of analog audio with digital control for key data.

No user station ports may use analog or digital interfaces. The standard analog wiring format provides wide audio bandwidth and easy connection over standard four-wire audio and data circuits. It allows user- to-matrix communication through repeaters and over satellite links, TI circuits, and fiber optic systems.

The fully digital format uses only a single pair of wires that carry data plus digitized audio, simplifying installation. Additionally, digital audio cannot be tapped by unauthorized persons, it offers total noise immunity from external sources. The wiring is easy, it permits standard punch-block connection techniques using existing unshielded multipair telco or CAT 5 wiring.

The digital matrix system differs from partyline in several areas. In addition to partyline conversations (often called conferences in matrix systems) one may also conduct private point-to-point conversations between the panels. Through intuitive programming, members of the groups may be changed at will. This system easily integrates telephones, two-way radios, line-level audio in/out, voice over IP, GPIs, and relays, which allows incoming and outgoing telephone calls to be routed to particular panels or groups. Matrix systems are well suited for dynamic broadcast environments where IFB audio feeds must change quickly. A wireless intercom is easily integrated and can be treated as another panel depending on the wireless version in use. The connections between the mainframe and control panels are standard nonshielded Cat5 or 6 cables. Digital matrix systems by Clear-Com are shown in Figs. 47-12, 47-13, and 47-14.

Matrix systems use crosspoint and CPU circuitry. The mainframe serves as the central interconnection point for the control stations, interface modules, power supplies, the configuration computer, and external audio and control equipment. All signals, digital and analog, are processed in the mainframe and routed according to the current software configuration program.

The CPU operates the frames and control all of the system data communications. Crosspoint electronics contain microprocessors that communicate with the CPU and with the stations. The crosspoint circuitry supports individual ports that can connect to stations, interfaces, or to analog four-wire circuits and devices.

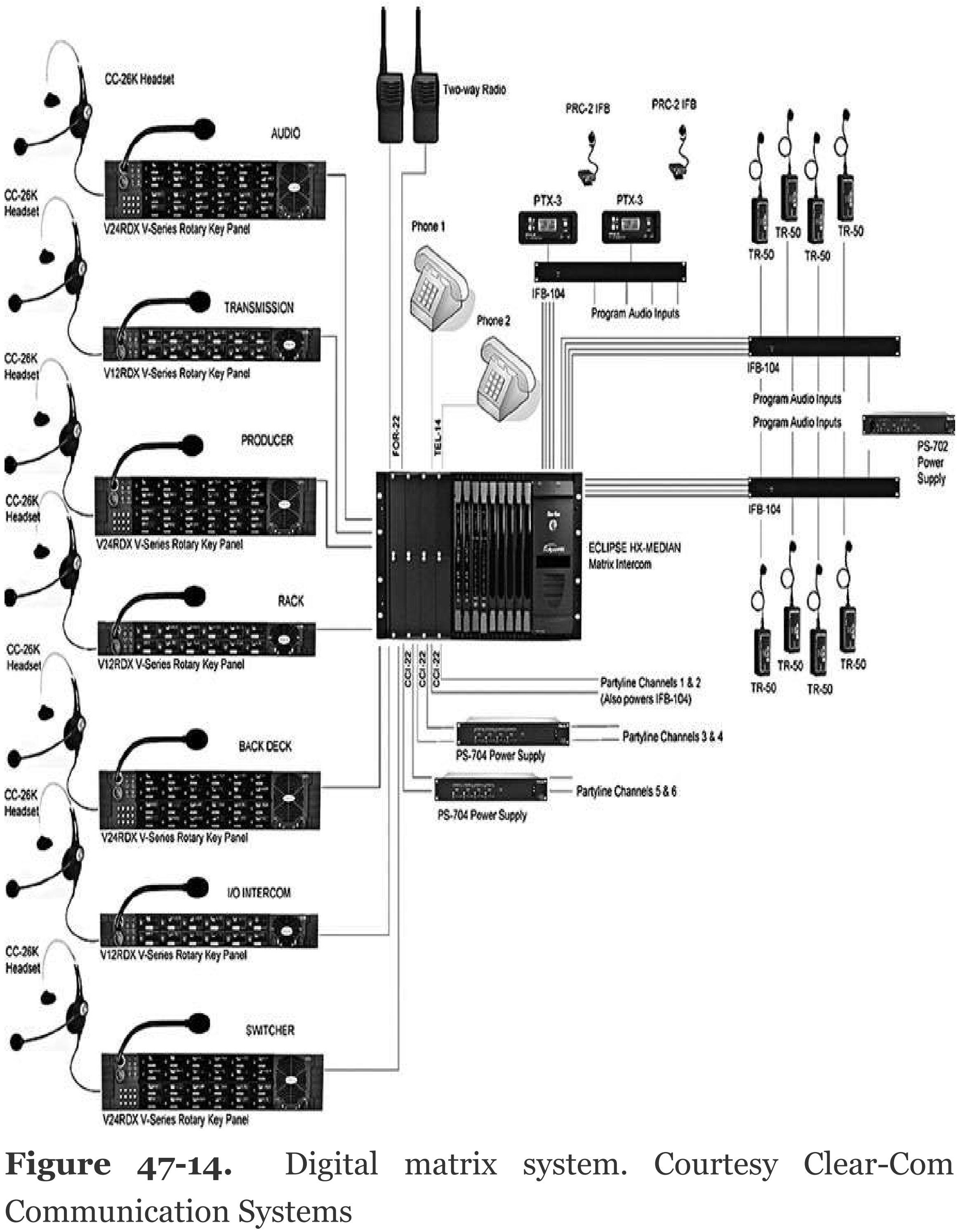


Figure 47-14. Digital matrix system. Courtesy Clear-Com Communication Systems

Interface circuitry couples the master system to external four-wire circuits providing the proper isolation, impedance matching, and level sets between systems. Additionally, it supports external relay activation and call sense circuitry. Typical circuits include 4-wire telephone lines, camera intercoms, two-way radios and microwave links, four-wire intercoms, fiber optic lines, and satellite links. Other types of four-wire circuits include IFB systems, ISO systems, and program audio in/out.

Many matrix systems have the capability of linking intelligently to other matrix systems. This allows stations in one system to communicate with stations in other systems and anything that can be selected or controlled in one system can be selected and controlled from any other linked system. This allows independent systems in remote locations, even in different cities, to operate a single system.

The station-to-matrix wiring in matrix systems can operate in either a three-pair mode or a four-pair mode. Using the four-pair scheme, remote station operation is possible from any location that can provide a standard four-wire audio circuit plus a four-wire RS-422 data circuit back to the central matrix. This can include transmission links such as satellite and fiber optic circuits, T-1 channel banks, and ISDN.

Port input signals are usually routed through software controllable *digital potentiometers*. This enables remote control of input levels from a PC, or directly from the intercom control stations, allowing instant on the fly adjustment of audio signals that are too loud or too soft within the user's mix.

Because the audio is digital, noise pickup is nonexistent, frequency response is typically 100Hz–20kHz ±1dB, distortion is less than 0.5%, and the *SNR* is greater than 68dB.

47.3.3 Program Interrupt (IFB)

IFB is a television production trade acronym that stands for *interrupted feed-back, interrupted fold-back, interrupted return-feed*, and in some cases *prompt-mute*. A normally made audio program feed is dimmed or muted when a caller talks to the IFB destination. In this way a single pair of wires carries the cue audio feed to the talent and intercom audio when needed. IFB plays an important role in the behind-the-scenes activities that make a production, as it permits the director or producer to talk to the talent during a voice-over commentary either on or off camera in a live or taped production. Sports broadcasts typically use many channels of IFB to communicate with announcers in various locations on the field and in booths. When recording live performances on stage or in a studio, musicians can hear direct queues from the conductor or producer as well as the individual music mix designated for them. For television broadcasting, IFB queues on-camera announcers and is used between on-scene reporters and the in-studio anchor and studio director. IFB is controlled by the talker or person in control while the listener has no control over IFB other than receive level.

47.3.4 Wireless Broadcast Intercom System

There are two types of wireless intercom systems. The first type provides a one-way, listen-only feature for the remote stations where the intercom is wired to the master transmitter. All communications on the intercom line are relayed to all of the wireless receivers. A one-way intercom is typically used for people who need to know what is going on, but don't need to talk back.

The second type of wireless intercom is a two-way system. In this configuration, the base unit and the field units can both talk and listen to each other in a full duplex mode. This requires two frequencies, a talk frequency and a listen frequency.

Wireless systems can stand alone, but when connected to a wired intercom system, the wireless link is virtually transparent to the user. The FCC has approved the 150–216MHz band for broadcast use with over 1700 possible frequencies available. This band is relatively free from external radio and electrical interference. Transmitter output power is limited to 50 mW. Operating distances between units vary with the environment. If it is in open areas, it is possible to have good reception up to 2000ft. However if the area includes walls, obstacles, and other radio transmitters, transmitting distances are more likely to be 150–300ft. Battery life on belt packs should be over 20h as these units are often used for extended periods. Long life is obtainable because the transmitters are only on when talking.

FM transmission has a characteristic that is called capture and is normally rated as capture ratio. It occurs when two or more transmitters are on the same frequency. The stronger of the two signals at the receiver captures the receiver, blanking out the other transmitters. If the transmitters are moving with respect to the receiver, the stronger transmitter would capture the receiver, so the communications could be bouncing between the various transmitters. For this reason all transmitters must be off when not in use and the new talker must monitor the channel before transmitting. If multiple transmitters and receivers are required, multiple frequencies are required.

47.3.5 Belt Packs

Belt packs are used in remote areas and when the person has to move around and/or the communications must not be heard by others such as for football communications between spotters and coaches, Fig. 47-15, or a theatrical performance, Fig. 47-16. They may be wired or wireless and incorporate either single or dual headsets. Belt packs often utilize noiseless, digital electronic switching on audio circuits. A push-pull amplifier supplies high levels of audio in the headset and a microphone limiter compensates for user voice variance. The belt pack may also include a two position gain switch for normal and high-noise environments. The *remote mic-kill* function at the base station enables belt pack microphones to be shut off from another location to conveniently eliminate microphone pickup. This is done by sending a 20–24kHz ultrasonic signal down the audio line, turning off the talk gate on each unit on the line. A visual call signal is provided by high-intensity LEDs on the belt pack to alert operators who have removed their headsets.

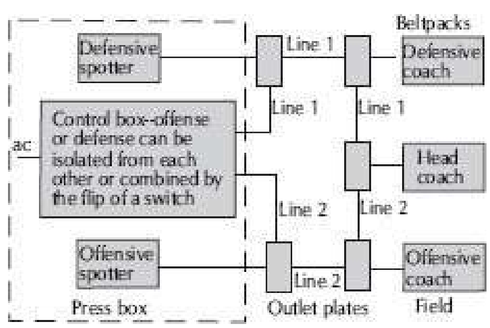


Figure 47-15. Football communication system. Courtesy Telex Communications, Inc.

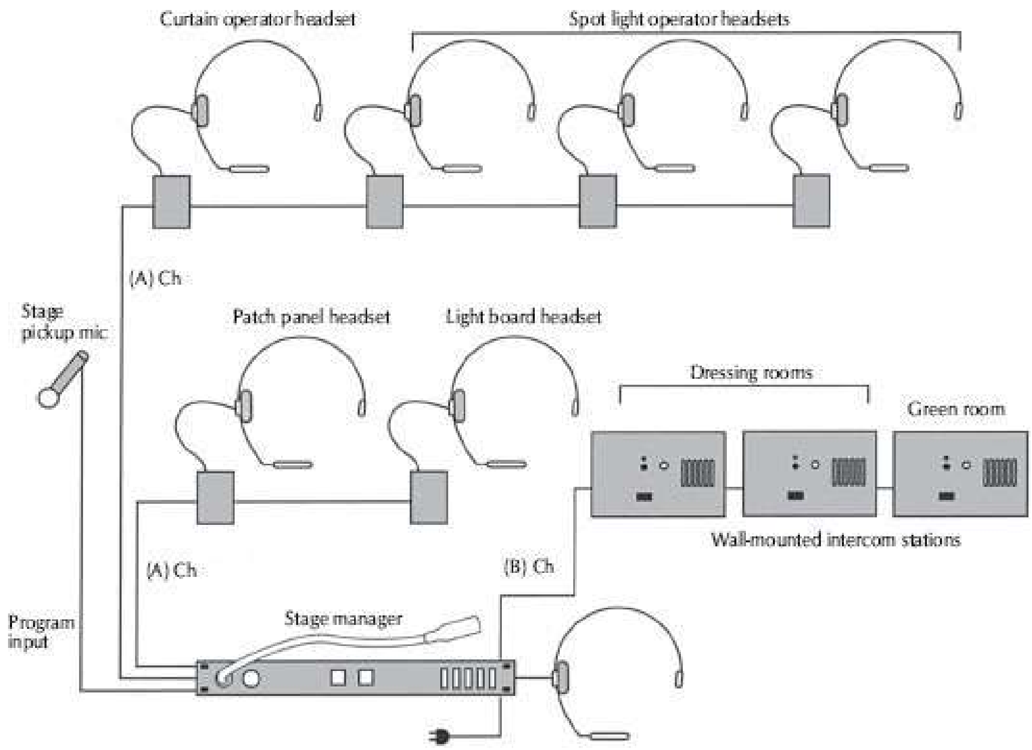


Figure 47-16. Theater intercom system. Courtesy Clear-Com Communication Systems.

47.3.6 Telephone Interface

Broadcast intercoms are often connected to a telephone line through a microprocessor-based telephone interface that provides communications between a wet dial-up phone line and an intercom system. The interface is ideal for the broadcast industry, and is specifically designed to connect a telephone line to ENG and EFP trucks, production studio consoles, and TV facilities. The device automatically answers incoming calls to the intercom system, and automatic forward-nulling circuitry adjusts internal hybrids on both sides of the line to achieve a null of up to 40dB in less than 0.1 s. The devices also include automatic gain control to insure that the incoming telephone audio remains at a constant level. It is also capable of automatic dial-up IFB, enabling a field crew to directly access a preset IFB circuit to immediately communicate with the studio crew.

When the device is set up for automatic answering, an incoming call is automatically answered, and the ringing is indicated on the master station and can illuminate a light at all intercom stations. The interface can automatically hang up or release the line if it detects a dial tone, resulting from either an intentional hang-up or a disconnection due to a line problem. If an audio program becomes too loud or distracting, it can be momentarily or permanently interrupted by any local intercom user.

Because telephone-to-intercom interfaces are used with standard telephone lines, frequency response is limited to 250Hz–3.4kHz ±3dB. Automatic volume control (AVC) is about 20dB and the depth of the null is greater than 30dB, 200Hz–8kHz.

Many of the digital matrix intercoms offer extensive direct dial-in access to the system from any touch-tone telephone in the world. They may include up to 50 two digit DTMF codes that can be used to select any station, group, program source, or IFB circuit in the system.

47.3.7 Headsets

Most headsets are designed to work with all major types of communication systems including partyline and matrix intercoms. They are used in intercom and sportscaster/announcer applications where audio quality, reliability, comfort, and the ability to hear and talk in noisy environments are of prime importance.

Light weight and comfort as well as durability are of major importance. Many headsets are made of flexible composite materials, which will not be damaged if dropped, thrown, or stepped on.

Attention must be paid to acoustical and electrical isolation between the microphone and earphone(s) to minimize the common problem of crosstalk in multiple channel intercom systems. Also by using dual chamber foam filled ear cushions, the acoustical isolation of earphones allows for comfort and low ear fatigue in high-noise environments.

Broadcast-quality microphones require noise rejecting abilities with wide-frequency response and good resistance to breath and wind noise as the microphone is usually very close to the lips. Boom micro

phones can be adjusted to any position, and located as a right or left hand headset with positive detent stops. In addition, the boom can be bent into any required position. Often swinging the boom up shuts off the microphone to eliminate feedback and unwanted noise.*

The headset cable is specially designed to minimize crosstalk between the microphone and the earphone. The wire stranding is a special composition for flexibility and resistance to breakage.

Earphones have specially contoured wide-band-width frequency response and a sensitivity of 94dB SPL with 1mW of power. This reduces amplifier power, increasing battery life in beltpacks.

* Clear-Com Eclipse can have up to 64 matrices intelligently linked with up to 256 stations per matrix.

* An example of a digital party line system is the Clear-Com HelixNet Network Party Line system.

* See Clear-Com's CC-300-X4 headset.

Chapter 48

The Fundamentals of Display Technologies

by Alan C. Brawn

48.1 Introduction

In this chapter we will explore the fundamentals of display technologies and how each technology does its work. We will begin by covering display specifications, video and computer signals, and finally the display technologies themselves providing the full context of how an image is produced on screen.

48.2 The Effect of Display Specification

In most audiovisual systems design, the display is the key focal point in the room. With this in mind, it is a requirement to match the display to the environment and the explicit needs set forth in the original sales proposal and system design. It is necessary to understand the specifications relating to display technologies, in order to properly design for individual applications. The key considerations are:

- Brightness.
- Contrast.
- Color.
- Resolution.
- Scaling.

Brightness is the element that we are most familiar with. It is the measurement of light falling on a surface as with video projectors, or light emitting from a source such as a plasma, LCD, OLED, or direct view LED flat panel display. It is most commonly stated in two units of measurement when relating to display technologies.

48.2.1 Lumens

A *lumen* is a measurement of light falling on a surface, such as a projector illuminating a screen. The measurement is taken using a light photometer pointed at the projector lens thus measuring the light output of the projector itself.

A lumen is equal to one foot-candle falling on one square foot of screen area. A standardized procedure for testing projectors was established back in 1992 by the American National Standards

Institute, defined in their IT7.215 document. The process involves averaging several measurements taken at 9 different positions, commonly called the checkerboard test pattern of alternating black and white segments. This measurement process has been re-validated in 2012 and adopted by the Society of Information Display in their Information Display Measurements Standard (IDMS). For marketing purposes, the luminous flux of projectors that have been tested according to this procedure may be still be quoted in "ANSI lumens", to distinguish them from those tested by other methods such as peak brightness. The original ANSI lumen measurement and more recent IDMS version are in general more accurate than the other measurement techniques used in the projector industry. This allows projectors to be more easily compared on the basis of their brightness specifications.

Since there are no mandatory/standardized methods of verifying lumen specifications from a given display manufacturer, the actual light output may vary as much as 20% less than the published specifications. This necessitates testing each display for actual lumen light output prior to specifying a specific projector in each application.

While lumen light output is the common specification, it is really footlamberts (fL) or the light reflected from the screen surface to the viewer that is most important. This necessitates taking the screen surface gain and ambient light in the room into consideration when specifying a projector.

Lumen light output will range from 100lm (lumens) in a tiny pico projector up to over 30,000lm in a large rental and staging or digital cinema projector.

Lumen light output will range from 10lm in a tiny pico projector up to over 40,000lm in a large rental and staging or digital cinema projector.

48.2.2 Candelas per Meter Squared (cd/m²), or Nits

Candelas per meter squared, or cd/m^2, is a unit of measure that may also be referred to as *nits* and is typically used in the measurement of light emitting from a flat panel display such as plasma, LCD, or OLED or a direct view LED display pointed directly at the viewer.

Broken down, candela, abbreviated as *cd*, is a term that originated in the days when candles were used in theaters.

For our purposes, candelas per meter squared measures the light properties radiating from a one-meter-square surface, providing a technical frame of reference for the performance of a display's black level, peak brightness, grayscale, and gamma readings.

Candelas per meter squared are more accurate than lumen light output measurements and can be measured using the same nine zone "ANSI" checkerboard test pattern. Since the screen reflectivity or gain is taken out of the equation it is less complex than a projector and screen combination. A light photometer is used to measure the light output of the display and each square on the checkerboard test pattern is measured and an average "overall brightness" is then calculated.

Typical candela per meter squared measurements or nits vary from $250cd/m^2$ for a 19 inch desktop monitor to the latest LCD displays providing $700–5000cd/m^2$ in sizes up to 108 inches diagonal and last but not least, direct view LED displays that can go up to 6,000nits for outdoor use.

It is generally agreed that contrast is the key element in a picture that provides the appearance of quality in an image. Poor contrast makes the image appear washed out whereas while good contrast gives us excellent depth of field and much more detail in a picture in the form of legible discrete information that we can see. It also gives us the appearance or perception of higher resolution while not actually providing more lines of resolution or more pixel density in the image.

48.2.3 Contrast

Contrast is the full range of light and dark values in a picture.

- It is stated as the ratio between maximum and the minimum brightness values, e.g., 500:1.
- Low contrast is shown mainly in shades of gray.
- High contrast is shown as blacks and whites with very little gray.
- In digital technologies, contrast is the difference between the luminance or brightness of an ON pixel and an OFF pixel.

Contrast is used as a marketing specification and is the most misstated of all specifications from manufacturers. The proper method to measure contrast is by using a sixteen zone alternating black and white “ANSI” checkerboard test pattern and comparing the average of the dark rectangles to the white rectangles to get the proper contrast ratio. A light photometer is used to measure the light output of each square on the checkerboard test pattern and an average picture contrast ratio is then calculated. As a point of reference, when measured in this manner, the most expensive digital cinema projectors in the world will produce approximately 500:1 contrast and a typical boardroom projector will produce less that 100:1 contrast in a typical lighting condition.

In 2010 InfoComm and their ANSI Standards development organization published a system contrast standard for projection. It is called the Projected Image System Contrast Ratio or PISCR for short. It established four viewing criteria made up of the following:

1. Passive Viewing
   a. The viewer is able to recognize what the images are on a screen and can separate the text or the main image from the background under typical lighting for the viewing environment. The content does not require assimilation and retention of detail, but the general intent is understood. There is passive engagement with the content (e.g., non-critical or informal viewing of video and data).
   b. Passive Viewing requires a minimum contrast ratio of 7:1

2. Basic Decision Making
   a. The viewer can make basic decisions from the displayed image. The decisions are not dependent on critical details within the image, but there is assimilation and retention of information. The viewer is actively engaged with the content (e.g., information displays, presentations containing detailed images, classrooms, boardrooms multi-purpose rooms, product illustrations).
   b. Basic Decision Making requires a minimum contrast ratio of 15:1.

3. Analytical Decision Making
   a. The viewer can make critical decisions by the ability to analyze details within the displayed image. The viewer is analytical and fully engaged with these details of the content

(e.g., architectural/engineering drawings, forensic evidence, medical imaging, and photographic image inspection).

b. Analytical Decision Making requires a minimum contrast ratio of 50:1.

4. Full Motion Video

a. The viewer is able to discern key elements present in the full motion video, including detail provided by the cinematographer or videographer necessary to support the story line and intent (e.g., home theater, business screening room, broadcast post-production).

b. Full Motion Video requires a minimum contrast ratio of 80:1.

Although the PISCR is the ANSI standard for viewing projected images in systems involving a projector, screen, and varying amounts of ambient light, it gives us a guide to what the human eye sees and relates to viewing flat panel displays as well.

48.3 Display Color

Each display technology produces the color we see in a different manner but they all utilize the primary colors of red, green, and blue as well as the secondary colors of cyan, magenta, and yellow to create the full color spectrum we see onscreen.

Going back to our physics class in high school, we remember that white light, when viewed through a prism, produces a veritable rainbow of colors, known as the full color or electromagnetic spectrum. The easiest way to remember the visible color spectrum is the name ROY G BIV (red, orange, yellow, green, blue, indigo, and violet).

Infrared and ultraviolet light are not visible, and fall at the far extremes of the spectrum.

In the world of professional audiovisual, you may encounter what is known as the CIE chromaticity diagram (color chart). This chart illustrates the full color spectrum, including wavelengths of light measured in nanometers and color temperature, measured in degrees Kelvin. As a specific point of reference, you can see the overlap of red, green, and blue, producing white light.

In modern display technologies, color is created in a variety of ways, by the manipulation of white light:

- Color is produced from the light of a projector lamp shining through a color wheel, in the case of single chip DLP (digital light processing).
- Color is produced from the light of a projector lamp or solid state illumination devices like an LED or laser phosphor array shining through a transmissive display device, or bouncing off a reflective display device, such as LCD, three chip DLP, or LCOS (liquid crystal on silicon).
- Color is produced from an emissive (light producing) display device, such as a CRT, plasma display, LED, or OLED.
- Color is produced from the backlight of a transmissive display device, such as a flat panel LCD.

Color and color space can be calibrated in the majority of displays by using instruments known as colorimeters. These devices measure the visible color spectrum and permit the technician to calibrate for a specific color temperature depending on the application. Most displays are set for 6500K (Kelvin), also known as D65, which replicates an image in a full daylight mode.

The concept and process of calibration is to match the output of the display image to the viewer with the content replicated as the graphic artist or videographer intended. For example, a brand logo must match their specific color perfectly, and the grass at the golf course must look like real grass and not a color never to be found in nature.

48.4 Display Resolution

The resolution of digital display technologies is fixed as in the reference to fixed matrix displays. Resolution relates directly to visual acuity and what the eye can see. Each display technology differs in the spaces between the pixels and this is called the fill factor. The displays with the highest fill factor appear to have less of what is known as the screen door effect thereby providing a look closer to that of an analog image of 35mm color film or CRT displays. The higher the number of pixels, the more detail in an image.

- In digital displays, resolution is the number of pixels (picture elements or individual points of color) contained on a display, expressed in terms of the number of pixels on the horizontal axis and the number of pixels on the vertical axis—e.g., 1920 × 1080.
- The sharpness of the image on a display depends on the resolution and the size of the display. The same pixel resolution will be sharper on a smaller display and gradually lose sharpness on larger display because the same number of pixels is being spread out over a larger number of inches.
- In terms of fill factor, LCD has the most space between pixels, with DLP providing more fill factor, and LCoS providing the

highest fill factor available today.

- The following are common display resolutions that are encountered in commercial audiovisual today:
  - VGA is 640 × 480
  - XGA is 1024 × 768
  - HD-720 (720p) is 1280 × 720
  - SXGA is 1280 × 1024
  - WXGA is 1280 × 800
  - UXGA is 1600 × 1200
  - HD-1080 (1080p, 1080i) is 1920 × 1080
  - WUXGA is 1920 × 1200
  - 4K UHDTV (2160p) is 3840 × 2160
  - 8K UHDTV (4320p) is 7680 × 4320

48.5 Display Scanning

Television signals and compatible displays are typically interlaced, and computer signals and compatible displays are typically progressive (noninterlaced). These two formats are incompatible with each other; one would need to be converted to the other before any common processing could be done.

Designed for analog NTSC television, interlaced scanning is where each picture, referred to as a frame, is made up of two separate subpictures, referred to as fields, so two fields make up a frame. An interlaced picture is drawn on the screen in two passes, by first scanning the horizontal lines of the first field and then retracing to the top of the screen and scanning the horizontal lines for the second field in between the first set. Field 1 consists of lines 1 through 262½, and field 2 consists of lines 262½ through 525. A television scans 60 fields every second (thirty odd and thirty even).

These two sets of thirty fields are combined to create a full frame every 1/30th of a second, resulting in a display of thirty frames per second. Drawbacks to interlaced scanning compared to progressive scanning include lower resolution, flicker, aliasing, and image artifact quality issues.

Progressive scan differs from interlaced scan in that each line (or row of pixels) in the signal is drawn in a sequential order rather than an alternate order, as is done with interlaced scan. In short, with progressive scan, the image lines (or pixel rows) are scanned in numerical order (1, 2, 3) down the screen from top to bottom, instead of in an alternate order as done in interlaced scanning. By progressively scanning the image onto a screen every 60th of a second rather than "interlacing" alternate lines every 30th of a second, a smoother, more detailed, image can be produced on the screen. The benefit is the viewing fine details, such as text, and is also less susceptible to interlace flicker and basically eliminates aliasing on the edges of objects in a picture. The drawback to progressive scan is that it requires more bandwidth to display the images onscreen.

48.6 Aspect Ratios and Screen Formats

Aspect ratio refers to the shape of the images we see on screen, but just what comprises an aspect ratio? Aspect ratio is typically described as the ratio of screen width to screen height. There are two common aspect ratios. The first is that of a standard television, which has a 4:3 (referred to as 4 by 3) aspect ratio. Also note that the television aspect ratio is listed as 1.33:1. This is another way of listing aspect ratios—dividing the width by the height, e.g., 4/3 = 1.33. This is referred to as 1.33:1 or 1.33 to 1. A widescreen display, such as a plasma panel, will usually have a 16 by 9 aspect ratio (16:9). Since 16/9 = 1.78, the aspect ratio is also known as 1.78:1 or 1.78 to 1. Common aspect ratios include:

- **4×3 (1.33:1).** This is the standard television format used throughout the second half of the 20th century. This is both typical computer and NTSC broadcast video.

  Note: 1280 × 1024 is actually 5:4 aspect, not 4:3.
- **16 × 9 (1.78:1).** This is the common format for widescreen DVD movies, HDTV (720p and 1080i) and widescreen computer resolutions (1280 × 720, 1920 × 1080, etc.).
- **13 × 7 (1.85:1).** This is the standard aspect ratio for theatrical release film prints.
- **29 × 9 (Cinemascope–2.35:1)**. A very wide screen format used for theatrical release movies, and some new DVDs.

48.7 Scaling

In the realm of digital display technologies, there is quite often a mismatch between the resolutions of the display itself and the signals or sources coming into the display. This mismatch necessitates the incorporation of a process known as scaling or scan conversion or also known as up scaling or down scaling. This refers to a process of taking a higher resolution signal, and modifying it to be displayed on a lower resolution device, or a lower resolution signal, and modifying or up scaling it to be displayed on a higher-resolution device.

While a scaler can be an outboard device, in most instances today, it is built into the display device.

Digital Fixed Matrix Display Scan Conversion

In a digital or fixed matrix display, where the pixels are in a fixed size and position, the input signal may or may not match. Scan conversion is a necessary process to fit the analog source image across multiple pixels in the display. Depending on the display technology, 20–30% of image information may be lost.

For example, if we have a fixed matrix display at 1024 × 768 and want to input a signal that is 800 × 600, a mathematical algorithm is employed where the signal information in the lower-resolution signal is reduced to a mathematical equation and fit into the higher-resolution display. The same process takes place where a higher-resolution image is fit mathematically into a lower-resolution display.

In all instances where scaling and scan conversion take place there is lost information. The quality of the scaler or scan converter varies with each display and the most fidelity in an image takes place where the signal and the display resolutions match each other.

48.8 Video Signals

In order to further the understanding of displays and display technologies, it is necessary to gain a basic comprehension of what comprises the different types of video signals in use today. We will now examine the core components of all signals, and their transmission standards.

48.8.1 What Comprises a Video Signal?

Chrominance noted as (C) is the hue or color with saturation in the red, green, and blue channels of a signal.

Luminance noted as (Y) is the amount of light in each red, green, and blue channel.

Without the chrominance in a signal, the picture is black and white.

48.8.2 Composite (aka NTSC)

An analog composite video signal is used in most home applications. It combines the chrominance and luminance, along with a sync signal into one cable. This facilitates the broadcast of the NTSC television signal to our homes.

48.8.3 Y/C (aka S-Video)

This is still a composite signal, but one that nearly separates luminance and chrominance to provide a more precise color reproduction on the screen.

48.8.4 Component Video

Commonly known as RGB (RGB sync, RGB with H and V sync, RGB sync on green).

This type of signal totally separates red, green, blue, and sync to give clearer definition to each of the color channels.

Component is never used for broadcast due to its excessive bandwidth requirement in the green channel. Note that the sync signal can be H/V (horizontal and vertical) or sync on green.

One version of a component signal is commonly known as YPbPr. In the broadcast community, this is known as a color difference signal. Since a normal RGB signal requires too much bandwidth to broadcast the dominant green channel, YPbPr, as a component signal, extrapolates the green signal by subtracting from the luminance channel (Y) both the blue component (Pb) and the red component (Pr), leaving the green component.

This allows the economical broadcast of a component signal by reducing the bandwidth needed by eliminating the dedicated green signal.

48.8.5 VGA (Video Graphics Array)

VGA is the analog display standard for the PC. VGA uses an analog monitor, and PC display adapters to output analog signals. All PC CRTs and most flat panel monitors accept VGA signals, although newer flat panels may also have a DVI interface for display adapters that output digital signals.

VGA may refer to the physical 15-pin VGA socket on a PC in order to contrast it with a digital DVI socket for flat panels. Or, VGA may also refer only to the original VGA resolution of 640 × 480 and 16 colors.

48.8.6 DVI (Digital Video Interface)

DVI is a multipin connection used for passing standard-definition and high-definition digital video signals, found on HDTV tuners, a growing number of DVD players, HDTV-ready televisions, and some computer displays. DVI connections transfer video signals in pure digital form, which is especially beneficial if you're using a fixed-pixel display (like a LCoS, plasma, LCD, or DLP TV). Signals are encrypted with HDCP high-bandwidth digital content protection) to prevent content from being re-recorded and pirated.

There are different kinds of DVI connections. DVI-D, which is the type of DVI connection found on most home video gear, carries digital-only signals. DVI-I, used with some computer video cards, is capable of passing both digital and analog video signals. Some TVs feature DVI-I inputs for greater hookup flexibility.

48.8.7 HDMI (High-Definition Multimedia Interface)

HDMI is the second generation digital interface that evolved out of the DVI standard.

HDMI is a multipin connection used for passing standard- and high-definition digital video signals, as well as multichannel digital audio, through a single cable. These connections are usually found on newer HDTV tuners, and a growing number of DVD players, HDTV-ready televisions, and home theater receivers. HDMI cable accommodates up to 5Gbps bandwidth, so it can simultaneously transfer pure digital video and audio signals without compression (even HDTV video).

HDMI works especially well with a fixed-pixel display (like a LCoS, plasma, LCD, or DLP TV), and is backwards compatible with most DVI connections. Signals are encrypted with HDCP (high-bandwidth digital content protection) to prevent recording.

Although many first generation HDMI-equipped components only pass two-channel audio signals, HDMI can carry up to eight discrete audio channels, making it forward-compatible with 7.1 sound systems. That means you can pass digital video and multichannel audio signals between newer HDMI-equipped components along a single cable.

48.9 Digital Display Technologies

In the early days of the audiovisual industry, it was necessary to immerse oneself in the tiniest details of technology and how it operated. In today's market, it is necessary to understand the basic function of various technologies, and more specifically, how the basic functions affect the final design, and the solutions presented to the client and for the specific project.

We will examine the characteristics and basic functions and operation of the following:

- PDP (plasma).
- DLP (digital light processing).
- LCD (liquid crystal display).
- LCOS (liquid crystal on silicon).
- OLED (organic light emitting diode).
- LED (light emitting diode).

48.9.1 Plasma Display Technology

Of all fixed matrix display technologies, plasma, or PDP displays most closely replicate the smooth image from a 35mm film projector and a CRT. Plasma displays are emissive in nature, and utilize a similar rare earth phosphor to a CRT to provide color saturation for the display, Fig. 48-1.

48.9.1.1 PDP Characteristics

- 3 to 4 inch thick displays (wall or base mount).
- 60 to 500 pounds.
- Panel sizes 37 inch, 40 inch, 42 inch, 43 inch, 46 inch, 50 inch, 55 inch, 60 inch, 61 inch, 63 inch, 71 inch, 103 inch, and 150 inch.

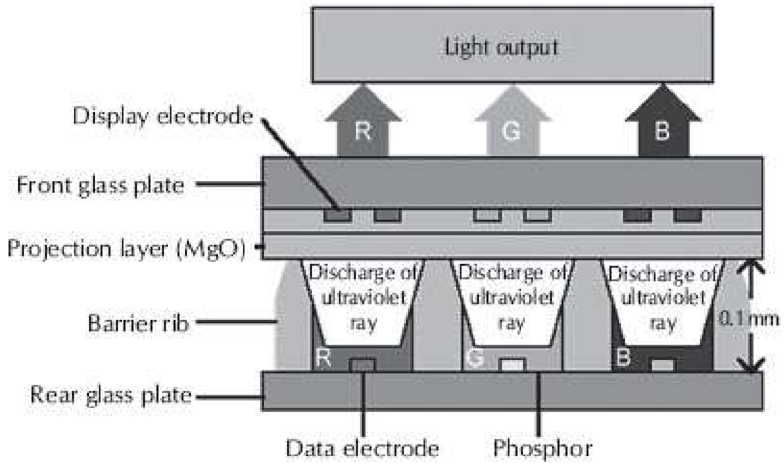


Figure 48-1. Plasma monitor.

- 16:9 aspect ratio panels.
- PDP combines the pixel structure of LCD with the color generation of a CRT.
- No radiation or high voltage emissions.
- Fast response time.
- High contrast.
- Deep color saturation.

48.9.1.2 How PDP Operates

- The cells are filled with a xenon and neon gas mixture.
- A controlled current is passed through the gas.
- Ultraviolet rays are produced by the current energizing the gas, creating a plasma.
- Ultraviolet rays hit the red, green, and blue phosphors applied inside the cells.
- Visible light is produced by the ultraviolet rays exciting the rare earth phosphors.
- Voltage is applied to one of three terminals on a pixel. The voltage discharges through the pixel to a second electrode ionizing a rare

gas (creating a plasma) in the process. The ionization creates UV light, which excites an RGB phosphor causing it to glow (like a CRT). Brightness variation is achieved by controlling the number of pulses of light that our eyes integrate to produce impression of dim or bright areas.

48.9.2 Liquid Crystal Displays

Liquid crystal displays have become ubiquitous. As the foundation for modern computer and cell phone displays, LCD technology is used for large flat panel displays as well as three chip LCD projectors. No matter the application, LCD technology and how it works is similar in the way it fundamentally operates.

48.9.2.1 LCD Characteristics

- 3 to 4 inch thick displays (wall or base mount).
- 60 to 400 pounds.
- Panel sizes ranging from a few inches diagonal up to 110 inch panels.
- 4:3, 16:9, and 16:10 aspect ratio panels.
- No radiation or high-voltage emissions.
- Low power consumption.
- High resolution, up to 4 × HDTV.
- Ideal for computer display and digital signage.

48.9.2.2 How LCD Operates

There’s far more to building an LCD than simply creating a sheet of liquid crystals. The combination of four facts makes LCDs possible:

- Light can be polarized.

- Liquid crystal can transmit polarized light or change the plane of polarization.
- The structure of liquid crystals can be changed by electric field.
- There are transparent substances that can conduct electricity.

To create an LCD, you take two pieces of glass with polarizing films applied.

A polyimide film is applied to the liquid crystal side of the glass and then mechanically rubbed to produce microgrooves.

The two glass plates are assembled together with a carefully controlled gap dimension.

When LC material is introduced to this cell, the layers adjacent to the polyimide will align with the microgroove directions resulting in a helical structure of LC molecules between the two glass plates, Fig. 48-1.

Liquid crystal displays come in two basic configurations: flat panel displays and projection displays. Both variations utilize the same basic LCD principle, but differ in the way that they are illuminated.

In the flat panel, or desktop display, the illumination comes from bright cold cathode fluorescent lights behind the display.

In projection LCD displays, the illumination comes from a bright lamp reflecting off of the LCD and onto the screen.

LCD monitors make use of thin film transistors (TFT). TFTs are small switch transistors and capacitors that sit on a glass substrate in the LCD structure, Fig. 48-2.

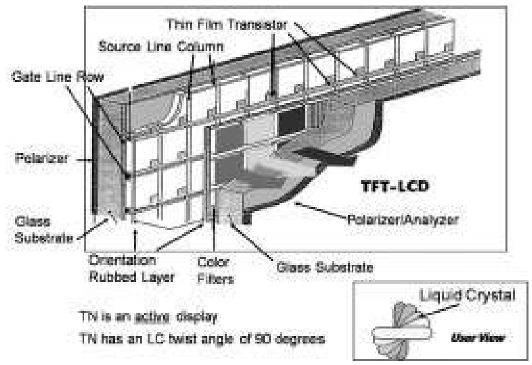


Figure 48-2. TFT-LCD technology.

Each pixel is controlled by one up to four of these TFTs. To ignite a particular pixel, power is applied to the correct column and row (just like passive matrix).

Any pixels on the same row and column that are not targeted simply pass the current on. The transistor at the target pixel stops the current. The capacitor takes the current and stores it. It is then able to hold that charge until the next screen refresh.

Also, by adjusting the amount of voltage to each pixel, you can control the amount that the crystals will untwist, thereby allowing varying degrees of color.

For an LCD monitor to produce color, each pixel on the screen has to have three subpixels, each being a primary color (red, blue, and green). In this aspect, color LCDs work the same way as the color CRT. By taking each of the three colors, each having 256 possible shades, and blending it all together, the color active matrix LCD has a possible palette of 16.8 million colors. Each subpixel has a transistor/capacitor and with this design process, one can see that there are millions of transistors necessary to formulate a full TFT screen.

In an LCD monitor, the light source is behind the panel and illuminates the display from behind. Typically the lighting is a florescent type but the most recent development in illuminations is via side emitting LED display, which improve uniformity, durability, brightness, and the life of the backlight.

LCD projectors utilize three LCD panels or chips as the imaging devices but unlike LCD monitors, they differ in the way color and illumination are derived. By looking at the LCD light path illustration, Fig. 48-3, we begin with a metal halide lamp for illumination. The lamp approximates pure white light from which the colors of the spectrum can be extrapolated. Color (RGB) is achieved by incorporating dichroic mirrors or filters into the light path. The dichroic mirrors filter out all of the unwanted color spectrum and pass on a narrow band of color coordinates of red, green, and blue, permitting each of those colors in a pure form to be transferred to the main optical prism or combiner just behind the projection lens.

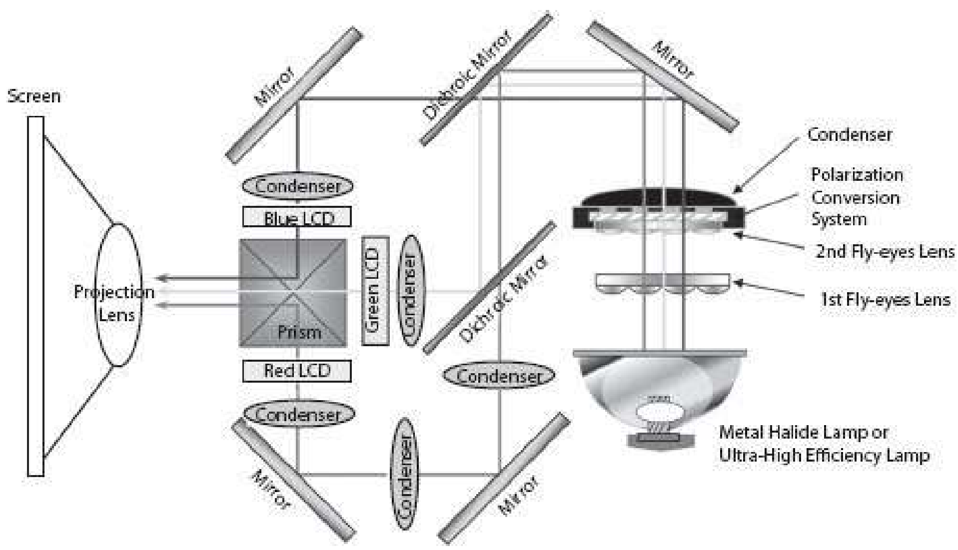


Figure 48-3. LCD projection TV optical path.

LCD projectors come in various sizes, shapes, light outputs, and resolutions.

Typical native resolutions for commercial LCD projectors are 640 × 480, 1024 × 768, 1280 × 800, 1280 × 1024 and 1920 × 1080.

In terms of weight, the brighter the projector, the bigger the lamp housing requirement and hence the heavier the projector.

Modern LCD projectors vary in size from 5lbs to over 50lbs for the high-brightness models.

Brightness has long been the holy grail of projectors and LCD with flat panel brightness from 250 to 5000cd/m^2 and projector brightness achieving 16,000lm.

48.9.3 Digital Light Processing

Digital light processing was developed by Dr. Larry Hornbeck of Texas Instruments and brought to market in the mid-1990s. It is fundamentally a digital light switch that is used in projection applications as far reaching as tiny pico projectors to be inserted in cell phones all the way to digital cinema projectors replacing 35mm film in movie theaters. Its compact size along with single chip and three chip variations make it unique in the world of display technologies, Fig. 48-4.

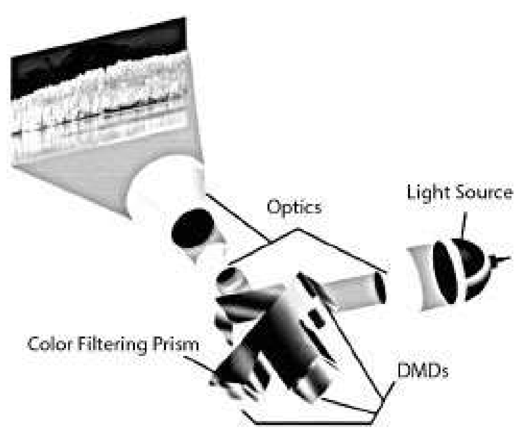


Figure 48-4. Three chip DLP projection TV.

48.9.3.1 DLP Characteristics

- Projection technology, no fixed screen size.
- Single or three chip configurations.
- 16:9 and 16:10 aspect ratio panels.
- No radiation or high-voltage emissions.
- Low power consumption.
- High resolution, up to 4 K.
- High brightness and contrast.
- Does not require polarized light.

48.9.3.2 How DLP Operates

- DLP™ is based on an optical semiconductor called a digital micromirror device, or DMD.
- The DMD is an extremely precise light switch that enables light to be modulated digitally via millions of microscopic mirrors arranged in a rectangular array.
- Each mirror is spaced less than 1 micron apart.
- These mirrors are literally capable of switching on and off thousands of times per second and are used to direct light toward, and away from, a dedicated pixel space.
- When the display is off, all of the mirrors are flat.
- When the display is turned on and the chip begins transmitting the signal, the mirrors flip back and forth thousands of times per second.
- Mirrors in the on position reflect the light through a projection lens and onto the screen. The longer a mirror is in the on position, the lighter the pixel it creates. Mirrors that are off for longer periods create darker pixels, and mirrors that are always off

create black pixels. By varying the length of time that the mirrors point toward the projection lens, the DMD creates up to 1024 shades of gray.

- The gray pixels combine on the screen to create a progressive, fully digital monochrome image.
- To add color to the picture, the single chip DLP system uses a color wheel, Fig. 48-5.

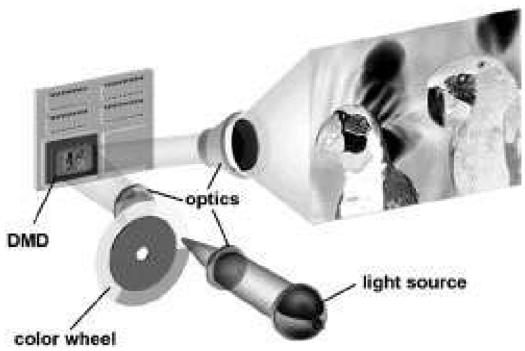


Figure 48-5. Single chip DLP technology.

- The color wheel is a transparent, spinning wheel with red, green, and blue. The light passing through each section turns red, green, or blue.
- The system's processor synchronizes the spinning of the wheel with the action of the mirrors. Together, the DMD and the color wheel can create 256 shades of each primary color.
- Each pixel of light on the screen is red, green, or blue at any given moment. The colors are then blended to create the desired colors of the image.
- With DLP projectors, a small number of people might experience a rainbow effect when watching a DLP projection, especially when they change their focus from one part of the image to another, seeing the individual component colors.
- This happens only in DLP systems that use a segmented color

wheel, not in systems that use one DLP chip for each primary color.

- A number of home theater systems use color wheels with additional segments, two segments of each color, or sequential color recapture (primary colors arranged in a spiral instead of in segments) in order to reduce the appearance of the rainbow effect.
- New BrilliantColor™ color wheel technology reduces the appearance of the rainbow effect. In the newer DMD generations, a light-eating dark metal coat is applied to the interior of each chip, preventing stray light from traveling to the screen when mirrors are switched off. This improvement increases contrast ratios from 1200:1 to >2000:1 and higher.
- Increased mirror tilt angle (from ±10 to ±12°), brings 20% more light to the screen for greater brightness.
- Double data rate technology allows a DMD chip to tilt toward or away from its light source twice as fast, allowing more accurate grayscale reproduction.

48.9.3.3 New DLP BrilliantColor™ Color Wheel Technology

Historically, most display devices would render a scene using the three primary colors, red, green, and blue.

This limits available colors that can be displayed, making it difficult to display brilliant yellows, magentas, and cyans that are commonly found in natural scenes.

BrilliantColor™ technology adds yellow, cyan, and magenta colors to the color wheel, maintaining bright whites while providing deeper red, green, and blue colors.

BrilliantColor™ provides brightness increases in nonprimary colors and boosts overall color intensity.

BrilliantColor™ provides flexibility in color wheel design allowing for bright, large color gamuts and differentiation from OEMs.

48.9.4 Liquid Crystal on Silicon

Liquid crystal on silicon combines the best of both worlds of LCD and DLP. It is a reflective technology like DLP but uses liquid crystals instead of moving mirrors to control the light transmission levels of the individual pixels. The benefit of LCoS is that it has excellent color and contrast capabilities and has the highest fill factor of any current digital display. It is also capable of 4K resolution and is a competitor with DLP for digital cinema applications, Fig. 48-6.

48.9.4.1 LCoS Characteristics

- Projection technology, no fixed screen size.
- Three chip configuration.
- 16:1 aspect ratio panels.
- No radiation or high voltage emissions.
- Low power consumption.
- High resolution, up to 8 K.
- High brightness and contrast.
- High fill factor.

48.9.4.2 How LCoS Operates

- LCoS technology is a reflective liquid crystal modulator where electronic signals are directly addressed to the device.
- The LCoS device has an X-Y matrix of pixels configured on a

CMOS single crystal silicon substrate mounted behind the liquid crystal layer using a planar process that is standard in IC technology.

- The liquid crystal is placed on top of the CMOS substrate on an array of aluminum mirrors that define each pixel.
- A glass counter electrode covers the liquid crystal to complete the structure.
- A voltage is applied to a selected pixel of the matrix in accordance with the input signal, making the liquid crystal change birefringence, thus changing the polarization direction of the incident projection light.
- The nonactive area between the pixel mirrors is minimal, only serving to separate each pixel; the rest of the electrode is active as a reflective surface, thereby providing a high aperture fill ratio.
- Although having the highest overall performance of any current projector technology it is held back by manufacturing yield issues and cost of components that impede its progress.

48.9.5 Organic Light Emitting Diode

Organic light emitting diode (OLED) is the newest display technology and a direct competitor for other flat panel displays such as LCD and plasma. The most obvious benefit is the nearly paper thinness of the technology. Since it is an emissive technology that does not require separate lighting it can be manufactured to create a display the thickness of a credit card. It can be made transparent and even flexible. It also has advantages in the area of low power consumption and excellent picture performance dynamics. The big issues facing OLED are manufacturing costs, and panel life, both of which are in the process of being addressed.

48.9.5.1 OLED Characteristics

- Thinnest and lightest display technology.
- Fast response time.
- High brightness.
- Low power consumption.
- Can be made transparent or flexible.

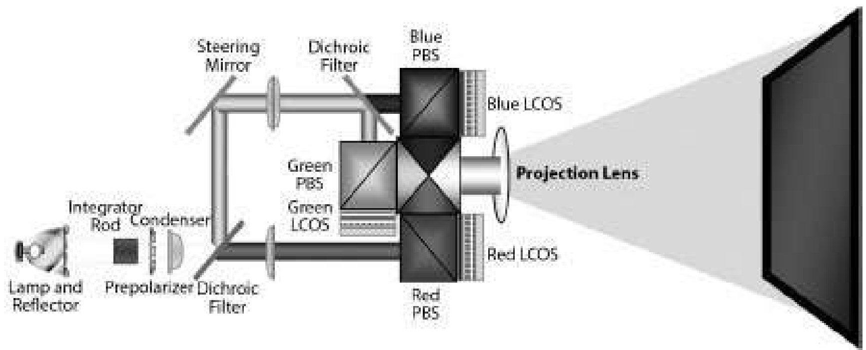


Figure 48-6. LCoS projection system.

48.9.5.2 How OLED Works

- The basic OLED cell structure consists of a stack of thin organic layers sandwiched between a transparent anode and a metallic cathode.
- The organic layers comprise a hole-injection layer, a hole-transport layer, an emissive layer, and an electron-transport layer.
- When an appropriate voltage (typically a few volts) is applied to the cell, the injected positive and negative charges recombine in the emissive layer to produce light (electroluminescence).
- The structure of the organic layers and the choice of anode and cathode are designed to maximize the recombination process in the emissive layer, thus maximizing the light output from the

OLED device.

OLEDs are typically fabricated on a transparent substrate on which the first electrode (usually indium-tin-oxide which is both transparent and conductive) is first deposited.

Then one or more organic layers are coated by either thermal evaporation in the case of small organic dye molecules, or spin coating of polymers. In addition to the luminescent material itself, other organic layers may be used to enhance injection and transport of electrons and/or holes.

The total thickness of the organic layers is of order 100nm.

Lastly, the metal cathode (such as magnesium-silver alloy, lithium-aluminum, or calcium) is evaporated on top.

The two electrodes add perhaps 200nm more to the total thickness of the device. Therefore the overall thickness (and weight) of the structure is mostly due to the substrate itself.

OLEDs can be manufactured in several different types, classified by the size of molecule they use, and the type of substrate they are manufactured on. Some examples are:

- TOLED—Transparent OLED. This is manufactured on a clear substrate suitable for applications such as heads-up displays.
- FOLED—Flexible OLED. This type of OLED is manufactured into a sealed flexible substrate that can be curved, rolled, or bent.

48.9.6 Light Emitting Diode

Light emitting diodes are popping up everywhere due to their high light output and relatively low power consumption. They are finding uses in homes, automobiles, and of course high brightness outdoor displays. Their newest application is as backlight illumination for LCD flat panel displays and as a light source for small pico projectors utilizing DLP and LCoS chips.

48.9.6.1 LED Characteristics

- Extremely high brightness.
- Relatively low maintenance.
- Long life, >50,000 hours.
- Outdoor/indoor capability.
- Modular construction with scalable display sizes.
- 1.9 to 25mm pixel pitches available.

48.9.6.2 How LED Operates

The phenomenon of *electroluminescence* was discovered in 1907 by Henry Joseph Round.

British experiments in the 1950s led to the first modern red LED, which appeared in the early 1960s.

By the mid-1970s LEDs could produce a pale green light. LEDs using dual chips (one in red and one in green) were able to emit yellow light.

The early 1980s brought the first generation of super bright LEDs, first in red, then yellow, and finally green, with orange-red, orange, yellow, and green appearing in the 1990s.

The first significant blue LEDs also appeared at the start of the 1990s, and high-intensity blue and green in the mid-1990s.

The ultra bright blue chips became the basis of white LEDs, in which the light emitting chip is coated with fluorescent phosphors.

This same technique has been used to produce virtually any color of visible light and today there are LEDs on the market, which can produce previously exotic colors, such as aqua and pink.

Light emitting diodes (LEDs) are source of continuous light with a high efficiency.

- At the heart of a light emitting diode is a semiconductor chip, containing several very thin layers of material that are sequentially deposited onto a supporting substrate.
- The first semiconductor material that is deposited onto the substrate is doped with atoms containing excess electrons, and a second doped material, containing atoms having too few electrons, is then deposited onto the first semiconductor to form the diode. The region created between the doped semiconductor materials is known as the active layer.
- When a voltage is applied to the diode, holes (positive charges) and electrons (negative charges) meet in the active layer to produce light.
- The wavelength of light emitted by the diode is dependent on the chemical composition and relative energy levels of the doped semiconductor materials, and can be varied to produce a wide range of wavelengths.
- After being fabricated, the chip is mounted in a reflector cup connected to a lead frame, and is bonded with wire to the anode and cathode terminals.
- The entire assembly is then encased in a solid epoxy dome lens that enables emitted light to be focused, controlled by embedding tiny glass particles into the lens that scatter light and spread the light beam, or angled, via changing the shape of the lens, or the reflector cup.

48.10 Resolution

What is resolution? A simple definition of resolution is the degree of sharpness and clarity of a displayed image. In LED displays, resolution is determined by the matrix area and pitch.

The area, also known as the pixel matrix, corresponds to the number of pixels that make up the display area. In our industry, we express the matrix area in the number of pixels vertically by number of pixels horizontally, such as 16 × 64.

The pitch is defined as the distance between two pixels. The distance is measured from the center of one pixel to the center of the next pixel.

Pitch can also influence the pixel matrix for a given area. For example, a 16mm pitch will give you a 5 × 7 matrix area, while a 10mm pitch gives you a 8 × 11 matrix area for the same area.

Pitch determines the amount of empty space between the pixels. Therefore, the smaller the pitch and larger the matrix area, the greater the resolution.

48.11 Burn-In and Image Retention

One consideration when choosing which display technology is appropriate for a given application is how the type of usage and content will affect the display in terms of burn-in or image retention. Both LCD and plasma displays can suffer from this phenomenon.

48.11.1 Plasma Burn-In

Plasma displays use a rare earth phosphor, which is consumed over time as the UV light generated by the gas plasma excites the phosphor. This is the natural aging process of the display. When static images are left on the display for extended periods of time, this can result in uneven phosphor use, or degradation. This causes a luminance difference between the colors of each sub-pixel, or across an entire pixel. This luminance difference will appear as a darker or color shifted pixel, resulting in a "ghost" image of the static content shown on screen. Frequently moving content, or inverting of colors, as well as juxtaposing full motion content with partially static content can reduce this phenomenon, but it cannot be completely eliminated.

48.11.2 LCD Image Retention

Every display based on LCD technology, when run for an extended period of time with a fixed image on screen, can suffer from a phenomenon known as image retention. The LCD display operates by twisting the liquid crystal structure to block or transmit light through the panel. When a static image is left on the screen for prolonged periods of time, the twist in the liquid crystal is kept fixed in a single position. When the liquid crystal molecules are kept still partially twisted for long periods of time, they can "stick" in that position, due to storing extra energy from the voltage forcing them to twist. This may leave a ghost image trapped on the screen. Over time this effect may trail off as the energy trapped in the liquid crystal drains away, however it can happen repeatedly.

48.11.3 LCD Response Time

Response time is the amount of time, measured in milliseconds (ms) a pixel in an LCD monitor takes to go from one brightness level to another and back again. Lower numbers mean faster transitions and therefore fewer visible image artifacts. Older monitors with long response times would create a smear or blur pattern around moving objects, making them unacceptable for moving video. Current LCD monitor models (such as those with 120Hz, 240Hz, or 480Hz refresh rates) have improved to the point that this is only seen with extreme contrasts. For an LCD display, typical response times are 8 to 16ms for black-white-black, or 2 to 6ms for grey-to-grey. The response time was traditionally recorded at the full black to white transition, which became the ISO standard for this specification on LCDs.

48.11.4 Display Refresh Rate

Refresh rate is distinctly different from response time. Refresh rate is the rate at which a display can draw frames onto the screen, expressed in cycles per second, or Hertz (Hz). Typical television video has al-ways been displayed at 30frames per second, at a rate of 60Hz. However, many display manufacturers are beginning to advertise faster refresh rates as a major selling feature, particularly 120Hz, 240Hz, and 480Hz. Why is this? It relates directly to film. Film, since the days of hand cranked cameras and Thomas Edison, has been recorded at 24 frames per second, providing smooth, clean images. Try dividing 24 frames into 60Hz. The math isn't pretty, and does not work out cleanly. Thus, film video must be modified to run on 60Hz displays. This modification is called 3:2 pulldown. What it means, in short, is that video frames for film, instead of being played 2 times per second each, like television, must be played in a sequence with a frame being shown 3 times, then a frame being shown 2 times, then a frame at 3 times, and so on. This results in a motion jitter and blur. The new high refresh rate standards allow elimination of the 3:2 pulldown, because 24 divides cleanly into 120, 240, and 480, respectively, resulting in smooth video playback

48.12 Multi Display Matrixes

Certain applications may require a configuration of display surface area that is not suited to a single large flat panel device. In these cases, a multi display matrix, commonly called a videowall, may be employed to provide the necessary image size or shape. These arrays can be configured into nearly any size or shape, and can be fit to the specific project requirements, such as covering a large surface area, filling a narrow strip or column, following a curved wall, or fitting in as a mosaic of flat panels.

Videowalls are commonly referred to as the number of vertical displays (rows) by the number of horizontal displays (columns), such as 2 × 2, 2 × 4, 1 × 8, etc. They are typically constructed from flat panel LCD displays, mounted with the display bezels directly against and aligned with each other to make a single solid display. Specialized wall, ceiling, and floor standing mounts are available to make installation, adjustment, and servicing of a videowall easier, although any standard mount may be used.

When specifying a videowall, four things need to be considered: Bezel size, resolution vs. viewing distance, color calibration, and processing.

48.12.1 Bezel Size

The size of the bezels between the displays (also known as mullions) is one of the more obvious considerations in videowall design.

Larger display bezels will produce a larger black “grid” in the image, which detracts from the quality of the display. Specialized videowall panels with bezels ranging from 15mm (per display) down to 1.75mm (per display) allow the designer to minimize the seams in the image, albeit at a higher display cost. Videowall displays may also reference the bezel size in bezel-to-bezel or screen-edge-to-screen-edge sizes. This means the total distance from one lit pixel on a display to the first lit pixel on its adjacent neighbor. Smaller sizes will produce a much more seamless image. The thinner the bezel, the more expensive the display. Best practice is to use the thinnest or slimmest bezel budget will allow. Another good rule of thumb is that the farther away the audience will be from the videowall, the larger the bezel may be. Distance will cause the bezel to blend into the image more.

48.12.2 Resolution

Resolution is not just a function of the number of pixels in an individual display in the videowall, nor is it correct to simply add up the number of pixels present in all displays in the videowall. The question of resolution is related to the number of available pixels, but also to processing and content. As viewers stand closer to a videowall, resolution needs to increase to reduce artifacting in the content. This means that content needs to increase in resolution (and may need to be specially prepared just for the videowall) and that the processor must be able to output enough resolution to meet the needs of the application.

48.12.3 Color Calibration

Because videowall displays may be manufactured at different times, the color calibration and luminance may differ due to changes in materials or manufacturing runs, not unlike purchasing paint at a home improvement center. However, unlike paint, we can calibrate the displays to match each other once they are installed. This should always be done in any videowall installation, to ensure that the color settings of each display match, and that the brightness of each panel matches. This allows the videowall to reproduce content seamlessly without looking like a patchwork quilt of different colors.

48.12.4 Processing

Processing should be prioritized as the key consideration in a videowall system's design. This is the single component which may make or break the videowall's success in the application. The processor is the device that accepts one or more input sources and produces a video output across the entire monitor array. They can vary in cost depending on number of inputs, outputs, and the features available. They will fall into three main categories of processor:

Daisy Chain Scalar. The daisy chain scalar is the most basic type of videowall processor. It offers the ability to take a single input being sent to all of the displays in an array, and scale that single image across multiple monitors. This type of processor is usually built directly into the flat panel, and thus does not have an extra cost. They are often combined with a daisy chain video pass-through to allow simplified installation and cabling. These processors are limited in features, but low to no cost. However, they cannot address anything other than the native resolution of one panel (typically 1080P) and a 16 × 9 aspect ratio. This limits them to being used with “linear” configurations that maintain the 16 × 9 aspect ratio, i.e. 2 × 2, 3 × 3, 4 × 4, etc. Also, they can only handle one input at a time, possibly requiring external switching.

Hardware Based Processor. A hardware based video-wall processor builds on the abilities of a daisy chain scalar, expanding the capabilities of the videowall. These devices are built with a group of video inputs, and a group of video outputs. The internal software allows the user to arrange multiple sources on the screen at once, sometimes incorporating animations and scheduling, depending on the manufacturer and model. On basic models, the number of inputs and outputs will be fixed, but some more advanced models will feature modular “frames” with slots for inputs and outputs. The modular units can be tailored to fit the application, with different input types and output types. Each display will connect to a separate output, and will require running a video cable or balun to connect. This may limit the distance a display can be from the processor, and will require more complex installation.

Software Based Processor. Software based videowall processors offer the same features as a hardware based processor, but replace the video infrastructure with off-the-shelf PC and networking hardware. Common Ethernet switches and cabling form the connection backbone. Each display has a PC connected that renders that portion of the video. Each source is captured by software on a PC, either streaming a desktop or the input from a capture card. This type of videowall can be less expensive and more expandable than a hardware based unit.

48.13 Conclusion

The one thing we can be certain of is that display technologies are constantly evolving with advances taking place in months not years. We can look forward to flat panel displays becoming thinner and lighter with significant power consumption reduction while producing brighter displays with longer panel life. Environmentally unfriendly CCFL backlights in LCD flat panel displays are being replaced with LED illumination and for projectors, hybrid LED/laser illumination will gain acceptance at the middle brightness levels under 6,000 lumens and reduce total cost of ownership by elimination conventional lamps. OLED is set to take on conventional LCD and plasma displays and will continue to decline in popularity over the next few years due to the advances seen in modern LCD displays that now rival plasma in terms of picture performance with less weight, more sizes, and less energy consumption. What we know is that the only constant is change in the world of display technologies and we all benefit in the end.

Part 8

Measurements

Chapter 49

Test and Measurement

by Pat Brown

49.1 -Test and Measurement

Technological advancements in the last few decades have given us a variety of useful measurement tools, and most manufacturers of these instruments provide specialized training on their use. This chapter will examine some principles of test and measurement that are common to virtually all measurement systems. If the measurer understands the principles of measurement, then most any of the mainstream measurement tools will suffice for the collection and evaluation of data. The most important prerequisite to performing meaningful sound system measurements is that the measurer has a solid understanding of the basics of audio and acoustics. The question “How do I perform a measurement?” can be answered much more easily than “What should I measure?” This chapter will touch on both, but readers will find their measurements skills will relate directly to their understanding of the basic physics of sound and the factors that produce good sound quality. The whole of this book will provide much of the required information.

Why Test?

Sound systems must be tested to assure that all components are functioning properly. The test and measurement process can be subdivided into two major categories—electrical tests and acoustical tests. Electrical testing mainly involves voltage and impedance measurements made at component interfaces. Current can also be measured, but since the setup is inherently more complex it is usually calculated from knowledge of the voltage and impedance using Ohm's Law. Also, current measurements are rarely needed, as sound reinforcement systems are interfaced to optimize voltage transfer. Acoustical tests are more complex by nature, but share the same fundamentals as electrical tests in that some time varying quantity (usually pressure) is being measured. The main difference between electrical and acoustical testing is that the interpretation of the latter must deal with the complexities of 3D space, not just amplitude versus time at one point in a circuit. In this chapter I define a loudspeaker as a number of components intentionally combined to produce a system that may then be referred to collectively as a loudspeaker. For example, a woofer, dome tweeter, and crossover network are individual components, but can be combined to form a loudspeaker system, or loudspeaker. Testing usually involves the measurement of systems, although a system might have to be dissected to fully characterize the response of each component, should that information be needed.

49.2 Electrical Testing

There are numerous electrical tests that can be performed on sound system components in the laboratory. The measurement system must have specifications that exceed those of the equipment being measured. Field testing need not be as comprehensive and the tests can be performed with less sophisticated instrumentation. The purpose for testing audio components in the field include:

1. To determine if all system components are operating properly.
2. To diagnose electrical problems in the system, which are usually manifested by some form of distortion.
3. To establish a proper system gain structure.

Electrical test instruments that the author feels are essential to the audio technician include:

- ac voltmeter.
- ac millivoltmeter.
- Oscilloscope.
- Impedance meter.
- Signal generator.
- Polarity test set.

It is important to note that most audio products have on-board metering and/or indicators that may suffice for setting levels, making measurements with stand-alone meters unnecessary. Voltmeters and impedance meters are often only necessary for troubleshooting a nonworking system, or checking the accuracy and calibration of the on-board metering.

There are a number of currently available instruments designed specifically for audio professionals that perform some or all of the functions listed. The days of carrying a separate instrument for each are mostly behind us. The instrument needs to have adequate bandwidth to cover the audible spectrum. Many general purpose "multi-meters" are designed primarily for ac power circuits and do not fit the wide bandwidth requirement for use in audio systems.

More information on electrical testing is included in the chapter on gain structure. The remainder of this chapter will be devoted to the acoustical tests that are required to characterize loudspeakers and rooms.

49.3 Acoustical Testing

The bulk of acoustical measurement and analysis today is being performed by instrumentation that includes or is controlled by a personal computer or tablet. Many excellent systems are available, and the would-be measurer should select the one that best fits their specific needs. As with loudspeakers, there is no clear-cut best choice or one size fits all instrument. Fortunately an understanding of the principles of operating one analyzer can usually be applied to another after a short indoctrination period. Measurement systems are like rental cars—you know what features are there; you just need to find them. In this chapter I will attempt to provide a sufficient overview of the various approaches to allow the reader to investigate and select a tool to meet his or her measurement needs and budget. The acoustical field testing of sound reinforcement systems mainly involves measurements of the pressure response produced by a loudspeaker(s) at various locations in the space. Microphone positions are selected based on the information that is needed. This could be the on-axis position of a loudspeaker for system alignment purposes, or a listener seat for measuring the clarity or intelligibility of the system. Measurements must be made to properly calibrate the sound system, which can include loudspeaker crossover settings, equalization, and the setting of signal delays. Acoustic waveforms are complex by nature, making them difficult to describe with one number readings for anything other than broadband level.

49.3.1 Sound Level Measurements

Sound level measurements are fundamental to all types of audio work. Unfortunately, the question "How loud is it?" does not have a simple answer. Instruments can easily measure sound pressure, but there are many ways to describe the results in ways relevant to human perception. Sound pressure is usually measured at multiple discrete listener positions. The sound pressure level may be displayed as is, integrated over a time interval, or frequency weighted by an appropriate filter. Fast meter response times produce information about peaks and transients in the program material, while slow response times yield data that correlates better with the perceived loudness and energy content of the sound.

A sound level meter consists of a pressure sensitive microphone, meter movement (or digital display), and some supporting circuitry, Fig. 49-1. It is used to observe the sound pressure on a moment-by-moment basis, with the pressure displayed as a level in decibels.

Figure 49-1. A sound level meter is basically a voltmeter that operates in the acoustic domain. Courtesy Galaxy Audio.

Few sounds will measure the same from one instant to the next. Complex sounds such as speech and music will vary dramatically, making their level difficult to describe without a graph of level versus time, Fig. 49-2. A sound level meter is basically a voltmeter that operates in the acoustic domain by incorporating a microphone.

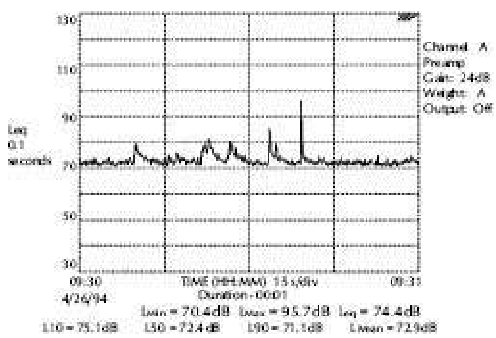


Figure 49-2. A plot of sound level versus time is the most complete way to record the level of an event. Courtesy Gold Line.

Sound pressure measurements are converted into decibels ref. 0.00002 pascals (Pa). See Chapter 5, *Fundamentals of Audio and Acoustics*, for information about the decibel. Twenty micropascals is used as the reference because it is the approximate threshold of pressure sensitivity for humans at midrange frequencies. Such measurements are referred to as *sound pressure level* or L_P (level of sound pressure) measurements, with L_P gaining acceptance among acoustics professionals because it is easily distinguished from L_W (sound power level) and L_I (sound intensity level) and a number of other L_X metrics used to describe sound levels. Sound pressure level is measured at a single point (the microphone position). Sound power measurements must consider all of the radiated sound from a device and sound intensity measurements must consider the sound power flowing through an area. Sound power and sound intensity measurements are usually performed by acoustical laboratories rather than in the field, so neither is considered in this chapter. All measurements described in this chapter will be measurements of sound pressure expressed as levels in dB ref. 0.00002 Pa.

Sound level measurements must usually be processed for the data to correlate with human perception. Humans do not hear all frequencies with equal sensitivity, and to complicate things further our response is dependent on the level that we are hearing. The well-known Fletcher-Munson curves describe the approximate frequency/level characteristics for an average listener, (see Chapter 5, *Fundamentals of Audio and Acoustics*). Sound level measurements are passed through weighting filters that make the meter "hear" with a response similar to a human. Each scale correlates with human hearing sensitivity at a different range of levels. For a sound level measurement to be meaningful, the weighting scale that was used must be indicated, in addition to the response time of the meter. Here are some examples of meaningful (if not universally accepted) expressions of sound level:

- The system produced an L_P = 100dBA (slow response) at mix position.
- The peak sound level was L_A = 115dB at my seat.
- The average sound pressure level was 100dBC at 30ft.
- The loudspeaker produced a continuous L_P of 100dB at one meter (assumes no weighting used).
- The equivalent sound level L_{EQ} was 90dBA at the farthest seat.

Level specifications should be stated clearly enough to allow someone to repeat the test from the description given. Because of the large differences between the weighting scales, it is meaningless to specify a sound level without indicating the scale that was used. An event that produces an L_P = 115dB using a C scale may only measure as an L_P = 95dB using the A scale.

The measurement distance should also be specified (but rarely is). Probably all sound reinforcement systems produce an L_P = 100dB at some distance, but not all do so at the back row of the audience!

L_{pk} is the level of the highest instantaneous peak in the measured time interval. Peaks are of interest because the sound system components must be able to pass them without clipping them. A peak that is clipped produces high levels of harmonic distortion that degrade sound quality. Also, clipping reduces the crest factor of the waveform, causing more heat to be generated in the loudspeaker causing premature failure. Humans are not extremely sensitive to peak levels because our auditory system integrates energy over time with regard to loudness. We are, unfortunately, susceptible to damage from peaks, so they should not be ignored. Research suggests that it takes the brain about 300ms to process sound information with regard to loudness (frequency-dependent), which means that sound events closer together than this are blended together. This is why your voice sounds louder in a small, hard room. It is also why the loudness of the vacuum cleaner varies from room to room. Short interval reflections are integrated with the direct sound by the ear/brain system. Most sound level meters have slow and fast settings that change the response time of the meter. The slow setting of most meters indicates the approximate root-mean-square sound level. This is the effective level of the signal, and should correlate well with its perceived loudness.

A survey of audio practitioners on the SynAudCon forum revealed that most accept an L_P =95dBA (slow response) as the maximum acceptable sound level of a performance at any listener seat for a broad age group audience. The A weighting is used because it considers the sound level in the portion of the spectrum where humans are most easily annoyed and damaged. The slow response time allows the measurement to ignore short duration peaks in the program. A measurement of this type will not indicate true levels for low-frequency information, but it is normally the mid-frequency levels that are of interest.

There exist a number of ways to quantify sound levels that are measured over time. They include:

- $\boldsymbol{L_{PK}}$—the maximum instantaneous peak recorded during the span.
- $\boldsymbol{L_{EQ}}$—the equivalent level (the integrated energy over a specified time interval).
- $\boldsymbol{L_N}$—where L is the level exceeded N percent of the time.
- $\boldsymbol{L_{DEN}}$—a special scale that weights the gathered sound levels based on the time of day. DEN stands for day-evening-night.
- ***DOSE***—a measure of the total sound exposure.

A variety of instruments are available to measure sound pressure level, ranging from the simple sound level meter (SLM) to sophisticated data-logging equipment. SLMs are useful for making quick checks of sound levels. Most have at least an A- and C-weighting scale, and some have octave band filters that allow band-limited measurements. A useful feature on an SLM is an output jack that allows access to the measured data in the form of an ac voltage. Software applications are available that can log the meter's response versus time and display the results in various ways. A plot of sound level versus time is the most complete way to record the level of an event. Fig. 49-2 is such a measurement. Note that a start time and stop time are specified. Such measurements usually provide statistical summaries for the recorded data. An increasing number of venues monitor the level of performing acts in this manner due to growing concerns over litigation about hearing damage to patrons. SLMs vary dramatically in price, depending on quality and accuracy.

All sound level meters provide accurate indications for relative levels. For absolute level measurements a calibrator must be used to calibrate the measurement system. Many PC-based measurement systems have routines that automate the calibration process. The calibrator is placed on the microphone, Fig. 49-3, and the calibrator level (usually 94 or 114dB ref. 20μPa) is entered into a data field. The measurement tool now has a true level to use as a reference for displaying measured data.

Noise criteria ratings provide a one-number specification for allowable levels of ambient noise. Sound level measurements are performed in octave bands, and the results are plotted on the chart shown in Fig. 49-4. The NC rating is read on the right vertical axis. Note that the NC curve is frequency-weighted. It permits an increased level of low-frequency noise, but becomes more stringent at higher frequencies. A sound system specification should include an acceptable NC rating for the space, since excessive ambient noise will reduce system clarity and require additional acoustic gain. This must be considered when designing the sound system. Instrumentation is available to automate noise criteria measurements.

Figure 49-3. A calibrator must be fitted with a disc to provide a snug fit to the microphone. Most microphone manufacturers can provide the disc.

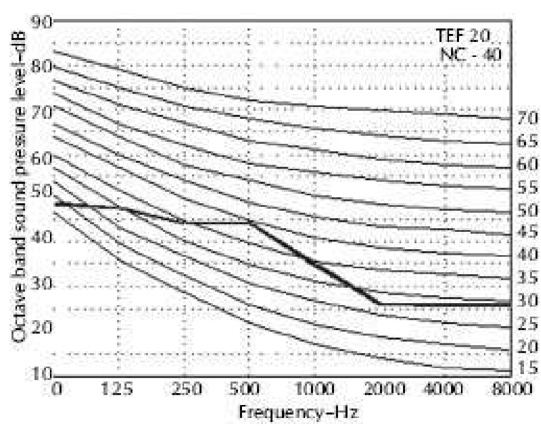


Figure 49-4. A noise criteria specification should accompany a sound system specification.

Stated sound level measurements are often so ambiguous as to be meaningless. It was recently reported in the news that the crowd level in a large stadium reached 137dB. No weighting scale or meter response time were given, making the number meaningless. When stating a sound level, it is important to indicate:

1. The sound pressure level.
2. Any weighting scale used.
3. Meter response time (fast, slow, or other).
4. The distance or location at which the measurement was made.
5. The type of program measured (i.e., music, speech, ambient noise).

Some correct examples:

- "The house system produced 90dBA-Slow in section C for broadband program."
- "The monitor system produced 105dBA-Slow at the performer's head position for broadband program."
- "The ambient noise with room empty was NC-35 with HVAC running."

In short, if you read the *SPL* number and have to request clarification then sufficient information has not been given. As you can see, one-number SPL ratings are rarely useful.

All sound technicians should own a sound level meter, and many can justify investment in more elaborate systems that provide statistics on the measured sound levels. From a practical perspective, it is a worthwhile endeavor to train one's self to recognize various sound levels without a meter, if for no other reason than to find an exit in a venue where excessive levels exist.

49.3.2 Detailed Sound Measurements

The response of a loudspeaker or room must be measured with appropriate frequency resolution to be properly characterized. It is also important for the measurer to understand what the appropriate response should be. If the same criteria were applied to a loudspeaker as to an electronic component such as a mixer, the optimum response would be a flat (minimal variation) magnitude and linear phase response at all frequencies within the required pass band of the system. In reality, we are usually testing loudspeakers to make sure that they are operating at their fullest potential. While flat magnitude and linear phase response are a noble objective, the physical reality is that we must often settle for far less in terms of accuracy. Notwithstanding, even with their inherent inaccuracies, many loudspeakers do an outstanding job of delivering speech or music to the audience. Part of the role of the measurer is to determine if the response of the loudspeaker, the room, or both are inhibiting the required system performance.

49.3.2.1 Sound Persistence in Enclosed Spaces

Sound system performance is greatly affected by the sound energy persistence in the listening space. One metric that is useful for describing this characteristic is the reverberation time, T_{30}. The T_{30} is the time required for an interrupted steady-state sound source to decay to inaudibility. This will be about 60dB of decay in most auditoriums with low ambient noise. The T_{30} designation comes from the practice of measuring 30dB of decay and then doubling the time interval to get the time required for 60dB of decay. A number of methods exist for determining the T_{30}, ranging from simple listening tests to sophisticated analytical methods. Fig. 49-5 shows a simple gated-noise test that can provide sufficient accuracy for designing systems. The bursts for this test can be generated with a WAV editor. Bursts of up to 5 seconds for each of eight octave bands should be generated. Octave band-limited noise is played into the space through a low directivity loudspeaker. The noise is gated on for one second and off for 1 second. The room decay is evaluated during the off span. If it decays completely before the next burst, the T_{30} is less than one second. If not, the next burst should be on for 2 seconds and off for 2 seconds. The measurer simply keeps advancing to the next track until the room completely decays in the off span, Figs. 49-6, 49-7, and 49-8. The advantages of this method include:

1. No sophisticated instrumentation is required.
2. The measurer is free to wander the space.
3. The nature of the decaying field can be judged.
4. A group can perform the measurement.

A test of this type is useful as a prelude to more sophisticated techniques.

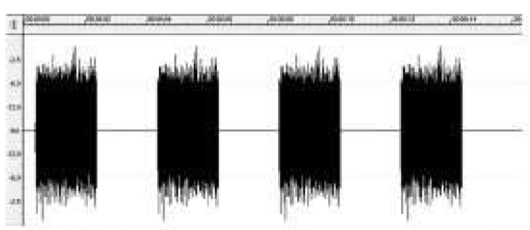


Figure 49-5. Level versus time plot of a one-octave band gated burst (2-second duration).

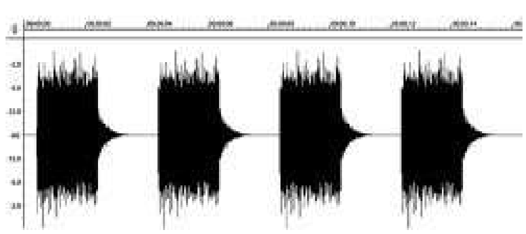


Figure 49-6. A room with RT <2 seconds.

Figure 49-7. A room with RT >2 seconds.

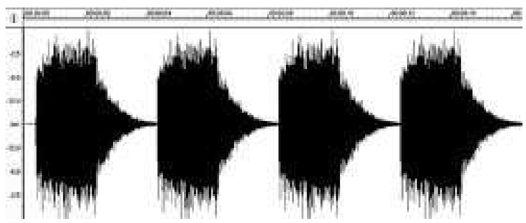


Figure 49-8. A room with RT = 2 seconds.

49.3.2.2 Amplitude versus Time

Fig. 49-9 shows an audio waveform displayed as amplitude versus time. The waveform shown is of a male talker recorded in an anechoic (echo-free) environment. The 0 line represents the ambient (no signal) state of the medium being modulated. This would be ambient atmospheric pressure for an acoustical wave, or zero volts or a dc offset for an electrical waveform measured at the output of a system component.

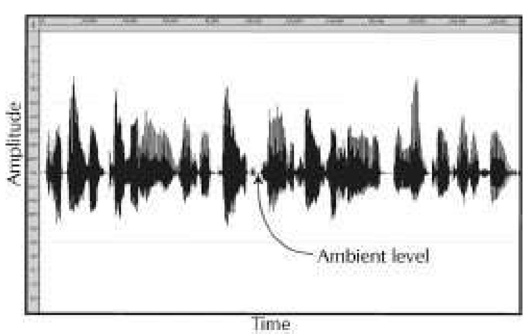


Figure 49-9. Amplitude versus time plot of a male talker made in

an anechoic environment.

Fig. 49-10 shows the same waveform, but this time played over a loudspeaker into a room and recorded. The waveform has now been encoded (convolved) with the response of the loudspeaker and room. It will sound completely different than the anechoic version.

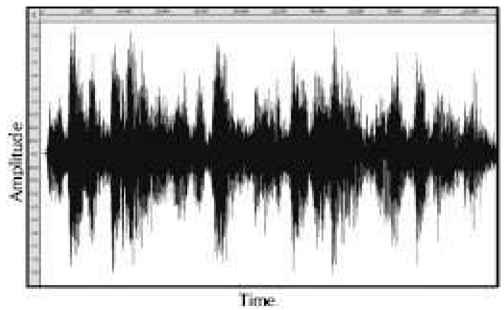


Figure 49-10. The voice waveform after encoding with the room response.

Fig. 49-11 shows an impulse response and Fig. 49-12 shows the envelope-time curve (ETC) of the loudspeaker and room. It is essentially the difference between Fig. 49-9 and Fig. 49-10 and fully characterizes any effect that the loudspeaker or room has on the electrical signal fed to the loudspeaker and measured at that point in space. Most measurement systems attempt to measure the impulse response, since knowledge of the impulse response of a system allows its effect on any signal passing through it to be determined, assuming the system is linear and time invariant. The impulse response is represented in the frequency domain as the transfer function of the system and includes both magnitude (level) and phase (timing) information for each frequency in the pass band. Both the loudspeaker and room can be considered filters that the energy must pass through en route to the listener. Treating them as filters allows their responses to be measured and displayed, and provides an objective benchmark to evaluate their effect. It also opens loudspeakers and rooms to evaluation by electrical network analysis methods, which are generally more widely known and better developed than acoustical measurement methods.

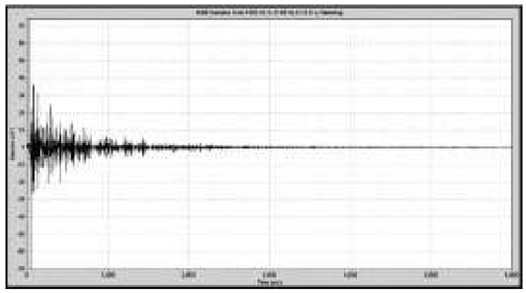


Figure 49-11. The impulse response of the acoustic environment.

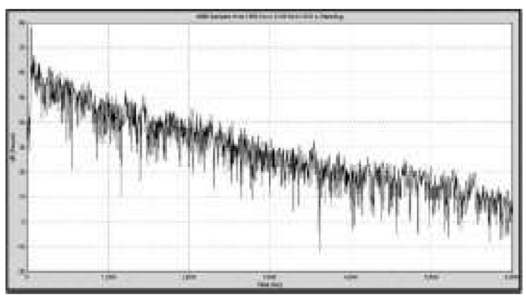


Figure 49-12. The envelope-time curve (ETC) of the same environment. It can be derived from the impulse response.

49.3.2.3 The Transfer Function

The effect that a filter has on a waveform is called its *transfer function*. A transfer function can be found by comparing the input signal and output signal of the filter. It matters little if the filter is an electronic component, loudspeaker, room, or listener. The time domain behavior of a system (impulse response) can be displayed in the frequency domain as a spectrum and phase (transfer function). Either the time or frequency description fully describes the filter. Knowledge of one allows the determination of the other. The mathematical map between the two representations is called a transform, specifically the Fourier transform. Transforms can be performed at amazingly fast speeds by computers. Fig. 49-13 shows a domain chart that provides a map between various representations of a system's response. The measurer must remember that the responses being measured and displayed on the analyzer are dependent on the test stimulus used to acquire the response. Appropriate stimuli must have adequate energy content over the pass band of the system being measured. In other words, we can't measure a subwoofer using a flute solo as a stimulus. With that criteria met, the response measured and displayed on the analyzer is independent of the program material that passes through a linear system. Pink noise and sine sweeps are common stimuli due to their broadband spectral content. The response of a linear system doesn't change relative to the nature of the program material. For a linear system, the transfer function is a summary that says, "If you put energy into this system, this is what will happen to it."

The domain chart provides a map between various methods of displaying the system's response. The utility of this is that it allows measurement in either the time or frequency domain. The alternate view can be determined mathematically by use of a transform. This allows frequency information to be determined with a time domain measurement, and time information to be determined by a frequency domain measurement. This important inverse relationship between time and frequency can be exploited to yield many possible ways of measuring a system and/or displaying its response. For instance, a noise immunity characteristic not attainable in the time domain may be attainable in the frequency domain. This information can then be viewed in the time domain by use of a transform. The Fourier Transform and its inverse are commonly employed for this purpose. Measurement software programs can display the signal in either domain, Fig. 49-14.

49.3.3 Measurement Systems

Any useful measurement system must be able to extract the system response in the presence of noise. In some applications, the signal-to-noise requirements might actually determine the type of analysis that will be used. Some of the simplest and most convenient tests have poor signal-to-noise performance, while some of the most complex and computationally demanding methods can measure under almost any conditions. The measurer must choose the type of analysis with these factors in mind. It is possible to acquire the impulse response of a filter without using an impulse. This is accomplished by feeding a known broadband stimulus into the filter and reacquiring it at the output. A complex comparison of the two signals (mathematical division) yields the transfer function, which is displayed in the frequency domain as a magnitude and phase or inverse-transformed for display in the time domain as an impulse response. The impulse response of a system answers the question, “If I feed a perfect impulse into this system, when will the energy exit the system?” A knowledge of “when” can characterize a system. After transformation, the spectrum or frequency response is displayed on a decibel scale. A phase plot shows the phase response of the device-under-test, and any phase shift versus frequency becomes apparent. If an impulse response is a measure of when, we might describe a frequency response as a measure of what. In other words, “If I input a broadband stimulus (all frequencies) into the system, what frequencies will be present at the output of the system and what will their phase relationship be?” A transfer function includes both magnitude and phase information.

Alternate Perspectives

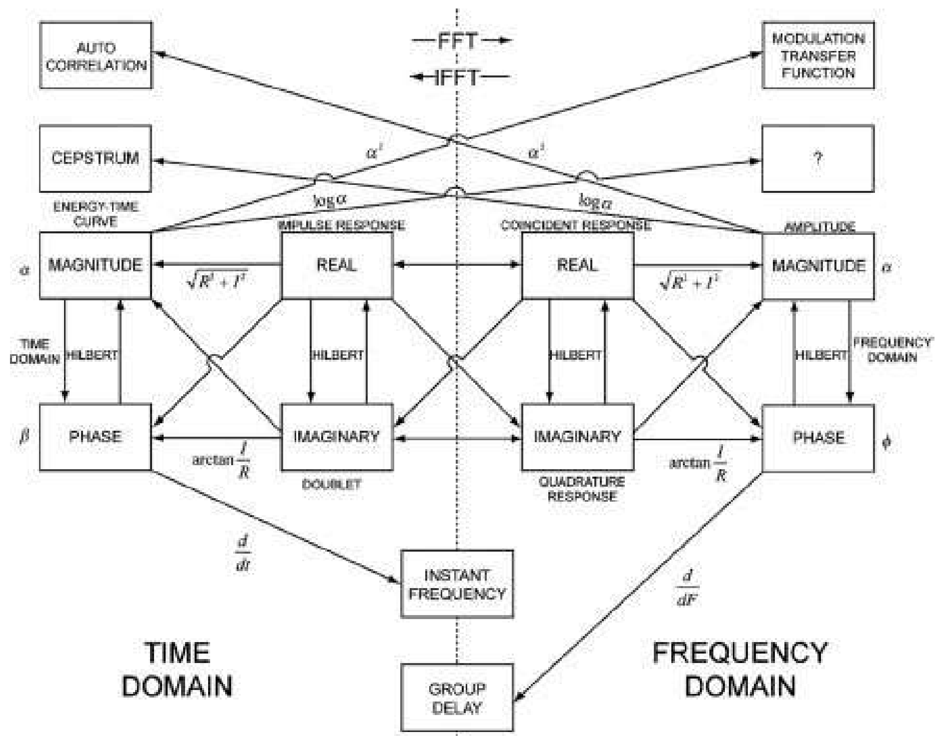


Figure 49-13. The domain chart provides a map between various representations of a system response. Courtesy Brüel and Kjaer.

The time and frequency views of a system’s response are mutually exclusive. By definition the time period of a periodic event is

$$T = \frac{1}{f} \qquad (49\text{-}1)$$

where,
T is time in s,
f is frequency in Hz.

Since time and frequency are reciprocals, a view of one excludes a view of the other. Frequency information cannot be observed on an impulse response plot, and time information can't be observed on a magnitude/phase plot. Any attempt to view both simultaneously will obscure some of the detail of both. Modern analyzers allow the measurer to switch between the time and frequency perspectives to extract information from the data.

49.3.4 Testing Methods

Compared to other components in the sound system, the basic design of loudspeakers and compression drivers has changed relatively little in the last 50 years. At over a half-century since their invention, we are still pushing air with pistons driven by voice coils suspended in magnetic fields. But the methods for measuring their performance have improved steadily since computers can now efficiently perform digital sampling and signal processing, and execute transforms in fractions of a second. Extremely capable measurement systems are now accessible and affordable to even the smallest manufacturers and individual audio practitioners. A common attribute of systems suitable for loudspeaker testing is the ability to make reflection-free measurements indoors, without the need for an anechoic chamber. Anechoic measurements in live spaces can be accomplished by the use of a time window that allows the analyzer to collect the direct field response of the loudspeaker while ignoring room reflections. Conceptually, a time window can be thought of as an accurate switch that can be closed as the desired waves pass the microphone and opened prior to the arrival of undesirable reflections from the environment. A number of implementations exist, each with its own set of advantages and drawbacks. The potential buyer must understand the trade-offs and choose a system that offers the best set of compromises for the intended application. Parameters of interest include signal-to-noise ratios, speed, resolution, and price.

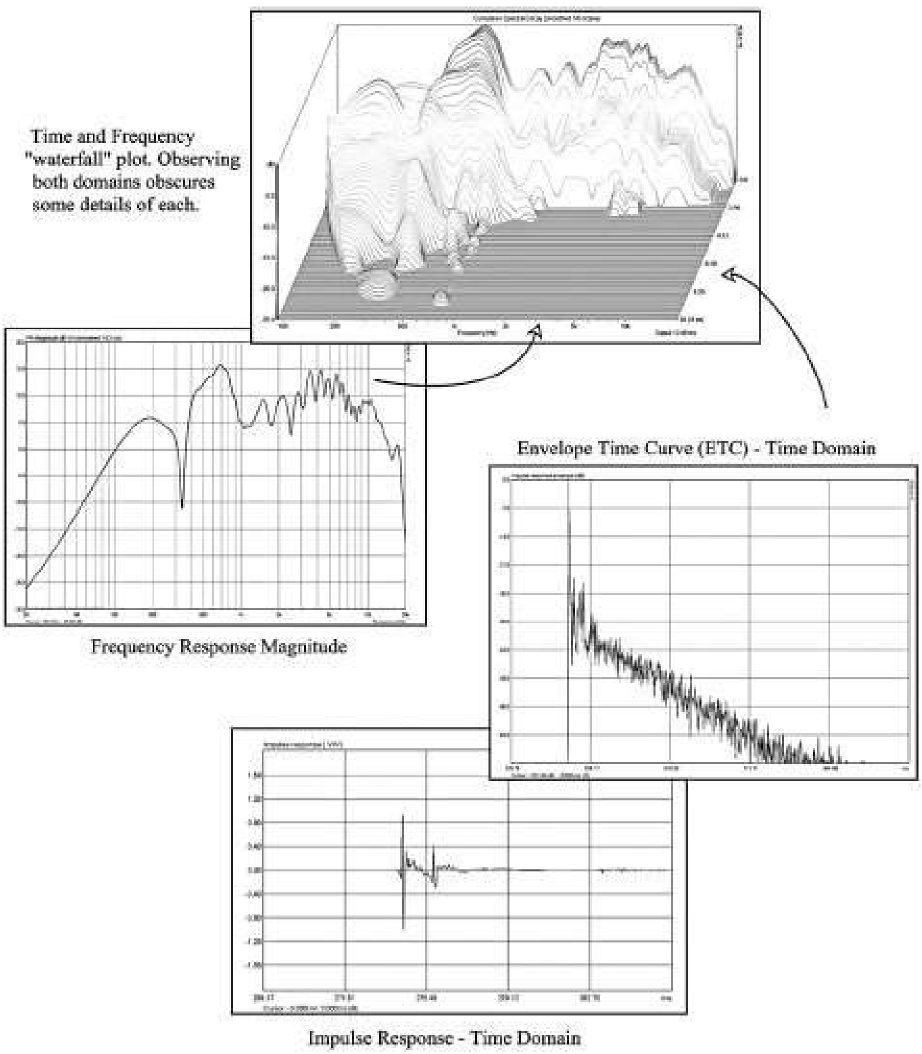


Figure 49-14. The FFT can be used to view the spectral content of a time domain measurement, Arta 1.2.

49.3.4.1 FFT Measurements

The Fourier Transform is a mathematical filtering process that determines the spectral content of a time domain signal. The Fast

Fourier Transform, or FFT, is a computationally efficient version of the same. Most modern measurement systems make use of the computer's ability to quickly perform the FFT on sampled data. The cousin to the FFT is the IFFT, or Inverse Fast Fourier Transform. As one might guess, the IFFT takes a frequency domain signal as its input and produces a time domain signal. The FFT and IFFT form the bedrock of modern measurement systems. Many fields outside of audio use the FFT to analyze time records for periodic activity, such as utility companies to find peak usage times or an investment firm to investigate cyclic stock market behavior. Analyzers that use the Fast Fourier Transform to determine the spectral content of a time-varying signal are collectively called FFTs. If a broadband stimulus is used, the FFT can show the spectral response of the device under test (DUT). One such stimulus is the unit impulse, a signal of theoretically infinite amplitude and infinitely small time duration. The FFT of such a stimulus is a straight, horizontal line in the frequency domain.

The time-honored hand clap test of a room is a crude but useful form of impulse response. The hand clap is useful for casual observations, but more accurate and repeatable methods are usually required for serious audio work. The drawbacks of using impulsive stimuli to measure a sound system include:

1. Impulses can drive loudspeakers into nonlinear behavior.
2. Impulse responses have poor signal-to-noise ratios, since all of the energy enters the system at one time and is reacquired over a longer span of time along with the noise from the environment.
3. There is no way to create a perfect impulse, so there will always be some uncertainty as to whether the response characteristic is

that of the system, the impulse, or some nonlinearity arising from impulsing a loudspeaker.

Even with its drawbacks, impulse testing can provide useful information about the response of a loudspeaker or room.

49.3.4.2 Dual-Channel FFT

When used for acoustic measurements, dual-channel FFT analyzers digitally sample the signal fed to the loudspeaker, and also digitally sample the acoustic signal from the loudspeaker at the output of a test microphone. The signals are then compared by division, yielding the transfer function of the loudspeaker. Dual-channel FFTs have the advantage of being able to use any broadband stimulus as a test signal. Pink noise and swept sines provide excellent stability and noise immunity. It is a computationally intense method since both the input and output signal must be measured simultaneously and compared, often in real time. For a proper comparison to yield a loudspeaker transfer function, it is important that the signals being compared have the same level, and that any time offsets between the two signals be removed if accurate phase information is needed. Dual-channel FFT analyzers have set up routines that simplify the establishment of these conditions. Portable computers have A/D converters as part of their on-board sound system, as well as a microprocessor to perform the FFT. With the appropriate software and sound system interface they form a powerful, low-cost and portable measurement platform.

49.3.4.3 Maximum-Length Sequence

The maximum-length sequence (MLS) is a pseudorandom noise test stimulus. The MLS overcomes some of the shortcomings of the dual-channel FFT, since it does not require that the input signal to the system be measured. A binary string (ones and zeros) is fed to the device under test while simultaneously being stored for future correlation with the loudspeaker response acquired by the test microphone. The pseudorandom sequence has a white spectrum (equal energy per Hz), and is exactly known and exactly repeatable. Comparing the input string with the string acquired by the test microphone yields the transfer function of the system. The advantage of the MLS is its excellent noise immunity and fast measurement time, making it a favorite of loudspeaker designers. A disadvantage is that the noise-like stimulus can be annoying, sometimes requiring that measurements be done after hours. The use of MLS has waned in recent years to log-swept sine measurements made on dual-channel FFT analyzers.

49.3.4.4 Time-Delay Spectrometry (TDS)

TDS is a fundamentally different method of measuring the transfer function of a system. Richard Heyser, a staff scientist at the Jet Propulsion Laboratories, invented the method. An anthology of Mr. Heyser's papers on TDS is available in the reference. Both the dual-channel FFT and MLS methods involve digital sampling of a broadband stimulus. TDS uses a method borrowed from the world of sonar, where a single-frequency sinusoidal "chirp" signal is fed to the system under test. The chirp slowly sweeps through the frequencies being measured, and is reacquired with a tracking filter by the TDS analyzer. The reacquired signal is then mixed with the outgoing signal, producing a series of sum and difference frequencies, each frequency corresponding to a different arrival time of sound at the microphone. The difference frequencies are transformed to the time domain with the appropriate transform, yielding the Envelope Time Curve (ETC) of the system under test. TDS is based in the frequency domain, allowing the tracking filter to be tuned to the desired signal while ignoring signals outside of its bandwidth. TDS offers excellent noise immunity, allowing good data to be collected under near-impossible measurement conditions. Its downside is that good low-frequency resolution can be difficult to obtain without extended measurement times, plus the correct selection of measurement parameters requires a knowledgeable user. In spite of this, it is a favorite among contractors and consultants, who must often perform sound system calibrations in the real world of air conditioners, vacuum cleaners, and building occupants.

While other measurement methods exist, the ones outlined above make up the majority of methods used for field and lab testing of loudspeakers and rooms. Used properly, any of the methods can provide accurate and repeatable data. Many audio professionals have several measurement platforms and exploit the strengths of each when measuring a sound system.

49.3.5 Preparation

There are many measurements that can be performed on a sound system. A prerequisite to any measurement is to answer the following questions:

1. What am I trying to measure?
2. Why am I trying to measure it?
3. Is it audible?

4. Is it relevant?

Failure to consider these questions can lead to hours of wasted time and a hard drive full of meaningless data. Even with the incredible technologies that we have available to us, the first part of any measurement session is to listen. It can take many hours to determine what needs to be measured to solve a sound system problem, yet the actual measurement itself can often be completed in seconds. Using an analogy from the medical field, the physician must query the patient at length to narrow down the ailment. The more that is known about the ailment, the more specific and relevant the tests that can be run for diagnosis. There is no need to test for tonsillitis if the problem is a sore back!

1. What am I measuring? A fundamental decision that precedes a meaningful measurement is how much of the room's response to include in the measured data. Modern measurement systems have the ability to perform semi-anechoic measurements, and the measurer must decide if the loudspeaker, the room, or the combination needs to be measured. If one is diagnosing loudspeaker ailments, there is little reason to select a time window long enough to include the effects of late reflections and reverberation. A properly selected time window can isolate the direct field of the loudspeaker and allow its response to be evaluated independently of the room. If one is trying to measure the total decay time of the room, the direct sound field becomes less important, and a microphone placement and time window are selected to capture the entire energy decay. Most modern measurement systems acquire the complete impulse response, including the room decay, so the choice of the time window size

can be made after the fact during post-processing.

2. Why am I measuring? There are several reasons for performing acoustic measurements in a space. An important reason for the system designer is to characterize the listening environment. Is it dead? Is it live? Is it reverberant? These questions must be considered prior to the design of a sound system for the space. While the human hearing system can provide the answers to these questions, it cannot document them and it is easily deceived. Measurements might also be performed to document the performance of an existing system prior to performing changes or adding room treatment. Customers sometimes forget how bad it once sounded after a new or upgraded system is in place for a few weeks. The most common reason for performing measurements on a system is for calibration purposes. This can include equalization, signal alignment, crossover selection, and a multiplicity of other reasons. Since loudspeakers interact in a complex way with their environment, the final phase of any system installation is to verify system performance by measurement.
3. Is it audible? Can I hear what I am trying to measure? If one cannot hear an anomaly, there is little reason to attempt to measure it. The human hearing system is perhaps the best tool available for determining what should be measured about a sound system. The human hearing system can tell us that something doesn't sound right, but the cause of the problem can be revealed by measurement. Anything you can hear can be measured, and once it is measured it can be quantified and possibly manipulated.
4. Is it relevant? Am I measuring something that is worth measuring? If one is working for a client, time is money.

> Measurements must be prioritized to focus on audible problems. Endless hours can be spent “chasing rabbits” by measuring details that are of no importance to the client. This is not necessarily a fruitless process, but it is one that should be done on your own time. I have on several occasions spent time measuring and documenting anomalies that had nothing to do with the customer’s reason for calling me. All venues have problems that the owner is unaware of. Communication with the client is the best way to avoid this pitfall.

49.3.5.1 Dissecting the Impulse Response

The audio practitioner is often faced with the dilemma of determining whether the reason for bad sound is the loudspeaker system, the room, or an interaction of the two. The impulse response can hold that answer to these and other perplexing questions. The impulse response in its amplitude versus time display is not particularly useful for other than determining the polarity of a system component, Fig. 49-15. A better representation comes from squaring impulse response (making all deflections positive) and displaying the square root of the result on a logarithmic vertical axis. This log-squared response allows the relative levels of energy arrivals to be compared, Fig. 49-16.

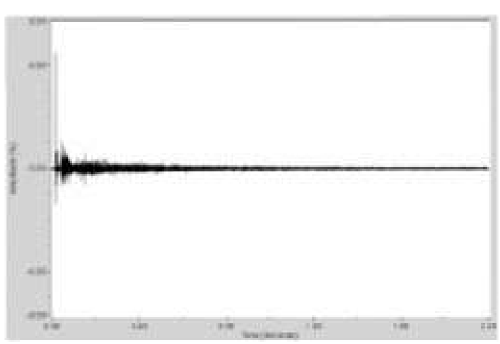


Figure 49-15. The impulse response, SIA-SMAART.

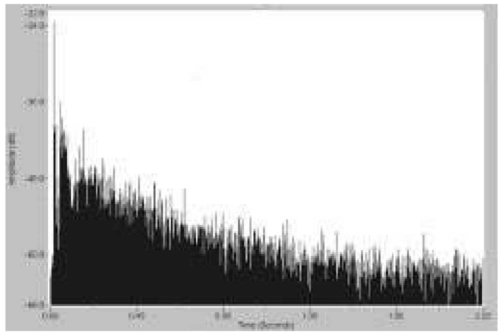


Figure 49-16. The log-squared response, SIA-SMAART.

49.3.5.2 The Envelope-Time Curve

Another useful way of viewing the impulse response is in the form of the Envelope Time Curve, or ETC. The ETC is also a contribution of Richard Heyser. It takes the real part of the impulse response and combines it with a 90° phase shifted version of the same, Fig. 49-17. One way to get the shifted version is to use the Hilbert Transform. The complex combination of these two signals yields a time domain waveform that is often easier to interpret than the impulse response. The ETC can be loosely thought of as a smoothing function for the log-squared response, showing the envelope of the data. This can be more revealing as to the audibility of an event. The impulse response, log-squared response, and envelope-time curve are all different ways to view the time domain data.

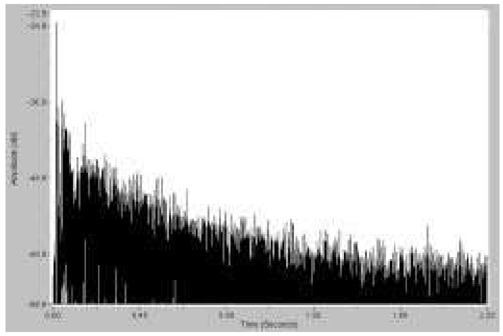


Figure 49-17. The Envelope Time Curve (ETC), SIA-SMAART.

49.3.5.3 A Global Look

When starting a measurement session, a practical approach is to first take a global look and measure the complete decay of the room. The measurer can then choose to ignore part of the time record by using a time window to isolate the desired part during post-processing. The length of the time window can be increased to include the effects of more of the energy returned by the room. The time window can also be used to isolate a reflection and view its spectral content. Just like your life span represents a time window in human history, a time window can be used to isolate parts of the impulse response from the total time record.

49.3.5.4 Time Window Length

The impulse response can be sub-divided to identify the arrivals that can be attributed to the loudspeaker and those which can be attributed to the room. It must be emphasized that there is a rather gray and frequency-dependent line between the two, but for this discussion we will assume that we can clearly separate them. The direct field is the energy that arrives at the listener prior to any reflections from the room. The division is fairly distinct if neither the loudspeaker nor microphone is placed near any reflecting surfaces, which, by the way, is a good system design practice. At long wavelengths (low frequencies) the direct field may include the effects of boundaries near the loudspeaker and microphone. As frequency increases, the sound from the loudspeaker becomes less affected by boundary effects (due in part to increased directivity) and can be measured independently of them. Proper loudspeaker placement produces a time gap between the sound energy arrivals from the loudspeaker and the later arriving room response. We can use this time gap to aid in selecting a time window to separate the loudspeaker response from the room response for diagnosing system problems.

49.3.5.5 Acoustic Wavelengths

Sound travels in waves. The sound waves that we are interested in characterizing have a physical size. There will be a minimum time span required to observe the spectral content of a waveform. The minimum required length of time to view an acoustical event is determined by the longest wavelength (lowest frequency) present in the event. At the upper limits of human hearing, the wavelengths are only a few millimeters in length, but as frequency decreases the waves become increasingly larger. At the lowest frequencies that humans hear, the wavelengths are many meters long, and can actually be larger than the listening (or measurement) space. This makes it difficult to measure low frequencies from a loudspeaker independently of the listening space, since low frequencies radiated from a loudspeaker interact (couple) with the surfaces around them. In an ideally positioned loudspeaker, the first energy arrival from the loudspeaker at mid- and high frequencies has already dissipated prior to the arrival of reflections and can therefore often be measured independently of them. The human hearing system tends to fuse the direct sound from the loudspeaker with the early reflections from nearby surfaces with regard to level (loudness) and frequency (tone). It is usually useful to consider them as separate events, especially since the time offset between the direct sound and first reflections will be unique for each listening position. This precludes any type of frequency domain correction (i.e., equalization) of the room/loudspeaker response other than at frequencies where coupling occurs due to close proximity to nearby surfaces. While it is possible to compensate to some extent for room reflections at a point in space (acoustic echo cancellers used for conference systems), this correction cannot be extended to include an area. This inability to compensate for the reflected energy at mid/high frequencies suggests that their effects be removed from the loudspeaker's direct field response prior to meaningful equalization work by use of an appropriate time window.

49.3.5.6 Microphone Placement

A microphone is needed to acquire the sound radiated into the space from the loudspeaker at a discrete position. Proper microphone placement is determined by the type of test being performed. If one were interested in measuring the decay time of the room, it is usually best to place the microphone well beyond critical distance. This allows the build-up of the reverberant field to be observed as well as providing good resolution of the decaying tail. Critical distance is the distance from the loudspeaker at which the direct field level and reverberant field level are equal. It is described further in Section 49.3.5.7. If it's the loudspeaker's response that needs to be measured, then a microphone placement inside of critical distance will provide better data on some types of analyzers, since the direct sound field is stronger relative to the later energy returning from the room. If the microphone is placed too close to the loudspeaker, the measured sound levels will be accurate for that position, but may not accurately extrapolate to greater distances with the inverse-square law. As the sound travels farther, the response at a remote listening position may bear little resemblance to the response at the near field microphone position. For this reason, it is usually desirable to place the microphone in the far free field of the loudspeaker—not too close and not too far away. The approximate extent of the near field can be determined by considering that the path length difference from the measurement position (assumed axial) and the edge of the sound radiator should be less than one-quarter wavelength at the frequency of interest. This condition is easily met for a small loudspeaker that is radiating low frequencies. Such devices closely approximate an ideal point source. As the frequency increases the condition becomes more difficult to satisfy, especially if the size of the radiator also increases. Large radiators (or groups of radiators) emitting high frequencies can extend the near field to very long distances. Line arrays make use of this principle to overcome the inverse-square law. In practice, small bookshelf loudspeakers can be accurately measured at a few meters. About 8m is a common measurement distance for moderate-sized, full-range loudspeakers in a large space. Even greater distances are required for large devices radiating high frequencies. A general guideline is to not put the mic closer than three times the loudspeaker's longest dimension.

49.3.5.7 Estimate the Critical Distance D_C

Critical distance is easy to estimate. A quick method with reasonable accuracy requires a sound level meter and noise source. Ideally, the noise source should be band limited, as critical distance is frequency dependent. The 2kHz octave band is a good place to start when measuring critical distance. Proceed as follows:

1. Energize the room with pink noise in the desired octave band from the sound source being measured. The level should be at least 10dB higher than the background noise in the same octave band.
2. Using the sound level meter, take a reading near the loudspeaker (about 1m) and on-axis. At this distance, the direct sound field will dominate the measurement.
3. Move away from the loudspeaker while observing the sound level meter. The sound level will fall off as you move farther away. If you are in a room with a reverberant sound field, at some distance the meter reading will quit dropping. You have now moved beyond critical distance. Measurements of the direct field beyond this point will be a challenge for some types of analysis. Move back toward the loudspeaker until the meter begins to rise again. You are now entering a good region to perform acoustic measurements on loudspeakers in this environment. The above process provides an estimate that is adequate for positioning a measurement microphone for loudspeaker testing. With a mic placement inside of critical distance, the direct field is a more dominant feature on the impulse response and a time window will be more effective in removing room reflections.

At this point it is interesting to wander around the room with the sound level meter and evaluate the uniformity of the reverberant field. Rooms that are reverberant by the classical definition will vary little in sound level beyond critical distance when energized with a continuous noise source. Such spaces have low internal sound absorption relative to their volume.

49.3.5.8 Common Factors to All Measurement Systems

Let's assume that we wish to measure the impulse response of a loudspeaker/room combination. While it would not be practical to measure the response at every seat, it is good measurement practice to measure at as many seats as are required to prove the performance of the system. Once the impulse response is properly acquired, any number of post-processes can be performed on the data to extract information from it. Most modern measurement systems make use of digital sampling in acquiring the response of the system. The fundamentals and prerequisites are not unlike the techniques used to make any digital recording, where one must be concerned with the level of an event and its time length. Some setup is required and some fundamentals are as follows:

1. The sampling rate must be fast enough to capture the highest frequency component of interest. This requires at least two samples of the highest frequency present in the signal. If one wished to measure to 20kHz, the required sample rate would need to be at least 40kHz. Most measurement systems sample at 44.1kHz or 48kHz, more than sufficient for acoustic measurements.
2. The time length of the measurement must be long enough to allow the decaying energy curve to flatten out into the room noise floor. Care must be taken to not cut off the decaying

energy, as this will result in artifacts in the data, like a scratch on a phonograph record. If the sampling rate is 44.1kHz, then 44,100 samples must be collected for each second of room decay. A 3-second room would therefore require $44.1 \times 1000 \times 3$ or 128,000 samples. A hand clap test is a good way to estimate the decay time of the room and therefore the required number of samples to fully capture it. The time span of the measurement also determines the lowest frequency that can be resolved from the measured data, which is approximately the inverse of the measurement length. The sampling rate can be reduced to increase the sampling time to yield better low-frequency information.

3. The measurement must have a sufficient signal-to-noise ratio to allow the decaying tail to be fully observed. This often requires that the measurement be repeated a number of times and the results averaged. Using a dual-channel FFT or MLS, the improvement in *SNR* will be 3dB for each doubling of the number of averages. Ten averages is a good place to start, and this number can be increased or decreased depending on the environment. The level of the test stimulus is also important. Higher levels produce improved *SNR*, but can also stress the loudspeaker.
4. Perform the test and observe the data. It should fill the screen from top left to bottom right and should fully decay before reaching the right side of the screen. It should also be repeatable. Run the test several times to check for consistency. Background noise can dramatically affect the repeatability of the measurement and the validity of the data.

Once the impulse response is acquired, it can be further analyzed for spectral content, intelligibility information, decay time, etc. These are referred to as *metrics*, and some require some knowledge on the part of the measurer in properly placing markers (called *cursors*) to identify the parameters required to perform the calculations. Let us look at how the response of the loudspeaker might be extracted from the data just gathered.

The time domain data displays what would have resulted if an impulse were fed through the system. Don't try to correlate what you see on the analyzer with what you heard during the test. Most measurement systems display an impulse response that is calculated from a knowledge of the input and output signal to the system, and there is no correlation between what you hear when the test is run and what you are seeing on the screen, Fig. 49-18.

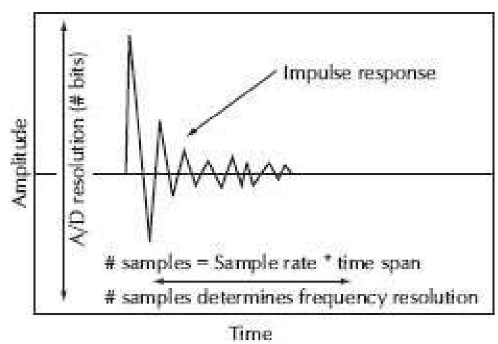


Figure 49-18. Many analyzers acquire the room response by digital sampling.

We can usually assume that the first energy arrival is from the loudspeaker itself, since any reflection would have to arrive later than the first wave front since it had to travel farther. Pre-arrivals can be caused by the acoustic wave propagating through a solid object, such as a ceiling or floor and reradiating near the microphone. Such arrivals are very rare and usually quite low in level. In some cases a reflection may actually be louder than the direct arrival. This could be due to loudspeaker design or its placement relative to the mic location. It's up to the measurer to determine if this is normal for a given loudspeaker position/seating position. All loudspeakers will have some internal and external reflections that will arrive just after the first wave front. These are actually a part of the loudspeaker's response and can't be separated from the first wave front with a time window due to their close proximity without an extreme compromise in frequency resolution. Such reflections are at least partially responsible for the characteristic sound of a loudspeaker. Studio monitor designers and studio control room designers go to great lengths to reduce the level of such reflections, yielding more accurate sound reproduction. Good system design practice is to place loudspeakers as far as possible from boundaries (at least at mid- and high frequencies). This will produce an initial time gap between the loudspeaker's response and the first reflections from the room. This gap is a good initial dividing point between the loudspeaker's response and the room's response, with the energy to the left of the dividing cursor being the response of the loudspeaker and the energy to the right the response of the room. The placement of this divider can form a time window by having the analyzer ignore everything later in time than the cursor setting. The time window size also determines the frequency resolution of the postprocessed data. In the frequency domain, improved resolution means a smaller number. For instance, 10Hz resolution is higher than 40Hz resolution. Since time and frequency have an inverse relationship, the time window length required to observe 10Hz will be much longer than the time window length required to resolve 40Hz. The resolution can be estimated by $f = 1/T$, where T is the length of the time window in seconds. Since a frequency magnitude plot is made up of a number of data points connected by a line, another way to view the frequency resolution is that it is the number of Hz between the data points in a frequency domain display.

The method of determination of the time window length varies with different analyzers. Some allow a cursor to be placed anywhere on the data record, and the placement determines the frequency resolution of the spectrum determined by the window length. Others require that the measurer select the number of samples to be used to form the time window, which in turn determines the frequency resolution of the time window. The window can then be positioned at different places on the time domain plot to observe the spectral content of the energy within the window, Figs. 49-19, 49-20, and 49-21.

For instance, a 1 second total time (44,100 samples) could be divided into about twenty two time windows of 2048 samples each (about 45ms). Each window would allow the observation of the spectral content down to 1000/45 or 22Hz. The windows can be overlapped and moved around to allow more precise selection of the time span to be observed. Displaying a number of these time windows in succession, each separated by a time offset, can form a 3D plot known as a waterfall.

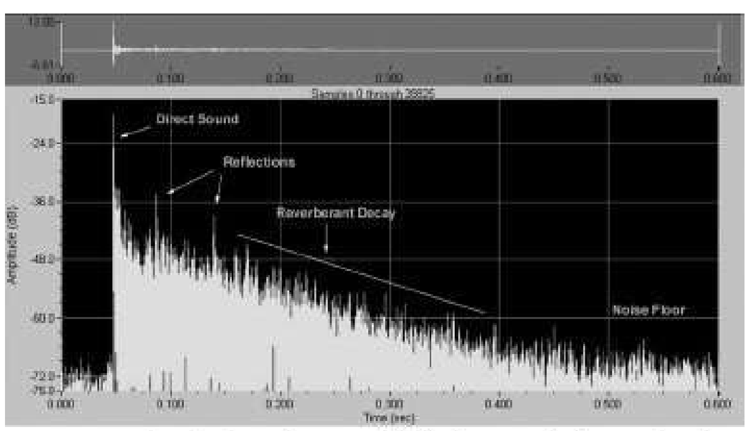


Figure 49-19. A room response showing the various sound fields that can exist in an enclosed space, SIA-SMAART.

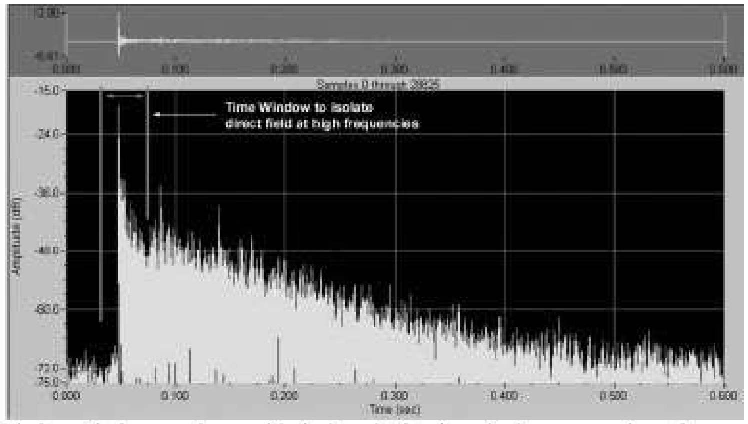


Figure 49-20. A time window can be used to isolate the loudspeaker's response from the room reflections.

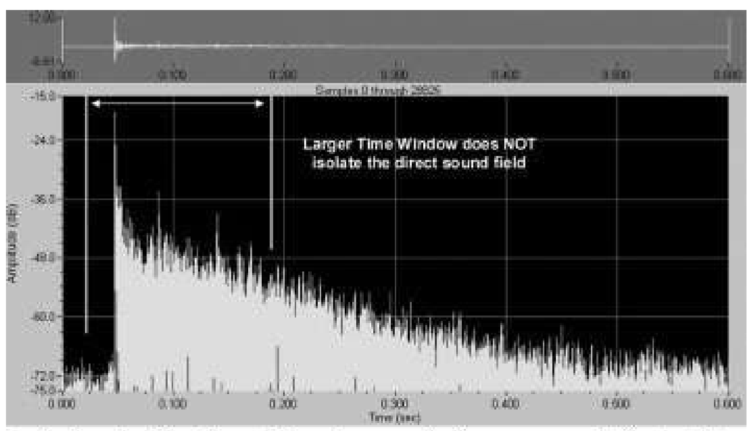


Figure 49-21. Increasing the length of the time window increases the frequency resolution, but lets more of the room into the measurement, SIA-SMAART.

49.3.5.9 Data Windows

There are some conditions that must be observed when placing cursors to define the time window. Ideally, we would like to place the cursor at a point on the time record where the energy is zero. A cursor placement that cuts off an energy arrival will produce a sharp rise or fall time that produces artifacts in the resultant calculated spectral response. Discontinuities in the time domain have broad spectral content in the frequency domain. A good example is a scratch on a phonograph record. The discontinuity formed by the scratch manifests itself as a broadband click during playback. If an otherwise smooth wheel has a discontinuity at one point, it would thump annoyingly when it was rolled on a smooth surface. Our measurement systems treat the data within the selected window as a continuously repeating event. The end of the event must line up with the beginning or a discontinuity occurs resulting in the generation of high-frequency artifacts called *spectral leakage*. In the same manner that a physical discontinuity in a phonograph record or wheel can be corrected by polishing, a discontinuity in a sampled time measurement can be remedied by tapering the energy at the beginning and end of the window to zero using a mathematical function. A number of data window shapes are available for performing the smoothing.

These include the Hann, Hamming, Blackman-Harris, and others. In the same way that a physical polishing process removes some good material from what is being rubbed, data windows remove some good data in the process of smoothing the discontinuity. Each window has a particular shape that leaves the data largely untouched at the center of the window but tapers it to varying degrees toward the edges. Half windows only smooth the data at the right edge of the time record while full windows taper both (start and stop) edges. Since all windows have side effects, there is no clear preference as to which one should be used. The Hann window provides a good compromise between time record truncation and data preservation. Figs. 49-22 and 49-23 show how a data window might be used to reduce spectral leakage.

49.3.5.10 A Methodical Approach

Since there are an innumerable number of tests that can be performed on a system, it makes sense to establish a methodical and logical process for the measurement session. One such scenario may be as follows:

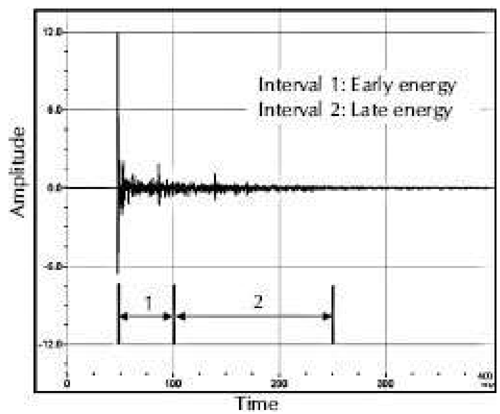


Figure 49-22. The impulse response showing both early and late energy arrivals.

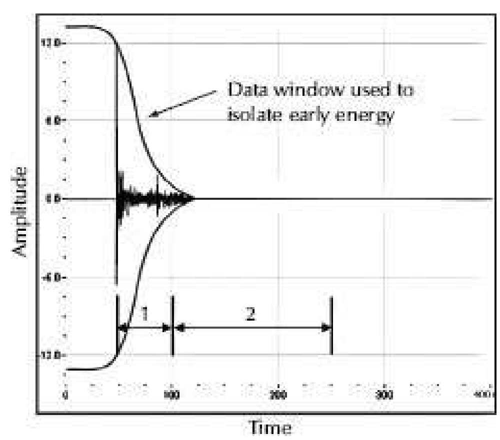


Figure 49-23. A data window is used to remove the effects of the later arrivals.

1. Determine the reason for and scope of the measurement session. What are you looking for? Can you hear it? Is it repeatable? Why do you need this information?
2. Determine what you are going to measure. Are you looking at the room or at the sound system? If it is the room, possibly the only meaningful measurements will be the overall decay time and the noise floor. If you are looking at the sound system, decide if you need to switch off or disconnect some loudspeakers. This may be essential to determine whether the individual components are working properly, or that an anomaly is the result of interaction between several components. “Divide and conquer” is the axiom.
3. Select the microphone position. I usually begin by looking at the on-axis response of the loudspeaker as measured from inside of critical distance. If multiple loudspeakers are on, turn all of them off but the one you are measuring. The microphone should be placed in the far free field of the loudspeaker as previously described. When measuring a loudspeaker’s response, care should be taken to eliminate the effects of early reflections on the measured data, as these will generate microphone position-dependent acoustic comb filters that can mask the true response of the loudspeaker. In most cases the predominant offending surface will be the floor or other boundaries near the microphone and loudspeaker. These reflections can be reduced or eliminated by using a ground plane microphone placement, a tall microphone stand (when the loudspeaker is overhead), or some strategically placed absorption. I prefer the tall microphone stand for measuring installed systems with seating present since it works most anywhere, regardless of the seating type. The idea is to intercept the sound on its way to a listener

position, but before it can interact with the physical boundaries around that position. These will always be unique to that particular seat, so it is better to look at the free field response, as it is the common denominator to many listener seats.

4. Begin with the big picture. Measure an impulse response of the complete decay of the space. This yields an idea of the overall properties of the room/system and provides a good point of reference for truncation by using smaller time windows. Save this information for documentation purposes, as later you may wish to reopen the file for further processing.
5. Reduce the size of the time window to eliminate room reflections. Remember that you are trading off frequency resolution when truncating the time record, Fig. 49-24. Be certain to maintain sufficient resolution to allow adequate low-frequency detail. In some cases, it may be impossible to maintain a sufficiently long window to view low frequencies and at the same time eliminate the effects of reflections at higher frequencies, Fig. 49-25. In such cases, the investigator may wish to use a short window for looking at the high-frequency direct field, but a longer window for evaluating the woofer. Windows appropriate for each part of the spectrum can be used. Some measurement systems provide variable time windows, which allow low frequencies to be viewed in great detail (long time window) while still providing a semi-anechoic view (short time window) at high frequencies. There is evidence to support that this is how humans process sound information, making this method particularly interesting, Fig. 49-26.
6. Are other microphone positions necessary to characterize this loudspeaker? The off-axis response of some loudspeakers is very similar to the on-axis response, reducing the need to measure at

many angles. Other loudspeakers have very erratic responses, and a measurement at any one point around the loudspeaker may bear little resemblance to the response at other positions. This is a design issue, but one that must be considered by the measurer.

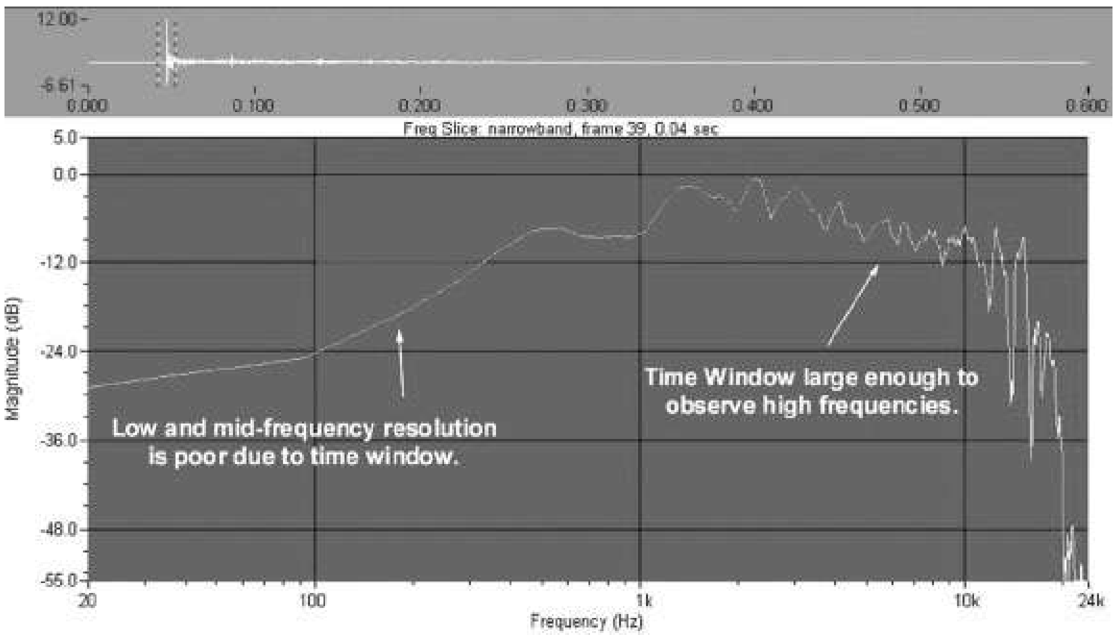


Figure 49-24. A short time window isolates the direct field at high frequencies at the expense of low-frequency resolution, SIA-SMAART.

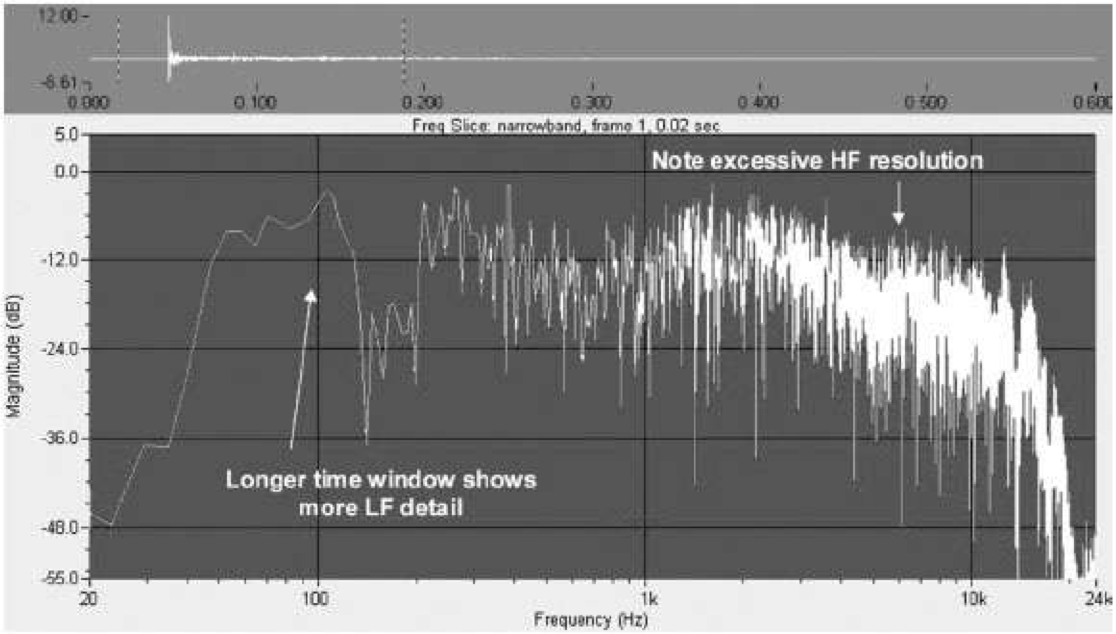


Figure 49-25. A long time window provides good low-frequency detail, SIA-SMAART.

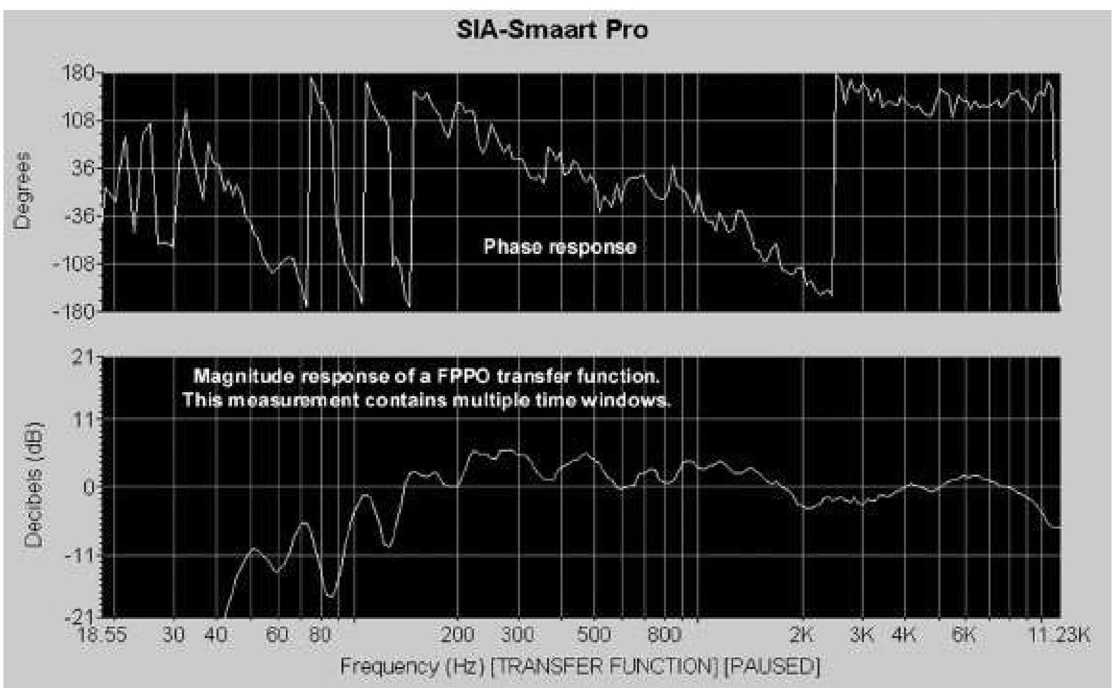


Figure 49-26. This time window increases with length as frequency decreases.

7. Once an accurate impulse response is measured, it can be post-processed to yield information on spectral content, speech intelligibility, and music clarity. There are a number of metrics that can provide this information. These are interpretations of the measured data and generally correlate with subjective perception of the sound at that seat.
8. An often overlooked method of evaluating the impulse response is the use of convolution to encode it onto anechoic program material. An excellent freeware convolver called GratisVolver is available from www.catt.se. Listening to the IR can often reveal subtleties missed by the various metrics, as well as provide clues as to what post-process must be used to observe the event of interest.

49.3.6 Human Perception

Useful measurement systems can measure the impulse response of a loudspeaker/room combination with great detail. Information regarding speech intelligibility and music clarity can be derived from the impulse response. In nearly all cases, this involves post-processing the impulse response using one of several clarity metrics.

49.3.6.1 Percentage Articulation Loss of Consonants (%Alcons)

For speech, one such metric is the percentage articulation loss of consonants, or *%Alcons*. Though not in widespread use today, a look at it can provide insight into the requirements for good speech intelligibility. A *%Alcons* measurement begins with an impulse response, which is usually displayed as a log-squared response or ETC. Since the calculation essentially examines the ratio between early energy, late energy, and noise, the measurer must place cursors on the display to define these parameters. These cursors may be placed automatically by the measurement program. The result is weighted with regard to decay time, so this too must be defined by the measurer. Analyzers such as the TEF25™ and EASERA include best guess default placements based on the research of Peutz, Davis, and others, Fig. 49-27.

These automatic placements were determined by correlating measured data with live listener scores in various acoustic environments, and represent a defined and orderly approach to achieving meaningful results that correlate with the perception of live listeners. The measurer is free to choose alternate cursor placements, but care must be taken to be consistent. Also, alternate cursor placements make it difficult if not impossible to compare your results with those obtained by other measurers. In the default %Alcons placement, the early energy (direct sound field) includes the first major sound arrival and any energy arrivals within the next 7-10ms. This forms a tight time span for the direct sound. Energy beyond this span is considered late energy and an impairment to communication. As one might guess, a later cursor placement yields better intelligibility scores, since more of the room response is being considered beneficial to intelligibility. As such, the default placement yields a worst-case scenario. The default placement considers the effects of the Early-Decay Time (EDT) rather than the classical T_{30} since short EDTs can yield good intelligibility, even in rooms with a long T_{30}. Again, the measurer is free to select an alternative cursor placement for determining the decay time used in the calculation, with the same caveats as placing the early-to-late dividing cursor. The *%Alcons* score is displayed instantly upon cursor placement and updates as the cursors are moved.

49.3.6.2 Speech Transmission Index (STI)

The STI can be calculated from the measured impulse response with a routine outlined by Schroeder and detailed by Becker in the reference. The STI is probably the most widely used contemporary measure of intelligibility. It is supported by virtually all measurement platforms, and some handheld analyzers are available for quick checks. In short, it is a number ranging from 0 to 1, with fair intelligibility centered at 0.5 on the scale. For more details on the Speech Transmission Index, and its variant, the Speech Transmission Index for Public Address Systems (STIPA), see Chapter 40, *Designing for Speech Intelligibility*.

49.3.7 Polarity

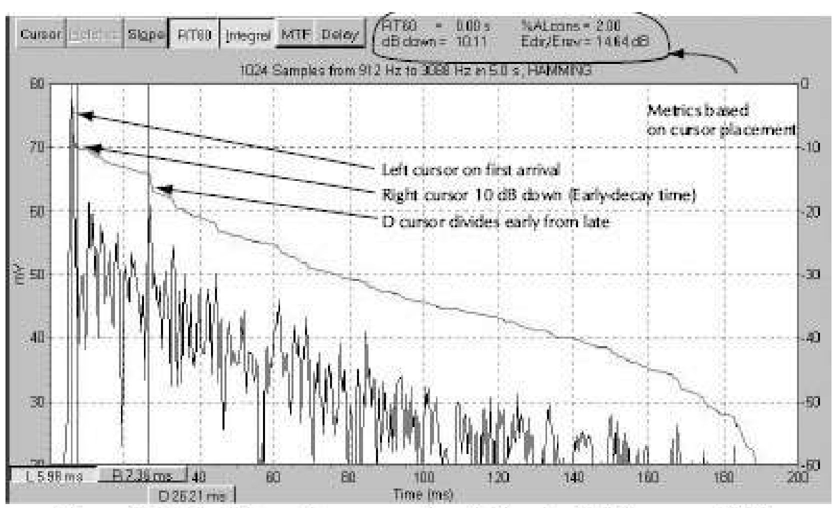


Figure 49-27. The ETC can be processed to yield an intelligibility score, TEF25.

Good sound system installation practice dictates maintaining proper signal polarity from system input to system output. An audio signal waveform always swings above and below some reference point. In acoustics, this reference point is the ambient atmospheric pressure. In an electronic device, the reference is the 0VA reference of the power supply (often called *signal ground*) in push-pull circuits or a fixed dc offset in class A circuits. Let's look at the acoustic situation first. An increase in the air pressure caused by a sound wave will produce an inward deflection of the diaphragm of a pressure microphone (the most common type) regardless of the microphone's orientation to the source. This inward deflection should cause a positive-going voltage swing at the output of the microphone on pin 2 relative to pin 3, as well as at the output of each piece of equipment that the signal passes through. Ultimately the electrical signal will be applied to a loudspeaker, which should deflect outward (toward an axial listener) on the positive-going signal, producing an increase in the ambient atmospheric pressure. Think of the microphone diaphragm and loudspeaker diaphragm moving in tandem and you will have the picture. Since most sound reinforcement equipment uses bipolar power supplies (allowing the audio signal to swing positive and negative about a zero reference), it is possible for signals to become inverted in polarity (flipped over). This causes a device to output a negative-going voltage when it is fed a positive-going voltage. If the loudspeaker is reverse-polarity from the microphone, an increase in sound pressure at the microphone (compression) will cause a decrease in pressure in front of the loudspeaker (rarefaction). Under some conditions, this can be extremely audible and destructive to sound quality. In other scenarios it can be irrelevant, but it is always good to check.

System installers should always check for proper polarity when installing the sound system. There are a number of methods, some simple and some complex. Let’s deal with them in order of complexity, starting with the simplest and least-costly method.

49.3.7.1 The Battery Test

Low-frequency loudspeakers can be tested using a standard 9V battery. The battery has a positive and negative terminal, and the spacing between the terminals is just about right to fit across the terminals of most woofers. The loudspeaker cone will move outward when the battery is placed across the loudspeaker terminals with the battery positive connected to the loudspeaker positive. While this is one of the most accurate methods for testing polarity, it doesn’t work for electronic devices or high-frequency drivers. Even so, it’s probably the least-costly and most accurate way to test a woofer.

49.3.7.2 Polarity Testers

There are a number of commercially available polarity test sets in the audio marketplace. The set includes a sending device that outputs a test pulse, Fig. 49-28, through a small loudspeaker (for testing microphones) or an XLR connector (for testing electronic devices) and a receiving device that collects the signal via an internal microphone (loudspeaker testing) or XLR input jack. A green light indicates correct polarity and a red light indicates reverse polarity. The receive unit should be placed at the system output (in front of the loudspeaker) while the send unit is systematically moved from device to device toward the system input. A polarity reversal will manifest itself by a red light on the receive unit.

Figure 49-28. A popular polarity test set.

49.3.7.3 Impulse Response Tests

The impulse response is perhaps the most fundamental of audio and acoustic measurements. The polarity of a loudspeaker or electronic device can be determined from observing its impulse response, Figs. 49-29 and 49-30. This is one of the few ways to test flown loudspeakers from a remote position. It is best to test the polarity of components of multi-way loudspeakers individually, since all of the individual components may not be polarized the same. Filters in the signal path (i.e., active crossover network) make the results more difficult to interpret, so it may be necessary to carefully test a system component (i.e., woofer) full-range for definitive results. Be sure to return the crossover to its proper setting before continuing.

Figure 49-29. The impulse response of a transducer with correct polarity.

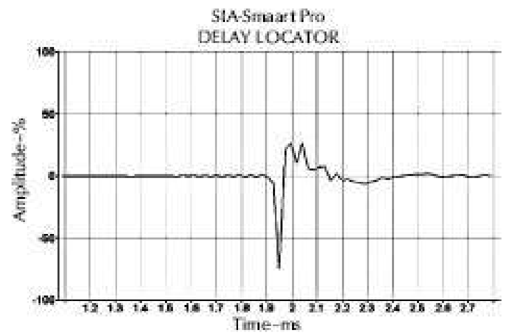


Figure 49-30. The impulse response of a reverse-polarity transducer.

49.4 Conclusion

The test and measurement of the sound reinforcement system are a vital part of the installation and diagnostic processes. The FFT and the analyzers that use it have revolutionized the measurement process, allowing sound practitioners to pick apart the system response and look at the response of the loudspeaker, room, or both. Powerful analyzers that were once beyond the reach of most technicians are readily available and affordable, and cost can no longer be used as an excuse for not measuring the system. The greatest investment by far is the time required to grasp the fundamentals of acoustics to allow interpretation of the data. Some of this information is general, and some of it is specific to certain measurement systems.

The acquisition of a measurement system is the first step in ascending the capability and credibility ladder. The next steps include acquiring proper instruction on its use by self-study or short course. The final and most important steps are the countless hours in the field required to correlate measured data with the hearing process. As proficiency in this area increases, the speed of execution, validity, and relevance of the measurements will increase also. While we can all learn how to make the measurements in a relatively short time span, the rest of our careers will be spent learning how to interpret what we are measuring.

Bibliography

D. Davis and C. Davis, *Sound System Engineering*, Focal Press, 1997, Boston, MA.

Farrel Becker, *A Do-It-Yourselfer's Guide to Computing the Speech Transmission Index*. SynAudCon Newsletter.

Joseph D'Appolitto, *Testing Loudspeakers*, Old Colony Sound Labs, Peterborough, NH.

Sam Berkow, *Understanding Sound System/Room Interactions*, Syn-Aud-Con Tech Topic, Volume 28, Number 2.

The Heyser Anthology, Audio Engineering Society, NY, NY.

Chapter 50

Fundamentals and Units of Measurement

by Glen Ballou

50.1 Units of Measurement

Measurements are the method we use to define all things in life. A dimension is any measurable extent such as length, thickness, or weight. A measurement system is any group of related unit names that state the quantity of properties for the items we see, taste, hear, smell, or touch.

A unit of measurement is the size of a quantity in the terms of which that quantity is measured or expressed, for instance, inches, miles, centimeters, and meters.

The laws of physics, which includes sound, are defined through dimensional equations that are defined from their units of measurements of mass, length, and time. For instance,

$$Area = L \times W \tag{50-1}$$

$$Velocity = \frac{D}{T} \tag{50-2}$$

where
L is length,
W is width,
D is distance,
T is time.

A physical quantity is specified by a number and a unit, for instance: 16ft or 5m.

50.1.1 SI System

The SI system (from the French Système International d'Unités) is the accepted international modernized metric system of measurement. It is used worldwide with the exception of a few countries including the United States of America.

The SI system has the following advantages:

1. Internationally accepted.
2. All values, except time, are decimal multiples or submultiples of the basic unit.
3. It is easy to use.
4. It is easy to teach.
5. It improves international trade and understanding.
6. It is coherent. All derived units are formed by multiplying and dividing other units without introducing any numerical conversion factor except one.
7. It is consistent. Each physical quantity has only one primary unit associated with it.

When using the SI system, exponents or symbol prefixes are commonly used. Table 50-1 is a chart of the accepted name of the number, its exponential form, symbol, and prefix name. (Note because of their size, the numbers from sextillion to centillion have not been shown in numerical form and symbols and prefix names have not been established for these numbers.)

Table 50-1. Multiple and Submultiple Prefixes

Name of Number	Number	Exponential Form	Symbol	Prefix
Centillion		1.0×10^{303}		
Googol		1.0×10^{100}		
Vigintillion		1.0×10^{63}		
Novemdecillion		1.0×10^{60}		
Octodecillion		1.0×10^{57}		
Septendecillion		1.0×10^{54}		
Sexdecillion		1.0×10^{51}		
Quindecillion		1.0×10^{48}		
Quattuordecillion		1.0×10^{45}		
Tredecillion		1.0×10^{42}		
Duodecillion		1.0×10^{39}		
Undecillion		1.0×10^{36}		
Decillion		1.0×10^{33}		
Nonillion		1.0×10^{30}		
Octillion		1.0×10^{27}		
Septillion		1.0×10^{24}	E	Exa-
Sextillion		1.0×10^{21}	P	Peta-
Trillion	1 000 000 000 000	1.0×10^{12}	T	Tera-
Billion	1 000 000 000	1.0×10^{9}	G	Giga-
Million	1 000 000	1.0×10^{6}	M	Mega-
Thousand	1000	1.0×10^{3}	k	Kilo-
Hundred	100	1.0×10^{2}	h	Hecto-
Ten	10	1.0×10^{1}	da	Deka-
Unit	1	1.0×10^{0}	—	—
Tenth	0.10	1.0×10^{-1}	d	Deci-
Hundredth	0.01	1.0×10^{-2}	c	Centi-
Thousandth	0.00	1.0×10^{-3}	m	Milli-
Millionth	0.000 001	1.0×10^{-6}	µ	Micro-
Billionth	0.000 000 001	1.0×10^{-9}	n	Nano-
Trillionth	0.000 000 000 001	1.0×10^{-12}	p	Pico-
Quadrillionth	0.000 000 000 000 001	1.0×10^{-15}	f	Femto-

50.1.2 Fundamental Quantities

There are seven fundamental quantities in physics: length, mass, time, intensity of electric current, temperature, luminous intensity, and molecular substance. Two supplementary quantities are plane angle and solid angle.

50.1.3 Derived Quantities

Derived quantities are those defined in terms of the seven fundamental quantities, for instance, speed = length/time. There are sixteen derived quantities with names of their own: energy (work, quantity of heat), force, pressure, power, electric charge, electric potential difference (voltage), electric resistance, electric conductance, electric capacitance, electric inductance, frequency, magnetic flux, magnetic flux density, luminous flux, illuminance, and customary temperature. Following are thirteen additional derived quantities that carry the units of the original units that are combined. They are area, volume, density, velocity, acceleration, angular velocity, angular acceleration, kinematic viscosity, dynamic viscosity, electric field strength, magnetomotive force, magnetic field strength, and luminance.

50.1.4 Definition of the Quantities

The quantities will be defined in SI units, and their US customary unit equivalent values will also be given.

Length (*L*). *Length* is the measure of how long something is from end to end. The meter (abbreviated m) is the SI unit of length. (Note: in the United States the spelling "meter" is retained, while most other countries use the spelling "metre.") The meter is the 1 650 763.73 wavelengths, in vacuum, of the radiation corresponding to the unperturbed transition between energy level $2P_{10}$ and $5D_5$ of the krypton-86 atom. The result is an orange-red line with a wavelength of 6057.802×10^{-10} meters. The meter is equivalent to 39.370 079 inches.

Mass (*M*). *Mass* is the measure of the inertia of a particle. The mass of a body is defined by the equation

$$M = \left(\frac{A_s}{a}\right)M_s \qquad (50\text{-}3)$$

where,

A_S is the acceleration of the standard mass M_S,

a is the acceleration of the unknown mass, M, when the two bodies interact.

The kilogram (kg) is the unit of mass. This is the only base or derived unit in the SI system that contains a prefix. Multiples are formed by attaching prefixes to the word gram. Small masses may be described in grams (g) or milligrams (mg) and large masses in megagrams. Note the term *tonnes* is sometimes used for the metric ton or megagram, but this term is not recommended.

The present international definition of the kilogram is the mass of a special cylinder of platinum iridium alloy maintained at the International Bureau of Weights and Measures, Sevres, France. One kilogram is equal to 2.204 622 6 avoirdupois pounds (lb). A liter of pure water at standard temperature and pressure has a mass of 1kg ± one part in 10^4.

Mass of a body is often revealed by its weight, which the gravitational attraction of the earth gives to that body.

If a mass is weighed on the moon, its mass would be the same as on earth, but its weight would be less due to the small amount of gravity.

$$M = \frac{W}{g} \qquad (50\text{-}4)$$

where,

W is the weight,
g is the acceleration due to gravity.

Time (*t*). *Time* is the period between two events or the point or period during which something exists, happens, etc.

The second (*s*) is the unit of time. Time is the one dimension that does not have powers of ten multipliers in the SI system. Short periods of time can be described in milliseconds (ms) and microseconds (μs). Longer periods of time are expressed in minutes (1min = 60s) and hours (1h = 3600s). Still longer periods of time are the day, week, month, and year. The present international definition of the second is the time duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the atom of Caesium 133. It is also defined as 1/86, 400 of the mean solar day.

Current (*I*). *Current* is the rate of flow of electrons. The ampere (*A*) is the unit of measure for current. Small currents are measured in milliamperes (mA) and microamperes (μA), and large currents are in kiloamperes (kA). The international definition of the ampere is the constant current that, if maintained in two straight parallel conductors of infinite length and negligible cross-sectional area and placed exactly 1 meter apart in a vacuum, will produce between them a force of 2×10^{-7}N/m^2 of length.

A simple definition of one ampere of current is the intensity of current flow through a 1 ohm resistance under a pressure of 1 volt of potential difference.

Temperature (*T*). *Temperature* is the degree of hotness or coldness of anything. The kelvin (K) is the unit of temperature. The kelvin is 1/273.16 of the thermodynamic temperature of the triple point of pure water.

Note: the term degree (°) is not used with the term kelvin as it is with other temperature scales.

Ordinary temperature measurements are made with the celsius scale on which water freezes at 0°C and boils at 100°C. A change of 1°C is equal to a change of 1 kelvin, therefore 0°C = 273.15 K: 0°C = 32°F.

Luminous Intensity (*I_L*). *Luminous intensity* is the luminous flux emitted per unit solid angle by a point source in a given direction. The candela (cd) is the unit of luminous intensity. One candela will produce a luminous flux of 1 lumen (lm) within a solid angle of 1 steradian (sr).

The international definition of the candela is the luminous intensity, perpendicular to the surface, of 1/600 000m^2 of a black body at the temperature of freezing platinum under a pressure of 101 325N/m^2 (pascals).

Molecular Substance (*n*). *Molecular substance* is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012kg of carbon 12.

The mole is the unit of molecular substance. One mole of any substance is the gram molecular weight of the material. For example, 1 mole (mol) of water (H_2O) weighs 18.016g.

$$
\begin{aligned}
H_2 &= 2 \text{ atoms} \times 1.008 \text{ atomic weight} \\
O &= 2 \text{ atoms} \times 16 \text{ atomic weight} \\
H_2O &= 18.016\,g
\end{aligned}
$$

Plane Angle (α). The *plane angle* is formed between two straight lines or surfaces that meet. The radian (rad) is the unit of plane angles. One radian is the angle formed between two radii of a circle and subtended by an arc whose length is equal to the radius. There are 2π radians in 360°.

Ordinary measurements are still made in degrees. The degree can be divided into minutes and seconds or into tenths and hundredths of a degree. For small angles, the latter is most useful.

$$\text{One degree of arc } (1^\circ) = \frac{\pi}{180}\text{Rad} \qquad (50\text{-}5)$$

$$1\,\text{Rad} = 57.2956^\circ$$

Solid Angle (A). (50-6)

A *solid angle* subtends three dimensions. The solid angle is measured by the area, subtended (by projection) on a sphere of unit radius by the ratio of the area A, intercepted on a sphere of radius r to the square of the radius (A/r^2).

The steradian (*sr*) is the unit of solid angle. The steradian is the solid angle at the center of a sphere that subtends an area on the spherical surface, which is equal to that of a square whose sides are equal to the radius of the sphere.

Energy (E). *Energy* is the property of a system that is a measure of its ability to do work. There are two main forms of energy—potential energy and kinetic energy.

1. Potential energy (U) is the energy possessed by a body or system by virtue of position and is equal to the work done in changing the system from some standard configuration to its present state. Potential energy is calculated with the equation

$$U = Mgh \qquad (50\text{-}7)$$

where,
M is the mass,
g is the acceleration due to gravity.
h is the height.

For example, a mass *M* placed at a height h above a datum level in a gravitational field with an acceleration of free fall (*g*), has a potential energy given by $U = mgh$. This potential energy is converted into kinetic energy when the body falls between the levels.

2. Kinetic energy (*T*) is the energy possessed by virtue of motion and is equal to the work that would be required to bring the body to rest. A body undergoing translational motion with velocity, *v*, has a kinetic energy given by

$$T = 0.5Mv^2 \qquad (50\text{-}8)$$

where,
M is the mass of the body,
v is the velocity of the body.

For a body undergoing rotational motion

$$T = 0.51I\omega^2 \qquad (50\text{-}9)$$

where,
I is the moment of inertia of the body about its axis of rotation,
ω is the angular velocity.

The joule (J) is the unit of energy. The mechanical definition is the work done when the force of one newton is applied for a distance of 1m in the direction of its application, or 1Nm. The electrical unit of energy is the kilowatt-hour (kWh), which is equal to 3.6×10^6J.

In physics, the unit of energy is the electron volt (eV), which is equal to $(1.602\ 10 \pm 0.000\ 07) \times 10^{-19}$J.

Force (*F*). *Force* is any action that changes, or tends to change, a body's state of rest or uniform motion in a straight line.

The newton (*N*) is the unit of force and is that force which, when applied to a body having a mass of 1 kg, gives it an acceleration of 1m/s^2. One newton equals 1J/m, 1kg(m)/s^2, 10^5dynes, and 0.224 809lb force.

Pressure. *Pressure* is the force (in a fluid) exerted per unit area on an infinitesimal plane situated at the point. In a fluid at rest, the pressure at any point is the same in all directions. A fluid is any material substance which in static equilibrium cannot exert tangential force across a surface but can exert only pressure. Liquids and gases are fluids.

The pascal (Pa) is the unit of pressure. The pascal is equal to the newton per square meter (N/m^2).

$$\begin{aligned} 1Pa &= 10^{-6}\,\text{bars} \\ &= 1.45038 \times 10^{-4}\ \text{lb/in}^2 \end{aligned} \tag{50-10}$$

Power (*W*). *Power* is the rate at which energy is expended or work is done. The watt (W) is the unit of power and is the power that generates energy at the rate of 1 J/s.

$$
\begin{aligned}
1\text{W} &= 1\text{J/s} \\
&= 3.141442\ \text{BTU/h} \\
&= 44.2537\ \text{ft-lb/min} \\
&= 0.00134102\ \text{hp}
\end{aligned}
\qquad (50\text{-}11)
$$

Electric Charge (*Q*). *Electric charge* is the quantity of electricity or electrons that flows past a point in a period of time. The coulomb (C) is the unit of electric charge and is the quantity of electricity moved in 1 second by a current of 1 ampere. The coulomb is also defined as 6.24196×10^{18} electronic charges.

Electric Potential Difference (*V*). Often called electromotive force (*emf*) and voltage (*V*), *electric potential difference* is the line integral of the electric field strength between two points. The volt (V) is the unit of electric potential. The volt is the potential difference that will cause a current flow of 1A between two points in a circuit when the power dissipated between those two points is 1W.

A simpler definition would be to say a potential difference of 1V will drive a current of 1A through a resistance of 1Ω.

$$
\begin{aligned}
Volt\ (V) &= \frac{W}{A} \\
&= \frac{J}{A(s)} \\
&= \frac{\text{kg}(\text{m}^2)}{s^3 A} \\
&= \text{A}\Omega
\end{aligned}
\qquad (50\text{-}12)
$$

Electric Resistance (*R*). *Electric resistance* is the property of conductors that, depending in their dimensions, material, and temperature, determines the current produced by a given difference of potential. It is also that property of a substance that impedes current and results in the dissipation of power in the form of heat.

The ohm (Ω) is the unit of resistance and is the resistance that will limit the current flow to 1A when a potential difference of 1 V is applied to it.

$$\begin{aligned} R &= \frac{V}{A} \\ &= \frac{kg(m^2)}{s^3 A^3} \end{aligned} \qquad (50\text{-}13)$$

Electric Conductance (*G*). *Electric conductance* is the reciprocal of resistance. The siemens (S) is the unit of electric conductance. A passive device that has a conductance of 1S will allow a current flow of 1A when 1V potential is applied to it.

$$\begin{aligned} S &= \frac{1}{\Omega} \\ &= \frac{A}{V} \end{aligned} \qquad (50\text{-}14)$$

Electric Capacitance (*C*). *Electric capacitance* is the property of an isolated conductor or set of conductors and insulators to store electric charge. The farad (F) is the unit of electric capacitance and is defined as the capacitance that exhibits a potential difference of 1V when it holds a charge of 1C.

$$\begin{aligned} \mathrm{F} &= \frac{C}{V} \\ &= \frac{AS}{V} \end{aligned} \qquad (50\text{-}15)$$

where,
C is the electric charge in C,
V is the electric potential difference in V,
A is the current in A,

S is the conductance in S.

Electric Inductance (L). *Electric inductance* is the property that opposes any change in the existing current. Inductance is only present when the current is changing. The henry (H) is the unit of inductance and is the inductance of a circuit in which an electromotive force of 1V is developed by a current change of 1A/s.

$$H = \frac{Vs}{A} \qquad (50\text{-}16)$$

Frequency (f). *Frequency* is the number of recurrences of a periodic phenomenon in a unit of time. The hertz (Hz) is the unit of frequency and is equal to one cycle per second, 1Hz = 1cps. Frequency is often measured in hertz (Hz), kilohertz (kHz), and megahertz (MHz).

Sound Intensity (W/m^2). *Sound intensity* is the rate of flow of sound energy through a unit area normal to the direction of flow. For a sinusoidally varying sound wave the intensity I is related to the sound pressure p and the density ß of the medium by

$$I = \frac{p^2}{\beta c} \qquad (50\text{-}17)$$

where,
c is the velocity of sound.

The watt per square meter (W/m^2) is the unit of sound intensity.

Magnetic Flux (ϕ). *Magnetic flux* is a measure of the total size of a magnetic field. The weber (Wb) is the unit of magnetic flux, and is the amount of flux that produces an electromotive force of 1V in a one-turn conductor as it reduces uniformly to zero in 1s.

$$\begin{aligned} Wb &= W(s) \\ &= 10^8 \text{ lines of flux} \\ &= \frac{kg(m^2)}{s^2 A} \end{aligned} \qquad (50\text{-}18)$$

Magnetic Flux Density (ß). The *magnetic flux density* is the flux passing through the unit area of a magnetic field in the direction at right angles to the magnetic force. The vector product of the magnetic flux density and the current in a conductor gives the force per unit length of the conductor.

The tesla (T) is the unit of magnetic flux density and is defined as a density of 1 Wb/m^2.

$$\begin{aligned} T &= \frac{Wb}{m^2} \\ &= \frac{V(s)}{m^2} \\ &= \frac{kg}{s^2 A} \end{aligned} \qquad (50\text{-}19)$$

Luminous Flux (Φ_v). *Luminous flux* is the rate of flow of radiant energy as evaluated by the luminous sensation that it produces. The lumen (lm) is the unit of luminous flux, which is the amount of luminous flux emitted by a uniform point source whose intensity is 1 steradian.

$$\begin{aligned} lm &= cd\left(\frac{sr}{m^2}\right) \\ &= 0.0795774 \text{ candlepower} \end{aligned} \qquad (50\text{-}20)$$

where,

cd is the luminous intensity in cd,
sr is the solid angle in sr.

Luminous flux Density (E_v). The *luminous flux density* is the luminous flux incident on a given surface per unit area. It is sometimes called illumination or intensity of illumination. At any point on a surface, the illumination is given by

$$E_v = \frac{d\Phi_v}{dA} \qquad (50\text{-}21)$$

The lux (lx) is the unit of luminous flux density, which is the density of radiant flux of lm/m^2,

$$\begin{aligned} lx &= \frac{lm}{m^2} \\ &= cd\,\frac{sr}{m^2} \qquad (50\text{-}22) \\ &= 0.0929030 \text{ fc} \end{aligned}$$

Displacement. *Displacement* is a change in position or the distance moved by a given particle of a system from its position of rest, when acted on by a disturbing force.

Speed/Velocity. *Speed* is the rate of increase of distance traveling by a body. Average speed is found by the equation

$$S = \frac{l}{t} \qquad (50\text{-}23)$$

where,
S is the speed,
l is the length or distance,
t is the time to travel.

Speed is a scalar quantity as it is not referenced to direction. Instantaneous speed = dl/dt. *Velocity* is the rate of increase of distance traversed by a body in a particular direction.

Velocity is a vector quantity as both speed and direction are indicated. The l/t can often be the same for the velocity and speed of an object, however, when speed is given, the direction of movement is not known. If a body describes a circular path and each successive equal distances along the path is described in equal times, the speed would be constant but the velocity would constantly change due to the change in direction.

Weight. *Weight* is the force exerted on a mass by the gravitational pull of the planet, star, moon, etc., that the mass is near. The weight experienced on earth is due to the earth's gravitational pull, which is 9.806 65m/s^2, and causes an object to accelerate toward earth at a rate of 9.806 65m/s^2 or 32ft/s^2.

The weight of a mass M is M(g). If M is in kg and g in m/s^2, the weight would be in newtons (N). Weight in the US system is in pounds (lb).

Acceleration. *Acceleration* is the rate of change in velocity or the rate of increase or decrease in velocity with time. Acceleration is expressed in meters per second squared (m/s^2), or ft/s^2 in the US system.

Amplitude. *Amplitude* is the magnitude of variation in a changing quantity from its zero value. Amplitude should always be modified with adjectives such as peak, rms, maximum, instantaneous, etc.

Wavelength (*M*). In a periodic wave, the distance between two points of the corresponding phase of two consecutive cycles is the *wavelength*. Wavelength is related to the velocity of propagation (c) and frequency (f) by the equation

$$\lambda = \frac{c}{f} \qquad (50\text{-}24)$$

The wavelength of a wave traveling in air at sea level and standard temperature and pressure (STP) is

$$\lambda = \frac{331.4\ \text{m/s}}{f} \qquad (50\text{-}25)$$

or

$$\lambda = \frac{1087.42\ \text{ft/s}}{f} \qquad (50\text{-}26)$$

For instance, the length of a 1000Hz wave would be 0.33m, or 1.09ft.

Phase. *Phase* is the fraction of the whole period that has elapsed, measured from a fixed datum. A sinusoidal quantity may be expressed as a rotating vector *OA*. When rotated a full 360°, it represents a sine wave. At any position around the circle, *OX* is equal in length but said to be *X* degrees out of phase with *OA*.

It may also be stated that the phase difference between *OA* and *OX* is α. When particles in periodic motion due to the passage of a wave are moving in the same direction with the same relative displacement, they are said to be in phase. Particles in a wave front are in the same phase of vibration when the distance between consecutive wave fronts is equal to the wavelength. The phase difference of two particles at distances X_1 and X_2 is

$$\alpha = \frac{2\pi(X_2 - X_1)}{\lambda} \quad (50\text{-}27)$$

Periodic waves, having the same frequency and waveform, are said to be in phase if they reach corresponding amplitudes simultaneously.

Phase Angle. The angle between two vectors representing two periodic functions that have the same frequency is the *phase angle*. Phase angle can also be considered the difference, in degrees, between corresponding stages of the progress of two cycle operations.

Phase Difference (ϕ). *Phase difference* is the difference in electrical degrees or time, between two waves having the same frequency and referenced to the same point in time.

Phase Shift. Any change that occurs in the phase of one quantity or in the phase difference between two or more quantities is the *phase shift*.

Phase Velocity. The *phase velocity* is when a point of constant phase is propagated in a progressive sinusoidal wave.

Temperature. *Temperature* is the measure of the amount of coldness or hotness. While kelvin is the SI standard, temperature is commonly referenced to °C (degrees Celsius) or °F (degrees Fahrenheit).

The lower fixed point (the ice point) is the temperature of a mixture of pure ice and water exposed to the air at standard atmospheric pressure.

The upper fixed point (the steam point) is the temperature of steam from pure water boiling at standard atmospheric pressure.

In the Celsius scale, named after Anders Celsius (1701–1744) and originally called Centigrade, the fixed points are 0°C and 100°C. This scale is used in the SI system.

The Fahrenheit scale, named after Gabriel Daniel Fahrenheit in 1714, has the fixed points at 32°F and 212°F.

To interchange between °C and °F, use the following equations.

$$
\begin{aligned}
°C &= (°F - 32°) \times \frac{5}{9} \\
°F &= \left(°C \times \frac{9}{5}\right) + 32°
\end{aligned}
\qquad (50\text{-}28)
$$

The absolute temperature scale operates from absolute zero of temperature. Absolute zero is the point where a body cannot be further cooled because all the available thermal energy is extracted.

Absolute zero is 0 kelvin (0K) or 0° Rankine (0°R). The Kelvin scale, named after Lord Kelvin (1850), is the standard in the SI system and is related to °C.

0°C = 273.15K

The Rankine scale is related to the Fahrenheit system.

32°F = 459.67°R

The velocity of sound is affected by temperature. As the temperature increases, the velocity increases. The approximate formula is

$$C = 331.4 \text{ m/s} + 0.607\,T \text{ SI units} \qquad (50\text{-}29)$$

where,
T is the temperature in °C.

$$C = 1052 \text{ ft/s} \times 1.106\, T \quad \text{US units} \qquad (50\text{-}30)$$

where,
T is the temperature in °F.

Another simpler equation to determine the velocity of sound is

$$C = 49.00\sqrt{459.69° + °F} \qquad (50\text{-}31)$$

Things that can affect the speed of sound are the sound wave going through a temperature barrier or going through a stream of air such as from an air conditioner. In either case, the wave is deflected the same way that light is refracted in glass.

Pressure and altitude do not affect the speed of sound because at sea level the molecules bombard each other, slowing down their speed. At upper altitudes they are farther apart so they do not bombard each other as often so they reach their destination at the same time.

Thevenin's Theorem. *Thevenin's Theorem* is a method used for reducing complicated networks to a simple circuit consisting of a voltage source and a series impedance. The theorem is applicable to both ac and dc circuits under steady-state conditions.

The theorem states: the current in a terminating impedance connected to any network is the same as if the network were replaced by a generator with a voltage equal to the open-circuit voltage of the network, and whose impedance is the impedance seen by the termination looking back into the network. All generators in the network are replaced with impedance equal to the internal impedances of the generators.

Kirchhoff's Laws. The laws of Kirchhoff can be used for both dc and ac circuits. When used in ac analysis, phase must also be taken into consideration.

Kirchhoff's Voltage Law (KVL). *Kirchhoff's voltage law* states that the sum of the branch voltages for any closed loop is zero at any time. Stated another way, for any closed loop, the sum of the voltage drops equal the sum of the voltage rises at any time.

In the laws of Kirchhoff, individual electric circuit elements are connected according to some wiring plan or schematic. In any closed loop, the voltage drops must be equal to the voltage rises. For example, in the dc circuit of Fig. 50-1, V_1 is the voltage source or rise such as a battery and V_2, V_3, V_4, and V_5 are voltage drops (possibly across resistors) so

$$V_1 = V_2 + V_3 + V_4 + V_5 \qquad (50\text{-}32)$$

or,

$$V_1 - V_2 - V_3 - V_4 - V_5 = 0 \qquad (50\text{-}33)$$

In an ac circuit, phase must be taken into consideration, therefore, the voltage would be

$$V_1 e^{j\omega t} - V_2 e^{j\omega t} - V_3 e^{j\omega t} - V_4 e^{j\omega t} - V_5 e^{j\omega t} = 0 \qquad (50\text{-}34)$$

where,
e^{jwt} is $\cos AT + j\sin AT$ or Euler's identity.

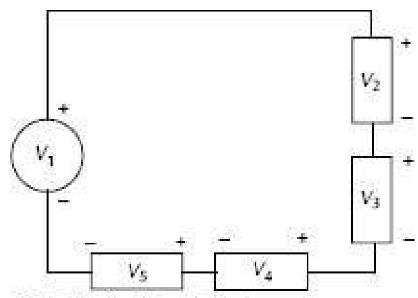


Figure 50-1. Kirchhoff's voltage law.

Kirchhoff's Current Law (KCL). *Kirchhoff's current law* states that the sum of the branch currents leaving any node must equal the sum of the branch currents entering that node at any time.

Stated another way, the sum of all branch currents incident at any node is zero.

In Fig. 50-2 the connection on node current in a dc circuit is equal to 0 and is equal to the sum of currents

$$I_1 = I_2 + I_3 + I_4 + I_5 \qquad (50\text{-}35)$$

or

$$I_1 - I_2 - I_3 - I_4 - I_5 = 0 \qquad (50\text{-}36)$$

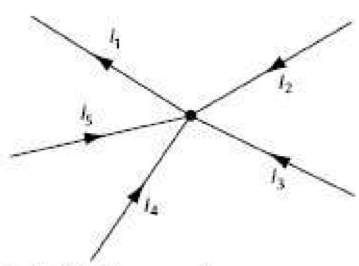


Figure 50-2. Kirchhoff's current law.

The current throughout the circuit is also a function of the current from the power source (V_1) and the current through all of the branch circuits.

In an ac circuit, phase must be taken into consideration, therefore, the current would be

$$I_1e^{j\omega t} - I_2e^{j\omega t} - I_3e^{j\omega t} - I_4e^{j\omega t} - I_5e^{j\omega t} = 0 \qquad (50\text{-}37)$$

where,
e^{jwt} is $\cos AT + j\sin AT$ or Euler's identity.

Ohm's Law. *Ohm's Law* states that the ratio of applied voltage to the resultant current is a constant at every instant and that this ratio is defined to be the resistance.

If the voltage is expressed in volts and the current in amperes, the resistance is expressed in ohms. In equation form it is

$$R = \frac{V}{I} \qquad (50\text{-}38)$$

or,

$$R = \frac{e}{i} \qquad (50\text{-}39)$$

where,
e and *i* are instantaneous voltage and current,
V and *I* are constant voltage and current,
R is the resistance.

Through the use of Ohm's Law, the relationship between voltage, current, resistance or impedance, and power can be calculated.

Power is the rate of doing work and can be expressed in terms of potential difference between two points (voltage) and the rate of flow required to transform the potential energy from one point to the other (current). If the voltage is in volts or *J/C* and the current is in amperes or *C/s*, the product is joules per second or watts:

$$P = VI \qquad (50\text{-}40)$$

or

$$\frac{J}{s} = \frac{J}{C}\left(\frac{C}{s}\right) \qquad (50\text{-}41)$$

where,
J is energy in J,
C is electric charge in C.

Fig. 50-3 is a wheel chart that relates current, voltage, resistance or impedance, and power. The power factor (PF) is cos *I* where *I* is the phase angle between *e* and *i*. A power factor is required in ac circuits.

50.2 Radio Frequency Spectrum

The *radio frequency spectrum* of 30Hz to 3,000,000MHz (3 × 10^{12}Hz) is divided into the various bands shown in Table 50-2.

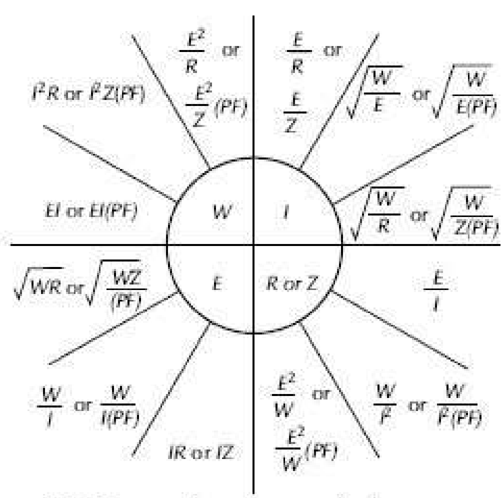


Figure 50-3. Power, voltage, current wheel.

Table 50-2. Frequency Classification

Frequency	Band No.	Classification	Abbreviation
30-300Hz	2	extremely low frequencies	ELF
300-3000Hz	3	voice frequencies	VF
3-30kHz	4	very low frequencies	VLF
30-300kHz	5	low frequencies	LF
300-3000kHz	6	medium frequencies	MF
3-30MHz	7	high frequencies	HF
30-300MHz	8	very high frequencies	VHF
30-3000MHz	9	ultrahigh frequencies	UHF
3-30GHz	10	super-high frequencies	SHF
30-300GHz	11	extremely high frequencies	EHF
300-3THz	12	-	-

50.3 Decibel (dB)

Decibels are a logarithmic ratio of two numbers. The decibel is derived from two power levels and is also used to show voltage ratios indirectly (by relating voltage to power). The equations or decibels are

$$Power\ dB = 10\log\frac{P_1}{P_2} \tag{50-42}$$

$$Voltage\ dB_v = 20\log\frac{E_1}{E_2} \tag{50-43}$$

Fig. 50-4 shows the relationship between the power, decibels, and voltage. In the illustration, “dBm” is the decibels referenced to 1mW.

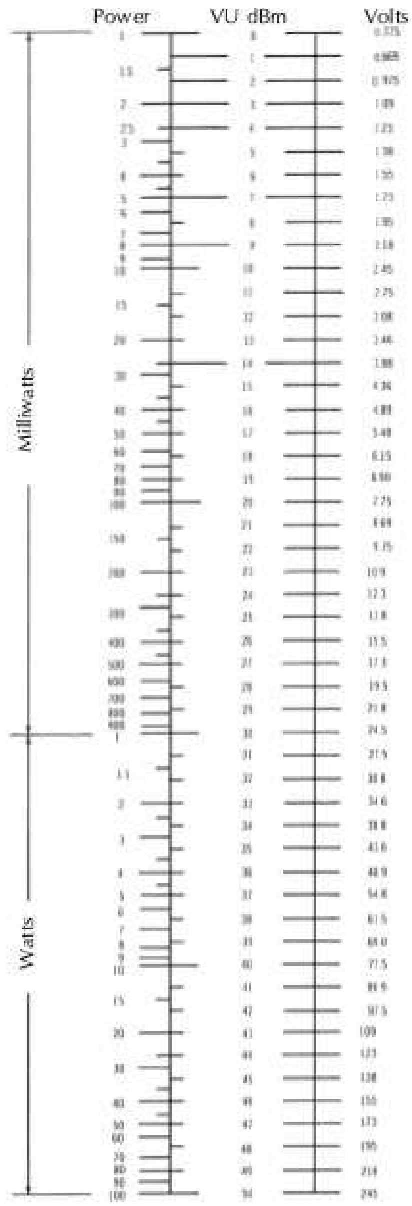


Figure 50-4. Relationship between power, dBm, and voltage.

Table 50-3 shows the relationship between decibel, current, voltage, and power ratios.

Volume unit (VU) meters measure decibels that are related to a 600Ω impedance, O VU is actually +4dBm (see Chapter 30, *Audio Output Meters and Devices*).

Table 50-3. Relationships between Decibel, Current, Voltage, and

Power Ratios

dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power
0.0	**1.0000**	**1.000**	**0.0**	**5.0**	**0.5623**	**1.778**	**0.50**	**10.0**	**0.3162**	**3.162**	**5.00**	**15.0**	**0.1778**	**5.623**	**0.50**
0.1	0.9886	1.012	0.05	0.1	0.5559	1.799	0.55	0.1	0.3126	3.199	0.05	0.1	0.1758	5.689	0.55
0.2	0.9772	1.023	0.10	0.2	0.5495	1.820	0.60	0.2	0.3090	3.236	0.10	0.2	0.1738	5.754	0.60
0.3	0.9661	1.035	0.15	0.3	0.5433	1.841	0.65	0.3	0.3055	3.273	0.15	0.3	0.1718	5.821	0.65
0.4	0.9550	1.047	0.20	0.4	0.5370	1.862	0.70	0.4	0.3020	3.311	0.20	0.4	0.1698	5.888	0.70
0.5	0.9441	1.059	0.25	0.5	0.5309	1.884	0.75	0.5	0.2985	3.350	0.25	0.5	0.1679	5.957	0.75
0.6	0.9333	1.072	0.30	0.6	0.5248	1.905	0.80	0.6	0.2951	3.388	0.30	0.6	0.1660	6.026	0.80
0.7	0.9226	1.084	0.35	0.7	0.5188	1.928	0.85	0.7	0.2917	3.428	0.35	0.7	0.1641	6.095	0.85
0.8	0.9120	1.096	0.40	0.8	0.5129	1.950	0.90	0.8	0.2884	3.467	0.40	0.8	0.1622	6.166	0.90
0.9	0.9016	1.109	0.45	0.9	0.5070	1.972	0.95	0.9	0.2851	3.508	0.45	0.9	0.1603	6.237	0.95
1.0	**0.8913**	**1.122**	**0.50**	**6.0**	**0.5012**	**1.995**	**3.00**	**11.0**	**0.2818**	**3.548**	**0.50**	**16.0**	**0.1585**	**6.310**	**8.00**
0.1	0.8810	1.135	0.55	0.1	0.4955	2.018	0.05	01	0.2786	3.589	0.55	01	0.1567	6.383	0.05
0.2	0.8710	1.148	0.60	0.2	0.4898	2.042	0.10	02	0.2754	3.631	0.60	02	0.1549	6.457	0.10
0.3	0.8610	1.161	0.65	0.3	0.4842	2.065	0.15	03	0.2723	3.673	0.65	03	0.1531	6.531	0.15
0.4	0.8511	1.175	0.70	0.4	0.4786	2.089	0.20	04	0.2692	3.715	0.70	04	0.1514	6.607	0.20
0.5	0.8414	1.189	0.75	0.5	0.4732	2.113	0.25	05	0.2661	3.758	0.75	05	0.1496	6.683	0.25
0.6	0.8318	1.202	0.80	0.6	0.4677	2.138	0.30	06	0.2630	3.802	0.80	06	0.1479	6.761	0.30
0.7	0.8222	1.216	0.85	0.7	0.4624	2.163	0.35	07	0.2600	3.846	0.85	07	0.1462	6.839	0.35
0.8	0.8128	1.230	0.90	0.8	0.4571	2.188	0.40	08	0.2570	3.890	0.90	08	0.1445	6.918	0.40
0.9	0.8035	1.245	0.95	0.9	0.4519	2.213	0.45	09	0.2541	3.936	0.95	09	0.1429	6.998	0.45
2.0	**0.7943**	**1.259**	**1.00**	**7.0**	**0.4467**	**2.239**	**0.50**	**12.0**	**0.2512**	**3.981**	**6.00**	**17.0**	**0.1413**	**7.079**	**0.50**
0.1	0.7852	1.274	0.05	0.1	0.4416	2.265	0.55	0.1	0.2483	4.027	0.05	0.1	0.1396	7.161	0.55
0.2	0.7762	1.288	0.10	0.2	0.4365	2.291	0.60	0.2	0.2455	4.074	0.10	0.2	0.1380	7.244	0.60
0.3	0.7674	1.303	0.15	0.3	0.4315	2.317	0.65	0.3	0.2427	4.121	0.15	0.3	0.1365	7.328	0.65
0.4	0.7586	1.318	0.20	0.4	0.4266	2.344	0.70	0.4	0.2399	4.169	0.20	0.4	0.1349	7.413	0.70
0.5	0.7499	1.334	0.25	0.5	0.4217	2.371	0.75	0.5	0.2371	4.217	0.25	0.5	0.1334	7.499	0.75
0.6	0.7413	1.349	0.30	0.6	0.4169	2.399	0.80	0.6	0.2344	4.266	0.30	0.6	0.1318	7.586	0.80
0.7	0.7328	1.365	0.35	0.7	0.4121	2.427	0.85	0.7	0.2317	4.315	0.35	0.7	0.1303	7.674	0.85
0.8	0.7244	1.380	0.40	0.8.	0.4074	2.455	0.90	0.8	0.2291	4.365	0.40	0.8	0.1288	7.762	0.90
0.9	0.7161	1.396	0.45	0.9	0.4027	2.483	0.95	0.9	0.2265	4.416	0.45	0.9	0.1274	7.852	0.95
3.0	**0.7079**	**1.413**	**0.50**	**8.0**	**0.3981**	**2.512**	**4.00**	**13.0**	**0.2239**	**4.467**	**0.50**	**18.0**	**0.1259**	**7.943**	**9.00**
0.1	0.6998	1.429	0.55	0.1	0.3936	2.541	0.05	0.1	0.2213	4.519	0.55	0.1	0.1245	8.035	0.05
0.2	0.6918	1.445	0.60	0.2.	0.3890	2.570	0.10	0.2	0.2188	4.571	0.60	0.2	0.1230	8.128	0.10
0.3	0.6839	1.462	0.65	0.3	0.3846	2.600	0.15	0.3	0.2163	4.624	0.65	0.3	0.1216	8.222	0.15
0.4	0.6761	1.479	0.70	0.4	0.3802	2.630	0.20	0.4	0.2138	4.677	0.70	0.4	0.1202	8.318	0.20
0.5	0.6683	1.496	0.75	0.5	0.3758	2.661	0.25	0.5	0.2113	4.732	0.75	0.5	0.1189	8.414	0.25
0.6	0.6607	1.514	0.80	0.6	0.3715	2.692	0.30	0.6	0.2089	4.786	0.80	0.6	0.1175	8.511	0.30
0.7	0.6531	1.531	0.85	0.7	0.3673	2.723	0.35	0.7	0.2065	4.842	0.85	0.7	0.1161	8.610	0.35
0.8	0.6457	1.549	0.90	0.8	0.3631	2.754	0.40	0.8	0.2042	4.898	0.90	0.8	0.1148	8.710	0.40
0.9	0.6383	1.567	0.95	0.9	0.3589	2.786	0.45	0.9	0.2018	4.955	0.95	0.9	0.1135	8.810	0.45
4.0	**0.6310**	**1.585**	**2.00**	**9.0**	**0.3548**	**2.818**	**0.50**	**1.4.0**	**0.1995**	**5.012**	**7.00**	**19.0**	**0.1122**	**8.913**	**0.50**
0.1	0.6237	1.603	0.05	0.1	0.3508	2.851	0.55	0.1	0.1972	5.070	0.05	0.1	0.1109	9.016	0.55
0.2	0.6166	1.622	0.10	0.2	0.3467	2.884	0.60	0.2	0.1950	5.129	0.10	0.2	0.1096	9.120	0.60
0.3	0.6095	1.641	0.15	0.3	0.3428	2.917	0.65	0.3	0.1928	5.188	0.15	0.3	0.1084	9.226	0.65
0.4	0.6026	1.660	0.20	0.4	0.3388	2.951	0.70	0.4	0.1905	5.248	0.20	0.4	0.1072	9.333	0.70

dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power
0.5	0. 5957	1. 679	0.25	0.5	0.3350	2.985	0.75	0.5	0.1884	5.309	0.25	0.5	0.1059	9.441	0.75
0.6	0.5888	1.698	0.30	0.6	0.3311	3.020	0.80	0.6	0.1862	5.370	0.30	0.6	0.1047	9.550	0.80
0.7	0. 5821	1.718	0.35	0.7	0.3273	3.055	0.85	0.7	0.1841	5.433	0.35	0.7	0.1035	9.661	0.85
0.8	0. 5754	1.738	0.40	0.8	0.3236	3.090	0.90	0.8	0.1820	5.495	0.40	0.8	0.1023	9.772	0.90
0.9	0.5689	1.758	0.45	0.9	0.3199	3.126	0.95	0.9	0.1799	5.559	0.45	0.9	0.1012	9.886	0.95

dB Voltage	Loss	Gain	dB Power	dB Voltage	Loss	Gain	dB Power
20.0	0.1000	10.0	10.0	60.0	0.001	1,000	30.00
	Use the same number as 0–20 dB but shift decimal point one step to the left. Thus since 10 dB = 0.3162 30 dB = 0.03162	Use the same number as 0–20 dB but shift decimal point one step to the right. Thus since 10 dB = 3.162 30 dB = 31.62	This column repeats every 10 dB instead of 20 dB		Use the same numbers as 0–20 dB but shift point three steps to the left. Thus since 10 dB = 0.3162 70 dB = 0.0003162	Use the same number as 0–20 dB column but shift point three steps to the right. Thus since 10 dB = 3.162 70 dB = 3162	This column repeats every 10 dB instead of 20 dB
40.0	0.01	100	20	80	0.0001	10,000	40.00
	Use the same number as 0–20 dB but shift point two steps to the left. Thus since 10 dB = 0.3162 50 dB = 0.003162	Use the same number as 0–20 dB but shift point two steps to the right. Thus since 10 dB = 3162 50 dB = 316.2	This column repeats every 10 dB instead of 20 dB		Use the same numbers as 0–20 dB but shift point four steps to the left. Thus since 10 dB = 0.3162 90 dB = 0.00003162	Use the same number as 0–20 dB but shift point four steps to the right. Thus since 10 dB = 3.162 90 dB = 31620	This column repeats every 10 dB instead of 20 dB
				100	0.00001	100,000	50.00

When measuring decibels referenced to 1 mW at any other impedance than 600Ω, use

$$\text{dBm at new } Z = \text{dBm}_{600\,\Omega} + 10\log\frac{600\ \Omega}{Z_{new}} \qquad (50\text{-}44)$$

Example: The dBm for a 32 Ω load is

$$\begin{aligned}\text{dBm}_{32} &= 4\ \text{dBm} + 10\log\frac{600\ \Omega}{32\ \Omega}\\ &= 16.75\ \text{dBm}\end{aligned}$$

This can also be determined by using the graph in Fig. 50-5.

To find the logarithm of a number to some other base than the base 10 and 2.718, use

$$n = b^L \tag{50-45}$$

A number is equal to a base raised to its logarithm,

$$\ln(n) = \ln(bL) \tag{50-46}$$

therefore,

$$\frac{\ln(n)}{\ln(b)} = L \tag{50-47}$$

The natural log is a number divided by the natural log of the base equals the logarithm.

Example: Find the logarithm of the number 2 to the base 10:

$$\frac{\ln 2}{\ln 10} = \frac{0.693147}{2.302585}$$
$$= 0.301030$$

In information theory work, logarithms to the base 2 are quite commonly employed. To find the $\log_2$ of 26

$$\frac{\ln 26}{\ln 2} = 4.70$$

To prove this, raise 2 to the 4.70 power

$2^{4.70} = 26$

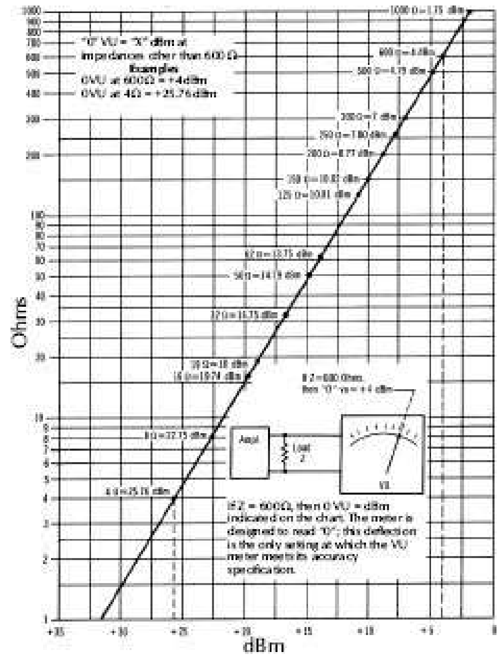


Figure 50-5. Relationship between VU and dBm at various impedances.

50.4 Sound Pressure Level

The *sound pressure level* (*SPL*) is related to acoustic pressure as seen in Fig. 50-6.

50.5 Sound System Quantities and Design Equations

Various quantities used for sound system design are defined as follows:

D_1. D_1 is the distance between the microphone and the loudspeaker, Fig. 50-7.

D_2. D_2 is the distance between the loudspeaker and the farthest listener, Fig. 50-7.

D_o. D_o is the distance between the talker (sound source) and the farthest listener, Fig. 50-7.

D_s. D_s is the distance between the talker (sound source) and the microphone, Fig. 50-7.

D_L. D_L is the *limiting distance* and is equal to 3.16 D_c for 15% *Alcons* in a room with a reverberation time of 1.6s. This means that D_2 cannot be any longer than D_L if *Alcons* is to be kept at 15% or less. As the *RT* increases or the required *%Alcons* decreases D_2 becomes less than D_L.

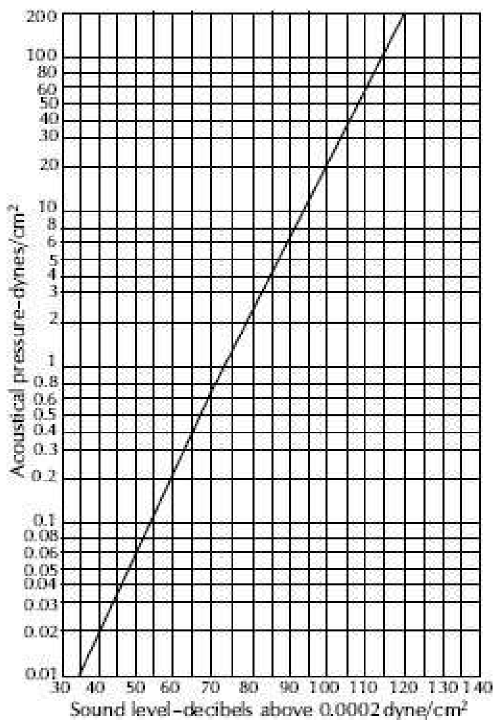


Figure 50-6. Sound pressure level versus acoustic pressure.

EAD. The *equivalent acoustic distance* (*EAD*) is the maximum distance from the talker that produces adequate loudness of the unamplified voice. Often an *EAD* of 8 ft is used in quiet surroundings as it is the distance at which communications can be understood comfortably. Once the *EAD* has been determined, the sound system is designed to produce that level at every seat in the audience.

D_c. *Critical distance* (D_c) is the point in a room where the direct sound and reverberant sound are equal. D_c is found by the equation

$$D_c = 0.141\frac{QRM}{N} \qquad (50\text{-}48)$$

where,
Q is the directivity of the sound source,
R is the room constant,
M is the critical distance modifier for absorption coefficient,
N is the modifier for direct-to-reverberant speaker coverage.

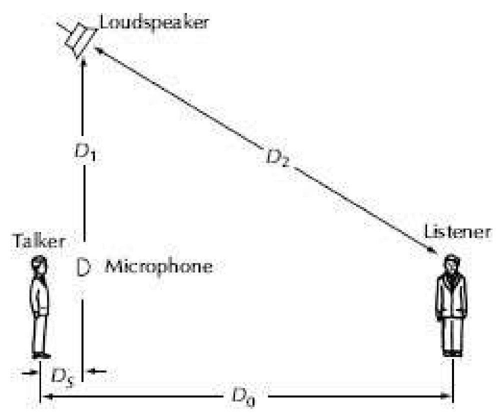


Figure 50-7. Definitions of sound system dimensions.

It can also be found with the equation

$$D_c = 0.3121^{**}\sqrt{\frac{Qv}{RT}} \qquad (50\text{-}49)$$

** 0.057 for SI units

M. The critical distance modifier (*M*) corrects for the effect of a different absorption coefficient within the path of the loudspeaker's coverage pattern.

$$M = \frac{1 - \bar{a}_{total\ room}}{1 - \bar{a}_{loudspeaker\ coverage\ area}} \qquad (50\text{-}50)$$

N. The critical distance modifier (*N*) corrects for multiple sound sources. *N* is the number describing the ratio of acoustic power going to the reverberant sound field without supplying direct sound versus the acoustic power going from the loudspeakers providing direct sound to a given listener position.

$$N = \frac{\text{Total number of loudspeakers}}{\text{Number providing direct sound}} \qquad (50\text{-}51)$$

%Alcons. The English language is made up of consonants and vowels. The consonants are the harsh letters that determine words. If the consonants of words are understood, the sentences or phrases will be understood. V. M. A. Peutz and W. Klein of Holland developed and published equations for the % articulation loss of consonants (*%Alcons*). The equation is

$$\%Alcons = 656^{**}\frac{RT_{60}{}^2 D_2{}^2 N}{VQM} \qquad (50\text{-}52)$$

** 200 for SI units

where,
Q is the directivity of the sound source,
V is the volume of the enclosure,
M is the critical distance modifier for absorption,

N is the critical distance modifier for multiple sources,
D_2 is the distance between the loudspeaker and the farthest listener.

When $D_c \geq D_L$, then $\%ALcons = 9RT_{60}$.

FSM. The *feedback stability margin* (*FSM*) is required to insure that a sound reinforcement system will not ring. A room and sound system, when approaching feedback, gives the effect of an long reverberation time. A room, for instance, with an *RT* of 3s could easily have an apparent *RT* of 6–12s when the sound system approaches feedback. To insure that this long reverberation time does not happen, a feedback stability margin of 6dB is added into the needed acoustic gain equation.

NOM. The *number of open microphones* (*NOM*) affects the gain of a sound reinforcement system. The system gain will be reduced by the following equation:

$$Gain\ reduction_{dB} = 10\log NOM \quad (50\text{-}53)$$

Every time the number of microphones doubles, the gain from the previous microphones is halved as the total gain is the gain of all the microphones added together.

NAG. The *needed acoustic gain (NAG)* is required to produce the same level at the farthest listener as at the *EAD*. *NAG* in its simplest form is

$$NAG = 20\log D_o - 20\log EAD \quad (50\text{-}54)$$

NAG, however, is also affected by the number of open microphones (*NOM*) in the system. Each time the *NOM* doubles, the *NAG* increases 3dB. Finally, a 6dB feedback stability margin (*FSM*) is added into the *NAG* formula to ensure that the system never approaches feedback. The final equation for *NAG* is

$$NAG = \Delta D_o - \Delta EAD + 10\log NOM + 6\text{ dB} + \text{FSM} \tag{50-55}$$

where,

ΔD_o and ΔEAD are the level change per the Hopkins-Stryker equation.

PAG. *The potential acoustic gain (PAG*) of a sound system is

$$PAG = \Delta D_o + \Delta D_1 - \Delta D_s - \Delta D_2 \tag{50-56}$$

where,

ΔD_o, ΔD_1, ΔD_s, and ΔD_2 are found as in *NAG*.

Q. The *directivity factor* (*Q*) of a transducer used for sound emission is the ratio of sound pressure squared, at some fixed distance and specified direction, to the mean sound pressure squared at the same distance averaged over all directions from the transducer. The distance must be great enough so that the sound appears to diverge spherically from the effective acoustic center of the source. Unless otherwise specified, the reference direction is understood to be that of maximum response.

Geometric *Q* can be found by using the following equations:

1. For rectangular coverage between 0° and 180°,

$$Q_{Geom} = \frac{180}{\arcsin\left(\frac{\sin\theta}{2}\right)\left(\frac{\sin\phi}{2}\right)} \tag{50-57}$$

2. For angles between 180° and 360° when one angle is 180°, and the other angle is some value between 180° and 360°

$$Q_{Geom} = \frac{360}{\text{angle}} \qquad (50\text{-}58)$$

3. For conical coverage,

$$Q_{Geom} = \frac{2}{1 - \cos\frac{\theta}{2}} \qquad (50\text{-}59)$$

C∠. *C*∠ is the *included angle* of the coverage pattern. Normally *C*∠ is expressed as an angle between the –6dB points in the coverage pattern.

EPR. *EPR* is the *electrical power required* to produce the desired SPL at a specific point in the coverage area. It is found by the equation

$$EPR_{watts} = 10^{\frac{\text{SPL}_{des} + 10\text{dB}_{crest} + \Delta D_2 - \Delta D_{ref} - L_{sens}}{10}} \qquad (50\text{-}60)$$

a. The *absorption coefficient* (*a*) of a material or surface is the ratio of absorbed sound to reflected sound or incident sound

$$a = \frac{I_A}{I_R} \qquad (50\text{-}61)$$

If all sound was reflected, *a* would be 0. If all sound were absorbed, *a* would be 1.

ā. The *average absorption coefficient* (*ā*) for all the surfaces together and is found by

$$\bar{a} = \frac{S_1 a_1 + S_2 a_2 + \ldots S_n a_n}{S} \qquad (50\text{-}62)$$

where,
$S_{1,2\ldots n}$ are individual surface areas,
$a_{1,2\ldots n}$ are the individual absorption coefficients of the areas,
S is the total surface area.

MFP. The *mean-free path (MFP*) is the average distance between reflections in a space. *MFP* is found by

$$MFP = \frac{4V}{S} \qquad (50\text{-}63)$$

where,
V is the space volume,
S is the space surface area.

ΔD_x. ΔD_x is an arbitrary level change associated with the specific distance from the Hopkins-Stryker equation so that

$$\Delta D_x = -10\log\left[\frac{Q}{4\pi D_x{}^2} + \frac{4N}{Sa}\right] \qquad (50\text{-}64)$$

In semireverberant rooms, Peutz describes ΔD_x as

$$\Delta D_x = -10\log\left[\frac{Q}{4\pi D_x{}^2} + \frac{4N}{Sa}\right] + \frac{0.734^{**}\sqrt{V}}{h\mathrm{RT}_{60}}\log\frac{D_x > D_c}{D_c} \qquad (50\text{-}65)$$

** 200 for SI units

where,
h is the ceiling height.

SNR. *SNR* is the acoustical *signal-to-noise ratio*. The signal-to-noise ratio required for intelligibility is

$$SNR = 35\left(\frac{2-\log\%Alcons}{2-\log 9RT}\right) \qquad (50\text{-}66)$$

SPL. *SPL* is the *sound pressure level* in dB-SPL re 0.00002N/m^2. *SPL* is also called L_p.

Max Program Level. *Max program level* is the maximum program level attainable at a specific point from the available input power. Max program level is

$$program\ level_{max} = 10\log\frac{watts_{avail}}{10} - (\Delta D_2 - \Delta D_{ref}) + L_{sens} \qquad (50\text{-}67)$$

L_{sens}. *Loudspeaker sensitivity* (L_{sens}) is the on-axis *SPL* output of the loudspeaker with a specified power input and at a specified distance. The most common L_{sens} are at 4 ft, 1W and 1m at 1W.

Sa. *Sa* is the *total absorption* in sabines of all the surface areas times their absorption.

dB-SPL$_T$ The *dB-SPL$_T$* is the talker's or sound source's *sound pressure level*.

dB-SPL$_D$. *The dB-SPL$_D$ is the desired sound pressure level.*

dB-SPL. The dB-SPL is the *sound pressure level* in decibels.

EIN. *EIN is the equivalent input noise.*

$$EIN = -198\text{ dB} + 10\log BW + 10\log Z + -6\text{ dB} - 20\log 0.775 \quad (50\text{-}68)$$

where,
BW is the bandwidth,
Z is the impedance.

Thermal Noise. *Thermal noise* is the noise produced in any resistance, including standard resistors. Any resistance that is at a temperature above absolute zero generates noise due to the thermal agitation of free electrons in the material. The magnitude of the noise can be calculated from the resistance, absolute temperature, and equivalent noise bandwidth of the measuring system. A completely noise-free amplifier whose input is connected to its equivalent source resistance will have noise in its output equal to the product of amplification and source resistor noise. This noise is said to be the *theoretical minimum.*

Fig. 50-8 provides a quick means for determining the rms value of thermal noise voltage in terms of resistance and circuit bandwidth.

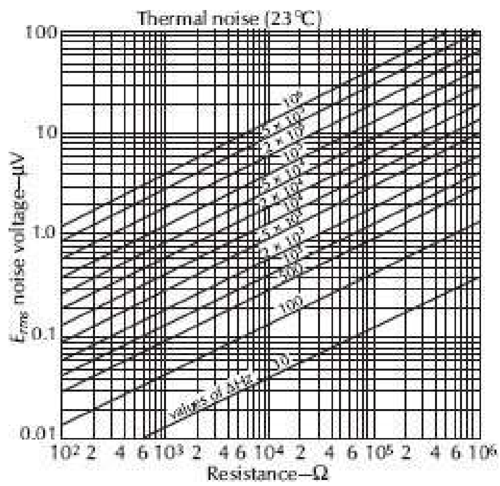


Figure 50-8. Thermal noise graph.

For practical calculations, especially those in which the resistive component is constant across the bandwidth of interest, use

$$E_{rms} = \sqrt{4 \times 10^{-23}(T)(f_1 - f_2)R} \qquad (50\text{-}69)$$

where,

f_1–f_2 is the 3 dB bandwidth,

R is the resistive component of the impedance across which the noise is developed,

T is the absolute temperature in K.

RT. *RT* is the time required for an interrupted steady-state signal in a space to decay 60dB. *RT* is normally calculated using one of the following equations: the classic Sabine method, the Norris Eyring modification of the Sabine equation, and the Fitzroy equation. The Fitzroy equation is best used when the walls in the *X, Y*, and *Z* planes have very different absorption materials on them.

Sabine:

$$RT_{60} = 0.049^{**}\frac{V}{S\bar{a}} \quad (50\text{-}70)$$

** 0.161 for SI units.

Norris Eyring:

$$RT_{60} = 0.049^{**}\frac{V}{-S\ln(1-\bar{a})} \quad (50\text{-}71)$$

** 0.161 for SI unit.

Fitzroy:

$$RT_{60} = \frac{0.049^{**}V}{S^2}\left[\frac{2XY}{-\ln(1-\bar{a}_{XY})} + \frac{2XZ}{-\ln(1-\bar{a}_{XZ})} + \frac{2YZ}{-\ln(1-\bar{a}_{YZ})}\right] \quad (50\text{-}72)$$

** 0.161 for SI units
where,
V is the room volume,
S is the surface area,
a is the total absorption coefficient,
X is the space length,
Y is the space width,
Z is the space height.

Signal Delay. *Signal delay* is the time required for a signal, traveling at the speed of sound, to travel from the source to a specified point in space

$$SD = \frac{Distance}{c} \quad (50\text{-}73)$$

where,
SD is the signal delay in ms,
c is the speed of sound.

50.6 ISO Numbers

From the American National Standards for Preferred Numbers

> *Preferred Numbers were developed in France by Charles Renard in 1879 because of a need for a rational basis for grading cotton rope. The sizing system that resulted from his work was based upon a geometric series of mass per unit length such that every fifth step of the series increased the size of rope by a factor of ten.*

This same system of preferred numbers is used today in acoustics. The one-twelfth, one-sixth, one-third, one-half, two-thirds, and one octave preferred center frequency numbers are not the exact *n* series number. The exact *n* series number is found by the equation

$$n \text{ Series number} = 10^{\frac{1}{n}}\left(10^{\frac{1}{n}}\right)\left(10^{\frac{1}{n}}\right)\ldots \qquad (50\text{-}74)$$

where,
n is the ordinal numbers in the series.

For instance, the third *n* number for a 40 series would be

$$10^{\frac{1}{40}}\left(10^{\frac{1}{40}}\right)\left(10^{\frac{1}{40}}\right) = 1.1885022$$

The preferred ISO number is 1.18. Table 50-4 is a table of preferred International Standards Organization (ISO) numbers.

Table 50-4. Internationally Preferred ISO Numbers

1/12 oct. 40 ser.	1/6 oct. 20 ser.	1/3 oct. 10 ser.	1/2 oct. 6⅓ ser.	2/3 oct. 5 ser.	1/1 oct. 3⅓ ser.	Exact value
1.00	1.00	1.00	1.00	1.00	1.00	1.000000000
1.06						1.059253725
1.12	1.12					1.122018454
1.18						1.188502227
1.25	1.25	1.25				1.258925411
1.32						1.333521431
1.40	1.40					1.412537543
1.50						1.496235654
1.60	1.60	1.60		1.60		1.584893190
1.70						1.678804015
1.80	1.80					1.778279406
1.90						1.883649085
2.00	2.00	2.00	2.00		2.00	1.995262310
2.12						2.113489034
2.24	2.24					2.238721132
2.36						2.371373698
2.50	2.50	2.50		2.50		2.511886423
2.65						2.660725050
2.80	2.80		2.80			2.818382920
3.00						2.985382606
3.15	3.15	3.15				3.162277646
3.35						3.349654376
3.55	3.55					3.548133875
3.75						3.758374024
4.00	4.00	4.00	4.00	4.00	4.00	3.981071685
4.25						4.216965012
4.50	4.50					4.466835897
4.75						4.731512563
5.00	5.00	5.00				5.011872307
5.30						5.308844410
5.60	5.60		5.60			5.623413217

6.00						5.956621397
6.30	6.30	6.30		6.30		6.309573403
6.70						6.683439130
7.10	7.10					7.079457794
7.50						7.498942039
8.00	8.00	8.00	8.00		8.00	7.943282288
8.50						8.413951352
9.00	9.00					8.912509312
9.50						9.440608688

50.7 Greek Alphabet

The *Greek alphabet* plays a major role in the language of engineering and sound. Table 50-5 shows the Greek alphabet and the terms that are commonly symbolized by it.

Table 50-5. Greek Alphabet

Name	Upper Case		Lower Case	
alpha	A		α	absorption factor, angles, angular acceleration, attenuation constant, common-base current amplification factor, deviation of state parameter, temperature coefficient of linear expansion, temperature coefficient of resistance, thermal expansion coefficient, thermal diffusivity
beta	B		ß	angles, common-emitter current amplification factor, flux density, phase constant, wavelength constant
gamma	Γ		γ	electrical conductivity, Grueneisen parameter
delta	Δ	decrement increment	δ	angles, damping coefficient (decay constant), decrement, increment, secondary-emission ratio
epsilon	E	electric field intensity	ε	capacitivity, dielectric coefficient, electron energy, emissivity, permittivity, base of natural logarithms (2.71828)
zeta	Z		ζ	chemical potential, dielectric susceptibility (intrinsic capacitance), efficiency, hysteresis, intrinsic impedance of a medium, intrinsic standoff ratio
eta	H		η	
theta	Θ	angles, thermal resistance	θ	angle of rotation, angles, angular phase displacement, reluctance, transit angle
iota	I		ι	

kappa	Κ	coupling coefficient	κ	susceptibility
lambda	Λ		λ	line density of charge, permeance, photosensitivity, wavelength
mu	Μ		μ	amplification factor, magnetic permeability, micron, mobility, permeability, prefix micro
nu	Ν		ν	reluctivity
xi	Ξ		ξ	
omicron	Ο		ο	
pi	Π		π	Peltier coefficient, ratio of circumference to diameter (3.1416)
rho	Ρ		ρ	reflection coefficient, reflection factor, resistivity, volume density of electric charge
sigma	Σ	summation	σ	conductivity, Stefan-Boltzmann constant, surface density of charge
tau	Τ	period	τ	propagation constant, Thomson coefficient, time constant, time-phase displacement, transmission factor
upsilon	Υ	admittance	υ	
phi	Φ	magnetic flux, radiant flux	φ	angles, coefficient of performance, contact potential, magnetic flux, phase angle, phase displacement
chi	Χ			angles
psi	Ψ	angles	ψ	dielectric flux, displacement flux, phase difference
omega	Ω	resistance	ω	angular frequency, angular velocity, solid angle

50.8 Audio Standards

Audio standards are defined by the Audio Engineering Society (AES), Table 50-6; and International Electrotechnical Commission (IEC) Table 50-7.

Table 50-6. AES Standards

Standards and Recommended Practices	
AES2-1984 (r2003)	AES recommended practice – Specification of loudspeaker components used in professional audio and sound reinforcement

AES2-2012	Revised Measuring the performance of loudspeakers – Drive units
AES3-2003	AES recommended practice for digital audio engineering – Serial transmission format for two-channel linearly represented digital audio data (Revision of AES3-1992, including subsequent amendments)
AES5-2003	AES recommended practice for professional digital audio – Preferred sampling frequencies for applications employing pulse-code modulation (revision of AES5-1997)
AES6-1982 (r2003)	Method for measurement of weighted peak flutter of sound recording and reproducing equipment
AES7-2000 (r2005)	AES standard for the preservation and restoration of audio recording – Method of measuring recorded fluxivity of magnetic sound records at medium wavelengths (Revision of AES7-1982)
AES10-2003	AES recommended practice for digital audio engineering – Serial Multichannel Audio Digital Interface (MADI) (Revision of AES10-1991)
AES11-2003	AES recommended practice for digital audio engineering – Synchronization of digital audio equipment in studio operations. (Revision of AES11-1997)
AES14-1992 (r2004)	AES standard for professional audio equipment – Application of connectors, part 1, XLR-type polarity and gender
AES15-1991 (w2002)	AES recommended practice for sound-reinforcement systems – Communications interface (PA-422) (Withdrawn: 2002)
AES17-1998 (r2004)	AES standard method for digital audio engineering – Measurement of digital audio equipment (Revision of AES17-1991)
AES18-1996 (r2002)	AES recommended practice for digital audio engineering – Format for the user data channel of the

	AES digital audio interface. (Revision of AES18-1992)
AES19-1992 (w2003)	AES-ALMA standard test method for audio engineering – Measurement of the lowest resonance frequency of loudspeaker cones (Withdrawn: 2003)
AES20-1996 (r2002)	AES recommended practice for professional audio – Subjective evaluation of loudspeakers
AES22-1997 (r2003)	AES recommended practice for audio preservation and restoration – Storage and handling – Storage of polyester-base magnetic tape
AES24-1-1999 (w2004)	AES standard for sound system control—Application protocol for controlling and monitoring audio devices via digital data networks – Part 1: Principles, formats, and basic procedures (Revision of AES24-1-1995)
AES24-2-tu (w2004)	PROPOSED DRAFT AES standard for sound system control – Application protocol for controlling and monitoring audio devices via digital data networks – Part 2, data types, constants, and class structure (for Trial Use)
AES26-2001	AES recommended practice for professional audio – Conservation of the polarity of audio signals (Revision of AES26-1995)
AES27-1996 (r2002)	AES recommended practice for forensic purposes – Managing recorded audio materials intended for examination
AES28-1997 (r2003)	AES standard for audio preservation and restoration – Method for estimating life expectancy of compact discs (CD-ROM), based on effects of temperature and relative humidity (includes Amendment 1-2001)
AES31-1-2001 z(r2006)	AES standard for network and file transfer of audio – Audio-file transfer and exchange Part 1: Disk format
AES31-2-2006	AES standard on network and file transfer of audio – Audio-file transfer and exchange – File format for transferring digital audio data between systems of different type and manufacture

AES31-2-2012	Revised Audio file format for interchange
AES31-3-1999	AES standard for network and file transfer of audio—Audio-file transfer and exchange—Part 3: Simple project interchange
AES32-tu	PROPOSED DRAFT AES standard for professional audio interconnections – Fibre optic connectors, cables, and characteristics (for Trial Use)
AES33-1999 (w2004)	AES standard – For audio interconnections – Database of multiple – program connection configurations (Withdrawn: 2004)
AES35-2000 (r2005)	AES standard for audio preservation and restoration – Method for estimating life expectancy of magneto-optical (M-O) disks, based on effects of temperature and relative humidity
AES38-2000 (r2005)	AES standard for audio preservation and restoration – Life expectancy of information stored in recordable compact disc systems – Method for estimating, based on effects of temperature and relative humidity
AES41-2000 (r2005)	AES standard for digital audio – Recoding data set for audio bit-rate reduction
AES41-1-2012	Revised Audio-embedded metadata – Part 1: General
AES41-2-2012	Revised Audio-embedded metadata – Part 2: MPEG-1 Layer II or MPEG-2 LSF Layer II
AES41-3-2012	Audio-embedded metadata – Part 3: AAC & HE-AAC
AES41-4-2012	Audio-embedded metadata – Part 4: Dolby E
AES41-5-2012	Audio-embedded metadata – Part 5: EBU loudness, true-peak, and downmix
AES42-2006	AES standard for acoustics – Digital interface for microphones
AES43-2000 (r2005)	AES standard for forensic purposes – Criteria for the authentication of analog audio tape recordings

AES45-2001	AES standard for single program connectors – Connectors for loudspeaker-level patch panels
AES46-2002	AES standard for network and file transfer of audio Audio-file transfer and exchange, Radio traffic audio delivery extension to the broad-cast-WAVE-file format
AES47-2006	AES standard for digital audio – Digital input-output interfacing – Transmission of digital audio over asynchronous transfer mode (ATM) networks
AES48-2005	AES standard on interconnections – Grounding and EMC practices – Shields of connectors in audio equipment containing active circuitry
AES49-2005	AES standard for audio preservation and restoration —Magnetic tape – Care and handling practices for extended usage
AES50-2005	AES standard for digital audio engineering – High-resolution multichannel audio interconnection
AES50-2011	Revised High-resolution multi-channel audio interconnection (HRMAI)
AES51-2006	AES standard for digital audio – Digital input-output interfacing – Transmission of ATM cells over Ethernet physical layer
AES52-2006	AES standard for digital audio engineering—Insertion of unique identifiers into the AES3 transport stream
AES53-2006	AES standard for digital audio – Digital input-output interfacing – Sample-accurate timing in AES47
AES55-2012	Revised MPEG Surround in an AES3 bitstream
AES57-2011	Metadata – Audio object structures for preservation and restoration
AES59-20122	5-way D-type connectors in balanced circuits
AES60-2011	Core audio metadata
AES62-2011	Modified XL3 Connector for Digital Audio
AES63-2012	Data connector in an XLR connector shell

AES64-2012	Networks: Command, control, and connection management
AES65-2012	Connector for surround microphones
AES66-2012	Miniature XLR-type polarity and gender
AES67-2013	Networks, Audio-over-IP interoperability
Duplicate entry	Duplicate entry
AES-1id-1991 (r2003)	AES information document – Plane wave tubes: design and practice
AES-2id-1996 (r2001)	AES information document for digital audio engineering—Guidelines for the use of the AES3 interface
AES-2id-2012	Revised Guidelines for the AES3 interface
AES-3id-2001 (r2006)	AES information document for digital audio engineering – Transmission of AES3 formatted data by unbalanced coaxial cable (Revision of AES-3id-1995)
AES-4id-2001	AES information document for room acoustics and sound reinforcement systems – Characterization and measurement of surface scattering uniformity
AES-5id-1997 (r2003)	AES information document for room acoustics and sound reinforcement systems – Loudspeaker modeling and measurement – Frequency and angular resolution for measuring, presenting, and predicting loudspeaker polar data
AES-6id-2000	AES information document for digital audio – Personal computer audio quality measurements
AES-10id-2005	AES information document for digital audio engineering – Engineering guidelines for the multichannel audio digital interface, AES10 (MADI)
AES-21id-2011	High-resolution audio on Blu-ray Disc™
AES-R1-1997	AES project report for professional audio –

	Specifications for audio on high-capacity media
AES-R2-2004	AES project report for articles on professional audio and for equipment specifications—Notations for expressing levels (Revision of AES-R2-1998)
AES-R3-2001	AES standards project report on single program connector – Compatibility for patch panels of tip-ring-sleeve connectors
AES-R4-2002	AES standards project report. Guidelines for AES Recommended practice for digital audio engineering – Transmission of digital audio over asynchronous transfer mode (ATM) networks
AES-R6-2005	AES project report–Guidelines for AES standard for digital audio engineering–High-resolution multichannel audio interconnection (HRMAI)
AES-R7-2006	AES standards project report – Considerations for accurate peak metering of digital audio signals

Table 50-7. IEC Standards

IEC Number	IEC Title
IEC 60038	IEC standard voltages
IEC 60063	Preferred number series for resistors and capacitors
IEC 60094	Magnetic tape sound recording and reproducing systems
IEC 60094-5	Electrical magnetic tape properties
IEC 60094-6	Reel-to-reel systems
IEC 60094-7	Cassette for commercial tape records and domestic use
IEC 60096	Radio-frequency cables
IEC 60098	Rumble measurement on vinyl disc turntables
IEC 60098	Rumble measurement on vinyl disc turntables
IEC 60134	Absolute maximum and design ratings of tube and semiconductor devices

IEC 60169	Radio-frequency connectors
IEC 60169-2	Unmatched coaxial connector (Belling-Lee TV Aerial Plug)
IEC 60169-8	BNC connector, 50 ohm
IEC 60169-9	SMC connector, 50 ohm
IEC 60169-10	SMB connector, 50 ohm
IEC 60169-15	N connector, 50 ohm or 75 ohm
IEC 60169-16	SMA connector, 50 ohm
IEC 60169-16	TNC connector, 50 ohm
IEC 60169-24	F connector, 75 ohm
IEC 60179	Sound level meters
IEC 60228	Conductors of insulated cables
IEC 60268	Sound system equipment
IEC 60268-1	General
IEC 60268-2	Explanation of general terms and calculation methods
IEC 60268-3	Amplifiers
IEC 60268-4	Microphones
IEC 60268-5	Loudspeakers
IEC 60268-6	Auxiliary passive elements
IEC 60268-7	Headphones and earphones
IEC 60268-8	Automatic gain control devices
IEC 60268-9	Artificial reverberation, time delay, and frequency shift equipment
IEC 60268-10	Peak program level meters
IEC 60268-11	Application of connectors for the interconnection of sound system components
IEC 60268-12	Application of connectors for broadcast and similar use
IEC 60268-13	Listening tests on loudspeakers
IEC 60268-14	Circular and elliptical loudspeakers; outer frame diameters and mounting dimensions
IEC 60268-16	Objective rating of speech intelligibility by speech

	transmission index
IEC 60268-17	Standard volume indicators
IEC 60268-18	Peak program level meters – Digital audio peak level meter
IEC 60297	19-inch rack
IEC 60386	Wow and flutter measurement (audio)
IEC 60417	Graphical symbols for use on equipment
IEC 60446	Wiring colors
IEC 60461	Time and control code
IEC 60574	Audio-visual, video, and television equipment and systems
IEC 60581	High fidelity audio equipment and systems: minimum performance requirements
IEC 60651	Sound level meters
IEC 60728	Cable networks for television signals, sound signals and interactive services
IEC 60899	Sampling rate and source encoding for professional digital audio recording
IEC 60908	Compact disk digital audio system
IEC 60958	Digital audio interfaces
IEC 61043	Sound intensity meters with pairs of microphones
IEC 61305	Household high-fidelity audio equipment and systems – Methods of measuring and specifying the performance
IEC 61603	Infrared transmission of audio or video signals
IEC 61329	Sound system equipment – Methods of measuring and specifying the performance of sounders
IEC 61606	Audio and audiovisual equipment – Digital audio parts – Basic measurement methods of audio characteristics
IEC 61966	Multimedia systems – Color measurement
IEC 61966-2-1	sRGB default RGB color space
IEC 62458	Sound system equipment – Electroacoustical

	transducers	

50.9 Audio Frequency Range

The audio spectrum is usually considered the frequency range between 20Hz and 20kHz, Fig. 50-9. In reality, the upper limit of hearing pure tones is between 12kHz and 18kHz, depending on the person's age and sex and how well the ears have been trained and protected against loud sounds. Frequencies above 20kHz cannot be heard as a sound, but the effect created by such frequencies, i.e., rapid rise time, can be heard.

The lower end of the spectrum is more often felt than heard as a pure tone. Frequencies below 20 Hz are difficult to reproduce. Often the reproducer actually reproduces the second harmonic of the frequency, and the brain translates it back to the fundamental.

50.10 Common Conversion Factors

Conversion from US to SI units can be made by multiplying the US unit by the conversion factors in Table 50-8. To convert from SI units to US units, divide by the conversion factor.

Table 50-8. US to SI Units Conversion Factors

US Units to	SI Unit	Multiply by
Length		
ft	m	$3.048\ 000 \times 10^{-1}$
mi	m	$1.609\ 344 \times 10^{3}$
in	m	$2.540\ 000 \times 10^{-2}$
Area		
ft^2	m^2	$9.290\ 304 \times 10^{-2}$
in^2	m^2	$6.451\ 600 \times 10^{-4}$
yd^2	m^2	$8.361\ 274 \times 10^{-1}$

Capacity/volume		
in^3	m^3	$1.638\ 706 \times 10^{-5}$
ft^3	m^3	$2.831\ 685 \times 10^{-2}$
liquid gal	m^3	$3.785\ 412 \times 10^{-3}$
Volume/mass		
ft^3/lb	m^3/kg	$6.242\ 796 \times 10^{-2}$
in^3/lb	m^3/kg	$3.612\ 728 \times 10^{-5}$
Velocity		
ft/h	m/s	$4.466\ 667 \times 10^{-5}$
in/s	m/s	$2.540\ 000 \times 10^{-2}$
mi/h	m/s	$4.470\ 400 \times 10^{-1}$
Mass		
oz	kg	$2.834\ 952 \times 10^{-2}$
lb	kg	$4.535\ 924 \times 10^{-1}$
Short Ton (2000lb)	kg	$9.071\ 847 \times 10^{2}$
Long Ton (2240 lb)	kg	$1.016\ 047 \times 10^{3}$
Mass/volume		
oz/in^3	kg/m^3	$1.729\ 994 \times 10^{3}$
lb/ft^3	kg/m^3	$1.601\ 846 \times 10^{1}$
lb/in^3	kg/m^3	$2.767\ 990 \times 10^{4}$
lb/US Gal	kg/m^3	$1.198\ 264 \times 10^{2}$
Acceleration		
ft/s^2	m/s^2	$3.048\ 000 \times 10^{-1}$
Angular Momentum		
lb f^2/s	$kg.m^2/s$	$4.214\ 011 \times 10^{-2}$
Electricity		
A•.8 or (tF −32h	C	$3.600\ 000 \times 10^{3}$
Gs	**T**	$1.000\ 000 \times 10^{-4}$
Mx	Wb	$1.000\ 000 \times 10^{-8}$
Mho	**S**	$1.000\ 000 \times 10^{0}$
Oe	A/m	$7.957\ 747 \times 10^{1}$
Energy (Work)		
Btu	J	$1.055\ 056 \times 10^{3}$
eV	J	$1.602\ 190 \times 10^{-19}$
W•h	J	$3.600\ 000 \times 10^{3}$
erg	J	$1.000\ 000 \times 10^{-7}$
Cal	J	$4.186\ 800 \times 10^{0}$

Force		
dyn	N	$1.000\ 000 \times 10^{-5}$
lbf	N	$4.448\ 222 \times 10^{0}$
pdl	N	$1.382\ 550 \times 10^{-1}$
Heat		
Btu/ft^2	J/m^2	$1.135\ 653 \times 10^{4}$
Btu/lb	J/hg	$2.326\ 000 \times 10^{3}$
Btu/(h•ft^2•°F) or k (thermal conductivity)	W/m•K	$1.730\ 735 \times 10^{0}$
Btu/(h•f^2•°F) or C (thermal conductance)	W/m^2•K	$5.678\ 263 \times 10^{0}$
Btu/(lb °F) or c (heat capacity)	J/kg•K	$4.186\ 800 \times 10^{3}$
°F•hft^2/Btu or R (thermal resistance)	K•m^2/W	$1.761\ 102 \times 10^{-1}$
cal	J	$4.186\ 000 \times 10^{0}$
cal/g	J/kg	$4.186\ 000 \times 10^{3}$
Light		
Cd (candle power)	cd (candela)	$1.000\ 000 \times 10^{0}$
fc	lx	$1.076\ 391 \times 10^{1}$
fL	cd/m^2	$3.426\ 259 \times 10^{0}$
Moment of Inertia		
lb•ft^2	kg•m^2	$4.214\ 011 \times 10^{-2}$
Momentum		
lbf/s	kg•m/s	$1.382\ 550 \times 10^{-1}$
Power		
Btu/h	W	$2.930\ 711 \times 10^{-1}$
erg/s	W	$1.000\ 000 \times 10^{-7}$
hp (550 ft/lb/s)	W	$7.456\ 999 \times 10^{2}$
hp (electric)	W	$7.460\ 000 \times 10^{2}$
Pressure		
atm (normal atmosphere)	Pa	$1.031\ 250 \times 10^{5}$
bar	Pa	$1.000\ 000 \times 10^{5}$
in Hg@ 60°F	Pa	$3.376\ 850 \times 10^{3}$
dyn/cm^2	Pa	$1.000\ 000 \times 10^{-1}$
cm Hg@ 0°C	Pa	$1.333\ 220 \times 10^{3}$

lbf/f^2	Pa	$4.788\ 026 \times 10^1$
pdl/ft^2	Pa	$1.488\ 164 \times 10^0$
Viscosity		
cP	Pa•s	$1.000\ 000 \times 10^{-3}$
lb/fhs	Pa•s	$1.488\ 164 \times 10^0$
ft^2/s	m^2/s	$9.290\ 304 \times 10^{-2}$
Temperature		
°C	K	$t_C + 273.15$
°F	K	$(t_F + 459.67)/1.8$
°R	K	$t_R/1.8$
°F	°C	$(t_F 32)/1.8$ or $(tF - 32) \times (5/9)$
°C	°F	$1.8\ (t_C) + 32$ or $(tc \times (9/5)) + 32$

50.11 Technical Abbreviations

Many units or terms in engineering have abbreviations accepted either by the US government or by the acousticians and audio consultants and engineers. Table 50-9 is a list of many of these abbreviations. Symbols for multiple and submultiple prefixes are shown in Table 50-1.

Table 50-9. Recommended Abbreviations

Symbol or Abbreviation	Unit or Term
... °	degree (plane angle)
...′ or ...′	minute (plane angle)
%Alcons	percentage of articulation loss for consonants degree (temperature)
°C	degree Celsius
°F	degree Fahrenheit
″	second (plane angle)

3PDT	three-pole, double-throw
3PST	three-pole, single-throw
4PDT	four-pole, double-throw
4PST	four-pole, single-throw
α	absorption coefficient
A	ampere
Å	angstrom
A/D	analog to digital
AAC	advanced audio coding
AACS	advanced access control system
A_{avg}	average amplitude
ABR	available bit rate
ac	alternating current
ACMC	alien crosstalk margin computation
ACR	attenuation-to-cross talk ratio
ADA	Americans with Disabilities Act
ADC	analog-to-digital converter
ADPCM	adaptive delta pulse code modulation
ADSL	asymmetric digital subscriber line
AE	audio erase
AES	Advanced Encryption standard, Audio Engineering Society
AF	audio frequency
AFC	automatic frequency control
AFEXT	alien far-end crosstalk
AFP	apple file protocol
AFTRA	American Federation of Television and Radio Artists
AGC	automatic gain control
Ah	ampere-hour
AHD	audio high density
AI	articulation index

AIP	available input power
ALC	automatic level control
ALD	assistive listening devices
ALS	assistive listening systems
AM	amplitude modulation
ANEXT	alien near-end crosstalk
ANL	ambient noise level
ANSI	American National Standards Institute
antilog	antilogarithm
AoIP	audio over IP
A_p	peak amplitude
A_{p-p}	peak-to-peak amplitude
APD	avalanche photodiodes
APF	all-pass filter
AR	assisted resonance
ARI	radio information for motorists
ASA	Acoustical Society of America, American Standards Association
ASCII	American Standard Code for Information Interchange
ASHRAE	American Society of Heating, Refrigeration and Air Conditioning Engineers
ASTM	American Society for Testing and Materials
At	ampere-turn
ATA	advanced technology attachment\
ATAC	Adaptive Transform Acoustic Coding
ATE	automated test equipment
ATM	asynchronous transfer mode
atm at	atmosphere normal atmosphere technical atmosphere
ATRAC	Adaptive Transform Acoustic Coding
ATSC	Advanced Television Systems Committee

AVB	Audio Video Bridging
AVC	automatic volume control
avg	average
AVR	audio/video receivers
AWG	American wire gauge
AWM	appliance wiring material
b	barn, bit
B	bel, flux density
B-Y	blue minus luminance
Balun	balanced to unbalanced (Bal-Un)
BAS	building automation systems
BBS	bulletin board service
BCA	balanced current amplifier
BCD	binary coded decimal
Bd	baud
BDP	bandwidth distance product
BER	bit error rate
BFO	beat-frequency oscillator
BIR	binaural impulse response
BJT	bipolar junction transistor
BLC	backlight compensation
BMG	background music
BOVP	battery overvoltage protection
BPF	bandpass filter
bps	bits per second
BPSK	binary phase shift keying
BRI	basic rate interface ISDN
BSI	British Standards Institution
Btu	British thermal unit
BV	breakdown voltage
BW	bandwidth
BWO	backward-wave oscillator

C	capacitance; capacitor, coulomb, coverage angle
CABA	Continental Automated Building Association
CAC	Ceiling Attenuation Class
CAD	computer aided design
CAF	Common Amplifier Format
cal_{IT}	calorie (International Table calorie)
cal_{th}	calorie (thermochemical calorie)
CAP	carrierless amplitude phase modulation
CAPS	constant-amplitude phase-shift
CAV	constant angular velocity
CB	citizens band, constant bandwidth, critical bands
CBF	constant bandwidth filter
CBR	constant bit rate
CCCA	Communications Cable and Connectivity Cable Association
CCD	charge coupled device
CCFL	cold cathod fluorescent lamps
CCIR	Comité Consultatif International des Radiocommunications
CCIR	International Radio Consultative Committee
CCITT	International Telegraph and Telephone Consultative Committee
CCNA	Cisco Certified Network Associate
CCNP	Cisco Certified Network Professional
CCT	correlated color temperature
CCTV	closed circuit television
cd	candela, candle
CD	compact disc
CD-DA	compact disc digital audio

CD-G	music cd plus graphics
CD-I	compact disc interactive
CD-UDF	cd universal device format
cd/ft^2	candela per square foot
cd/m^2	candela per square meter
CDR	clock and data recovery
CEA	Consumer Electronics Association
CEC	Canadian Electrical Code, Consumer Electronics Control
CEDIA	Custom Electronics Design and Installation Association
CFL	compact fluorescent lamps
CGS	centimeter-gram-second
Ci	curie
CIE	Commission Internationale de l’Eclairage
CIRC	Cross Interleave Reed Solomon Code
CLF	Common Loudspeaker Format
CLV	constant linear velocity
cm	centimeter
cm^3	cubic centimeter
cmil	circular mil
CMR, CMRR	common mode rejection or common mode rejection ratio
CMTS	cable modem termination systems
CO	central office
CODEC	compression/decompression algorithm
COLS	commercial online service
CP	consolidation point
CPB	constant percentage bandwidth
CPE	chlorinated polyethylene
CPRI	Common Public Radio Interface
CPU	central processing unit

CRC	cyclic redundancy check
CRO	cathode-ray oscilloscope
CRS	contact resistance stability
CRT	cathode-ray tube
CSA	Canadian Standards Association
CSMA/CD	carrier sense multiple access/collision detection
CSR	customer service representative
CSS	content scrambling system
CTD	charge transfer device
CTS	certified technology specialist
CW	continuous wave
D-VHS	digital VHS
D/A	digital to analog
DA	dielectric absorption
DAB	digital audio broadcasting
DAC	digital-to-analog converter
DASH	digital audio stationary head
DAT	digital audio tape
DAVIC	Digital Audio Video Council
DAW	digital audio workstations
dB	decibel
dBA	A-weighted sound-pressure level in decibels
dB_{DIR}	direct sound level in dB
dBm	decibel ref to one milliwatt
DBS	direct broadcast satellite
dB_{SPL}	sound pressure in dB
dBV	decibels with a reference of 1V
dc	direct current
Dc	critical distance
DCC	digital compact cassette

DCP	digital cinema packaging
DDC	display data channel
DDM	direct to disk mastering
DDR	double data rate
DDS	digital data storage
DEPIC	dual expanded plastic insulated conductor
DES	data encryption standard
DF	dissipation factor
DFP	digital phantom power
DFT	Discrete Fourier Transform
DHCP	dynamic host configuration protocol
DHS	digital home standard
DI	directivity index
DIN	Deutsche Industrie Normenausschuss
DIP	dual in-line package
DLP	digital light processing
DMA	direct memory access
DMD	digital micromirror device
DMM	direct metal mastering
DMT	discrete multitone
DNR	dynamic noise reduction
DOCSIS	Data Over Cable Service Interface Specification
DOD	depth of discharge
DoS	denial of service
DPC	deferred procedure calls
DPMS	display power management signaling
DPP	digital phantom power
DRC	digital room correction
DRM	digital rights management
DSB	direct satellite broadcast
DSB	double sideband

DSD	direct stream digital
DSL	Digital Subscriber Line
DSP	digital signal processing
DSS	digital satellite system
DST	direct stream transfer
DSV	digital sum value
ΔT	temperature differential
DTH	direct to home
DTL	direct time lock
DTLe	enhanced direct time lock
DTMF	dual-tone multifrequency
DTS	Digital Theater Systems
DTV	digital TV
DTVM	differential thermocouple voltmeter
DUT	device under test
DV	digital video
DVB	digital video broadcasting
DVD	digital versatile disc
DVI	Digital Video Interface
DVM	digital voltmeter
DVS	descriptive video service
DWDM	dense wave division multiplexing
dyn	dyne
E	voltage (electromotive force)
E.I.	electronic iris
EAD	equivalent acoustic distance
EBS	Emergency Broadcast System
EBU	European Broadcasting Union
ECC RAM	error checking and correcting random-access memory
ECM	electromagnetic compatibility
EDA	equipment distribution area

EDC	electronic dispersion compensation
EDID	Extended Display Identification Data
EDLC	electric double layer capacitor
EDO RAM	extended data out RAM
eDP	embedded DisplayPort
EDP	electronic data processing
EDTV	Enhanced Definition Television
EFC	energy frequency curve
E_{ff}	plate efficiency
EFM	eight-to-fourteen modulation
EFP	Electronic Field Production
EHF	extremely high frequency
EHV	extra-high voltage
EIA	Electronic Industries Alliance, Electronic Industries Association (obsolete)
EIDE	enhanced IDE
EIN	equivalent input noise
EKG	electrocardiograph
ELF	extremely low frequency
ELFEXT	equal level far end crosstalk
EMC	electromagnetic compatibility
EMD	electronic music distribution
emf	electromotive force
EMI	electromagnetic interference
EMP	electromagnetic pulse
emr	electromagnetic radiation
EMR	electromechanical relay
EMT	electrical metallic tubing
EMU	electromagnetic unit
En	noise voltage
EN	European Norm

ENG	electronic news gathering
Eo	open-circuit voltage
EOC	Emergency Operations Center
E_{OUT}	output voltage
E_p	plate voltage
EPA	Environmental Protection Agency
EPR	electrical power required
EQ	equalizer
ER	equipment room
ERB	equivalent rectangular bandwidth
ESD	electrostatic discharge
ESL	equivalent series inductance
ESR	equivalent series resistance
ESU	electrostatic unit
ETC	energy time curve
eV	electron volt
EVOM	electronic volt-ohm meter
f	frequency; force
F	farad
FAQ	frequently asked question
FAS	fire alarm and signal cable
f_{BP}	bandpass in hertz
fc	footcandle
FCC	Federal Communications Commission
FCFC	conductor flat cable
FCoE	Fiber Channel over Ethernet
FCSA	frame check sequence
FD	finite difference
FDDI	fiber data distributed interface
FDF	fiber distribution frame
FDFT	Fast Discrete Fourier Transform
FDTD	finite-difference time-domain

FEC	forward error correction
FET	field-effect transistor
FEXT	far end crosstalk
FFT	Fast Fourier Transform
FIP	function indicator panel
FIR	finite impulse response
fL	footlambert
FLPs	fast link pulses
FM	frequency modulation
FMG	foreground music
FO	fiber optics
FOC	fiber optic connector
FOLED	flexible OLED
FOTL	fiber optic transmission loss
FPGA	field programmable gate array
FRC	fractional part of
FS	full scale
FSK	frequency-shift keying
FSM	feedback stability margin
ft-pdl	foot poundal
ft·lbf	foot pound-force
ft/[′]	foot
ft/min	foot per minute
ft/s	foot per second
ft/s^2	foot per second squared
ft^3/min	cubic foot per minute
ft^3/s	cubic foot per second
FTC	frequency time curve
FTP	file transfer protocol, foil twisted pair
FTTC	fiber to the curb
FTTH	fiber to the home
g	gram

G	gauss
gal	gallon
gal/min	gallon per minute
Gb	gilbert
GeV	gigaelectronvolt
GHz	gigacycle per second, gigahertz
G_M	EIA microphone sensitivity rating
GMT	Greenwich Mean Time
GND	ground
GOF	glass optical fibers
GSA	General Services Administration
GWB	gypsum wallboard
h	hour
H	henry
HANA	Home Automation and Networking Association
HANs	home area networks
H_c	coercive force
HCP	horizontal connection point
HD-SDI	high definition – serial digital interface
HDAs	horizontal distribution areas
HDCP	high-bandwidth digital content protection
HDLCS	high-density linear converter system
HDMI	High Definition Multimedia Interface
HDSL	high bit-rate digital subscriber line
HDTV	high-definition television
HF	high frequency
HFC	hybrid fiber/coaxial
HIPPI	high-speed parallel network technology
HLAA	Hearing Loss Association of America
HOW	house of worship
hp	horsepower

HPF	high-pass filter
HRTF	Head Related Transfer Function
HSCDS	high-speed cable data service
HTML	hypertext markup language
HTTP	hypertext transfer protocol
HV	high voltage
HVAC	heating, ventilating, and air conditioning
Hz	hertz, cycle per second
I/O	input-output
Ia	acoustic intensity
IaaS	Infrastructure as a Service
Iac	ac current
IACC	interaural cross-correlation coefficient
IBC	International Building Code
IC	integrated circuit
ICAT	independent consultants in audiovisual technology
ICEA	Insulated Cable Engineers Association
ICIA	International Communication Industries Association
ID	inside diameter
Idc	dc current
IDC	insulation displacement connector
IDE	integrated device electronics
IDFT	inverse discrete Fourier transform
IDP	integrated detectors/preamplifiers
IDSL	ISDN digital subscriber line
IEC	International Electrotechnical Commission, integrated electronic component
IEEE	Institute of Electrical and Electronic Engineers

IETF	Internet Engineering Task Force
IF	intermediate frequency
IFB	interrupted feedback (foldback)
IGBT	insulated gate bipolar transistor
IGFET	insulated gate field effect transistor
IGMP	internet group management protocol
IGT	insulated gate transistor
IHC	inner hair cells
IIC	Impact Insulation Class
IID	interaural intensity difference
IIR	infinite impulse response
ILD	interaural level difference, injection laser diode
IM or IMD	intermodulation distortion, intermodulation
IMAX	ImageMAXimum
IMG	index matching gel
in, [″]	inch
in/s	inch per second
in^2	square inch
in^3	cubic inch
INMS	integrated network management system
INOVP	input overvoltage protection
IOC	integrated optical components
IOR	index of refraction
I^p	plate current
IP	internet protocol
IPCDN	IP over Cable Data Network
IPD	interaural phase difference
IPM™	intelligent power management system
IR	impulse response, infrared, insulation resistance

IRE	Institute of Radio Engineers
IRS	ignition radiation suppression
ISB	independent sideband
ISD	initial signal delay
ISDN	Integrated Services Digital Network
ISL	inverse square law
ISO	International Organization for Standardization
ISP	internet service provider
ISRC	International Standard Recording Code
ITD	interaural time difference
ITDG	initial time delay gap
ITE	in the ear
ITFS	instructional television fixed service
ITU	International Telecommunication Union
IXC	intermediate cross-connect
J	joule
J/K	joule per kelvin
JFET	junction field effect transistor
JND	just noticeable difference
JPEG	Joint Photographic Experts Group
JSA	Japanese Standards Association
K	kelvin
kcmil	thousand circular mils
keV	1000 electron volts, kiloelectronvolt
kg	kilogram
kG	kilogauss
kgf	kilogram-force
kHz	kilocycle per second, kilohertz
kJ	kilojoule
km	kilometer
km/h	kilometer per hour

kn	knot
kV	kilovolt (1000 volts)
kVA	kilovolt-ampere
kvar	kilovar
kO	kilohm
kW	kilowatt
kWh	kilowatthour
l	liter
L	inductance, inductor, lambert
l/s	liter per second
L_a	vibratory acceleration level
LAN	local area network
LAS	large area systems
LASER	light amplification by stimulated emission of radiation
lb	pound
lb·fft	pound-force foot
lbf	pound-force
lbf/in^2, psi	pound (force) per square inch. Although the use of the abbreviation psi is common, it is not recommended.
LC	inductance-capacitance
LCD	liquid crystal display
LCoS	liquid crystal on silicon
LCRS	Left, Center, Right, Surround
L_D	direct sound pressure level
LD	laser diode
LDR	light dependent resistor
L_E	energy level
LE	lateral efficiency
LEC	local exchange carrier
LED	light emitting diode

LEDE	live end–dead end
LEDR	Listening Environment Diagnostic Recording
LEED	Leadership in Energy and Environmental Design
LF	low frequency, lateral fraction
L_F	vibratory force level
LFE	low-frequency effects
LFE	low frequency effects
L_I	intensity level
lm	lumen
lm.s	lumen second
lm/ft^2	lumen per square foot
lm/m^2	lumen per square meter
lm/W	lumen per watt
LMDS	Local Multipoint Distribution Service
ln	logarithm, natural
log	logarithm
L_{out}	output level in dB
LP	long play
L_p, SPL	sound pressure level
LPF	low-pass filter
LPRS	Low Power Radio Services
L_R	reverberant sound level in dB
LSB	least significant bit, lower sideband
L_{sens}	loudspeaker sensitivity
LSHI	large-scale hybrid integration
LSI	large-scale integration
L_T	total sound level in dB
LTI	linear time invariant
LUT	look up table
Lv	vibratory velocity level

LVDS	low-voltage differential signaling
L_W	energy density level
L_w, dB-PWL	power level
lx	lux
m	meter, milli
M	mutual inductance, wavelength
M-JPEG	motion JPEG
m(F)	modulation reduction factor
m^2	square meter
M2M	machine-to-machine
m^3	cubic meter
m^3/s	cubic meter per second
mA	milliampere
MAC	media access control, multiplier/accumulator
MACs	moves, adds, and changes
MADI	multichannel audio digital interface
MAN	metropolitan area network
MAP	manufacturing automation protocol
MAS	medium area systems
MASH	multistage noise shaping
MATV	master antenna television
mb	millibarn
MB	megabyte
mbar	millibar
Mbps	megabits per second
MC	moving coil
MCNS	Multimedia Cable Network System Partners Ltd
MCR	multichannel reverberation
MCSE	Microsoft Certified Solutions Expert
MD	minidisc

MDA	main distribution areas, motion drive amplifier
MDA	Multi-Dimensional Audio
MDA	Multi-Dimensional Audio
mDP	mini DisplayPort
MDS	multipoint distribution system
MEMS	microelectromechanical systems
MeV	megaelectronvolt
MF	medium frequency
MFP	mean free path
mg	milligram
mGal	milligal
mH	millihenry
MHD	magneto hydrodynamics
MHz	megahertz, megacycle per second
mi	mile (statute)
mi/h	mile per hour
mic	microphone
MIDI	Musical Instrument Digital Interface
MIMO	multiple-in/multiple-out
min	minute (time)
MKS	meter-kilogram-second
ml	milliliter
mm	millimeter
MMB	Mass Media Bureau
MMF	magnetomotive force
mmHg	millimeter of mercury, conventional
MO	magneto-optics
mol	mole
MOR	magneto-optical recording
MOS	metal-oxide semiconductor
MOSFET	metal-oxide semiconductor field-effect

	transistor
MOV	metal-oxide varistor
MPEG	Motion Picture Experts Group
MRT	modified rhyme test
ms	millisecond
mS	millisiemens
MSB	most significant bit
MSO	multiple system operator
MTC	midi time code
MTF	modulation transfer function
MTP	mail transfer protocol
MUTOA	multiuser telecommunications outlet assembly
mV	millivolt
MV	megavolt
mW	milliwatt
MW	megawatt
MΩ	megohm
Mx	maxwell
MXC	main cross connect
N	newton
N·m	newton meter
N/D	neutral density filter
N/m^2	newton per square meter
nA	nanoampere
NA	numerical aperture
NAB	National Association of Broadcasters
NAD	National Association of the Deaf
NAG	needed acoustic gain
NAT	network address translation
NBR	Butadiene-acrylonitrile copolymer rubber
NBS	National Bureau of Standards

NEC	National Electrical Code
NECA	National Electrical Contractors Association
NEMA	National Electrical Manufacturers Association
NEP	noise equivalent power
NEXT	near end cross talk
nF	nanofarad
NF	noise figure; noise frequency
NFPA	National Fire Protection Association
NIC	network interface cards
NICAM	near-instantaneous companding
NIOSH	National Institute Of Occupational Safety And Health
NIST	National Institute of Standards and Technology
NLPs	normal link pulses
nm	nanometer
NOALA	noise-operated automatic level adjuster
NOC	network operations center
NOM	number of open microphone
Np	neper
NPN	negative-positive-negative
NRC	noise reduction coefficient
NRZI	non-return-to-zero inverted
ns	nanosecond
NSCA	National Systems Contractors Association
NTC	negative temperature coefficient
NTSC	National Television System Committee
nW	nanowatt
OC	optical carrier
OCIA	opposed current interleaved amplifier
OCMR	optimized common mode rejection

OCP	overcurrent protection
OD	outside diameter
Oe	oersted
OEIC	optoelectronic integrated circuit
OFDM	orthogonal frequency division multiplexing
OFHC	oxygen-free, high-conductivity copper
OHC	outer hair cells
OITC	Outdoor-Indoor Transmission Class
OLED	organic LED, organic light emitting diode
OLT	optical line terminal
ONTs	optical network terminals
OPC	optimum power calibration
ORTF	Office de Radiodiffusion Television Francois
OSI	open system interconnect
OSI	optimum source impedance
OSM™	on-screen manager
OSS	operations support systems
OTA	operational transconductance amplifier
OTDR	optical time domain reflectometer
oz	ounce (avoirdupois)
pA	picoampere
Pa	pascal
PaaS	Platform as a Service
PAG	potential acoustic gain
PAL	Phase Alternating Line, precision audio link
PAM	pulse-amplitude modulation
PAR	pinna acoustic response, parabolic aluminized reflector
PASC	precision adaptive subband coding
P_{avg}	average power

PB	phonemically balanced
PBO	power backoff
PBS	polarization beam splitter
PBX	private branch exchange
PC	printed circuit
PCA	power calibration area
PCM	pulse code modulation
PD	powered devices
pdl	poundal
PDM	pulse density modulation, pulse-duration modulation
PDP	plasma
PEM	pulse end modulation
PETC	Professional Education and Training Committee
pF	picofarad
PF	power factor
PFC	phase frequency curve, power factor correction
PFL	prefade listen
PFM	pulse-frequency-modulation
PGA	programmable gate array
PIC	plastic insulated conductor
PiMF	pairs in metal foil
PIN	positive, intrinsic, negative
PIP	picture in picture
piv	peak-inverse-voltage
PLD	programmable logic device
PLL	phase locked loop
PLS	personal listening systems
PM	phase modulation
PMD	physical medium dependent

PNP	positive-negative-positive
P°	power out
PoE	power over ethernet
PONs	passive optical networks
POTS	plain old telephone service
PPM	peak program meter, pulse-position modulation
PPP	point-to-point protocol
PRF	pulse-repetition frequency
PRI	primary rate interface ISDN
PRIR	parametric room impulse response
PROM	programmable read-only memory
PRR	pulse-repetition rate
ps	picosecond
PS-NEXT	power-sum near-end crosstalk
PSAELFEXT	power sum alien ELFEXT

PSAELFEXT power sum alien Equal Level Far-End crosstalk

PSANEXT	Power sum alien near-end crosstalk
PSE	power sourcing equipment
PSTN	public switched telephone network
PTM	pulse-time modulation
PTS	permanent threshold shift
PTZ	pan/tilt/zoom
PU	pickup
pW	picowatt
PWM	pulse-width modulation, pulse-width modulator
Q	directivity factor, quality factor
QAM	quandrature amplitude modulation
QDR	quad data rate

QoS	quality of service
QPSK	quaternary phase shift keying
QRD	quadratic residue diffuser
QWP	quarter wave plate
R	resistor, roentgen
R-DAT	rotary head digital audio tape
r/min, rpm	revolution per minute
r/s, rps	revolution per second
rad	radian
RADIUS	remote authentication dial-in user service
RAID	redundant array of independent disks
RAM	random access memory
RASTI	rapid speech transmission index
RC	resistance-capacitance, resistor-capacitor
RCDD	registered communication distribution designer
RCFC	round conductor flat cable
rd	rad
RDC	Regional Data Center
RDS	Radio Data Service, Radio Data System
R_{eq}	equivalent resistance
RF	radio frequency
RFI	radio-frequency interference
RFID	radio frequency identification
RFP	request for proposals
RFZ	reflection-free zone
RGB	red, green, blue
RIAA	Recording Industry Association of America
RIR	Room Impulse Response
RL	return loss
RLC	resistance-inductance-capacitance
R_M	matched resistance

RMA	Recording Management Area
rms	root-mean-square
RNG	random-noise generator
ROC	report on comments
RoHS	restriction of hazardous substances
ROM	read only memory
ROP	report on proposal
r_p	plate resistance
RPG	Reflection-Phase Gratings
RPS	reflections per second
R_s	source resistance
RSN	robust service network
RSVP	resource reservation protocol
RT, RT_{60}	reverberation time
RTA	real-time analyzer
RTL	resistor-transistor logic
RTP	real-time protocol
s	second (time)
S	siemens, total surface area
S-CDMA	synchronous code division multiple access
S-HDSL	single-pair high bit-rate digital subscriber line
S/H	sample and hold
Sa	room constant
SAA	sound absorption average
SaaS	Software as a Service
Sabin	unit of absorption
SACD	super audio CD
SAG	Screen Actors Guild
SAN	storage area network
SAP	second audio program

SAVVI	sound, audiovisual, and video integrators
SC	supercapacitor
SCA	Subsidiary Communications Authorization
SCIN	shield current induced noise
SCMS	serial copy management system
SCP	simple control protocol
SCR	silicon controlled rectifier
SCS	structured cabling system
SCSI	small computer system
ScTP	screened twisted pair
SCTP	Stream Control Transmission Protocol
SD	signal delay
SD-SDI	standard definition – serial digital interface
SDDS	Sony Dynamic Digital Sound
SDI	serial digital interface
SDMI	secure digital music initiative
SDR	single data rate
SDSL	symmetric digital subscriber line
SDTI	Serial Data Transport Interface
SDTV	standard definition television
SDV	serial digital video
SECAM	Sequence Electronique Couleur Avec Memoire
sensi	sensitivity
SHF	super-high frequency
SIP	single in-line package, Session Initiation Protocol
SLD	super-luminescent diode
SLM	sound level meter
SMA	subminiature A connector
SMART	self-monitoring analysis and reporting

	technology
SMB	subminiature B connector
SMC	subminiature C connector
SMPTE	Society of Motion Picture & Television Engineers
SNG	satellite news gathering
SNIR	signal-to-noise plus interference ratio
SNMP	simple network management protocol
SNR	signal-to-noise ratio
SoCs	systems-on-chips
SONET	synchronous optical network
SPDT	single-pole, double-throw
SPP	song position pointer
SPST	single-pole, single-throw
sr	steradian
SRAM	static RAM
SRC	sample-rate convertor
SRI	Sound Reduction Index
SRL	structural return loss
SRS	Sample-Rate Converter
SSB	single sideband
SSID	service station identifier
SSL	solid state lighting
SSM	solid state music
SSR	solid state relay
STC	sound transmission class
STI	speech transmission index
STP	shielded twisted pair(s)
STS	static transfer system
Super VHS	Super Video Home System
SW	short wave
SWG	stubs wire gage

SWR	standing-wave ratio
t	tonne or metric ton
T	tesla, time
TBC	timebase corrector
TC	thermocouple; time constant, telcom closet
TCP	Transmission Control Protocol
TCP/IP	transmission control protocol/internet protocol
TDM	time division multiplexing
TDMA	time division multiple access
TDS	time delay spectrometry
TE	telecommunications enclosure
TEF	time energy frequency
TEM	transverse electromagnetic
TFE	tetrafluoroethylene
TFT	thin film transistors
TGB	telecommunications grounding busbar
THD	total harmonic distortion
THD+N	total harmonic distortion plus noise
TIA	Telecommunications Industry Association
TIM	transient intermodulation distortion
TL	transmission loss
TM	transverse magnetic
TMDS	Transition Minimized Differential Signaling
TMV	transient minimum voltage
TN, i_{tn}	thermal noise
TOGAD	tone operated gain adjusting device
TOLED	transparent OLED
TP-PMD	twisted pair-physical medium dependent
TR	telecommunications room
TTL	transistor-transistor logic

TTS	temporary threshold shift
TV	television
TVI	television interference
TVRO	TV receive only
TVS	transient voltage suppression
TWT	traveling-wave tube
u	atomic mass unit (unified)
UBC	Uniform Building Code
UC	ultracapacitor
UDF	universal disc format
UDP	user datagram protocol
UDP	user defined protocol
UHF	ultrahigh frequency
UI	unit interval
UL	Underwriters Laboratories, Inc.
UPA	Universal Powerline Association
USB	universal serial bus, upper sideband
USOC	universal service order code
UTP	unshielded twisted pair(s)
UV	ultraviolet, unit interval
V	volt
VA	voltampere
Vac	ac volt
VC/MTM	variable constellation/multitone modulation
VCA	voltage-controlled amplifier
VCO	voltage-controlled oscillator
VCSEL	vertical-cavity surface-emitting laser
VCXO	voltage controlled crystal oscillator
V_{dc}	dc voltage
V_{dc}	direct current volts

VDSL	very high bit rate digital subscriber line
VESA	Video Electronics Standards Association
VFO	variable-frequency oscillator
VGA	video graphics array
VHF	very high frequency
VHS	video home system
VI	volume indicator
VLAN	virtual local area network
VLF	very low frequency
VLSI	very large scale integration
VOD, VoD	video on demand
VoIP	voice over internet protocol
VOM	volt-ohm-milliammeter
VoWi-Fi	voice over wireless fidelity
VP	velocity of propagation
VPN	virtual private networks
VRAM	video RAM
Vrms	root-mean-square voltage
VSO	variable speed oscillator
VSWR	voltage standing wave ratio
VTF	vertical tracking force
VTVM	vacuum-tube voltmeter
VU	volume unit
Ω	ohm
W	watt
W/(sr·m^2)	watt per steradian square meter
W/sr	watt per steradian
WAN	wide area network
WAP	wireless access point, wireless application protocol
Wb	weber
WCS	wireless communications service

WDM	wavelength division multiplexing
WEP	wired equivalent privacy
Wh	watt hour
WHO	World Health Organization
WiFi	wireless fidelity
WiMax	wireless microwave access
WMA	windows media audio
WMTF	weighted modulation transmission function
WO	write once
WORM	write once read many
X	reactance
X_C	capacitive reactance
X_L	inductive reactance
Y	admittance
Y	ripple factor
yd	yard
yd^2	square yard
yd^3	cubic yard
Z	impedance (magnitude)
ZDA	zone distribution area
μ	amplification factor, voltage gain
μA	microampere
μbar	microbar
μF	microfarad
μg	microgram
μH	microhenry
μm	micrometer
μmho	micromho
μs	microsecond
μS	microsiemens
μV	microvolt

μW	microwatt

50.12 Audio Frequency Range

The audio spectrum is usually considered the frequency range between 20Hz and 20kHz, Fig. 50-9. In reality, the upper limit of hearing pure tones is between 12kHz and 18kHz, depending on the person's age and sex and how well the ears have been protected against loud sounds. Frequencies above 20 kHz cannot be heard as a sound, but the effect created by such frequencies (i.e., rapid rise time) can be heard.

50.13 Surface Area and Volume Equations

To find the *surface area and volume* of complex areas, the area can often be divided into a series of simpler areas and handled one at a time. Figs. 50-10 to 50-17 are equations for various and unusual volumes.

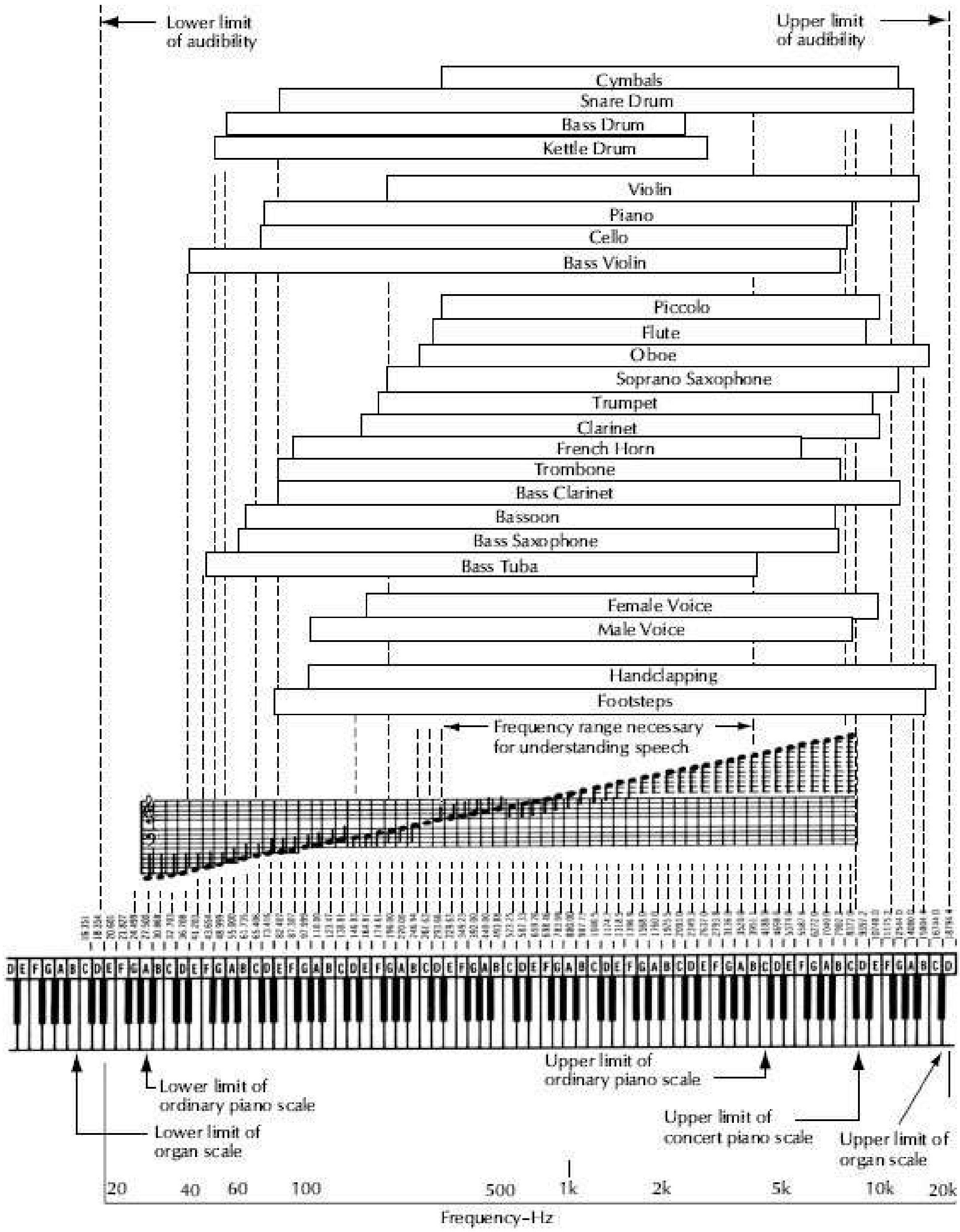


Figure 50-9. Audible frequency range.

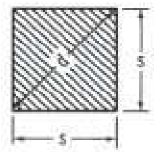
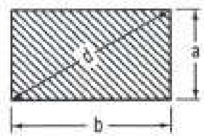
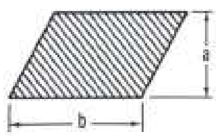
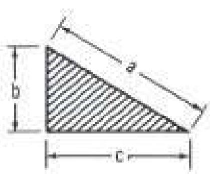
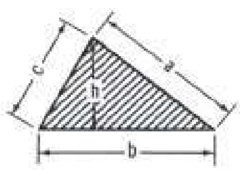
 Square	A = area. $A = s^2$ $A = \frac{1}{2} d^2$ $s = 0.7071\ d = \sqrt{A}$ $d = 1.414\ s = 1.414 \sqrt{A}$
 Rectangle	A = area. $A = ab$ $A = a\sqrt{d^2-a^2} = b\sqrt{d^2-b^2}$ $d = \sqrt{a^2+b^2}$ $a = \sqrt{d^2-b^2} = A \div b$ $b = \sqrt{d^2-a^2} = A \div a$
 Parallelogram	A = area. $A = ab$ $a = A \div b$ $b = A \div a$ Note that dimension a is measured at right angles to line b.
 Right-angled triangle	A = area. $A = \frac{bc}{2}$ $a = \sqrt{b^2+c^2}$ $b = \sqrt{a^2-c^2}$ $c = \sqrt{a^2-b^2}$
 Acute-angled triangle	A = area. $A = \frac{bh}{2} = \frac{b}{2}\sqrt{a^2 - \left(\frac{a^2+b^2-c^2}{2b}\right)^2}$ If $S = \frac{1}{2}(a+b+c)$, then $A = \sqrt{S(S-a)(S-b)(S-c)}$

Figure 50-10. Equations for finding surface areas for complex shapes.

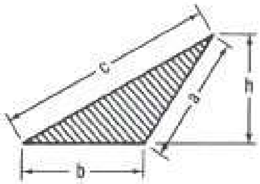
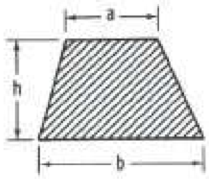
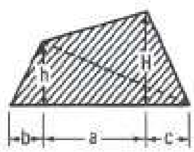
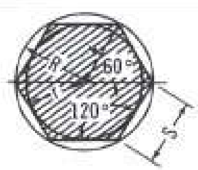
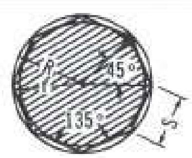
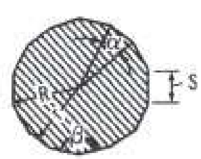
 Obtuse-angled triangle	A = area. $A = \frac{bh}{2} = \frac{b}{2}\sqrt{a^2 - \left(\frac{c^2-a^2-b^2}{2b}\right)^2}$ If $S = \frac{1}{2}(a+b+c)$, then $A = \sqrt{S(S-a)(S-b)(S-c)}$
 Trapezoid	A = area. $A = \frac{(a+b)h}{2}$
 Trapezium	A = area. $A = \frac{(H+h)a+bh+cH}{2}$ A trapezium can also be divided into two triangles as indicated by the dotted line. The area of each of these triangles is computed, and the results added to find the area of the trapezium.
 Regular hexagon	A = area; R = radius of circumscribed circle; r = radius of inscribed circle. $A = 2.598\ s^2 = 2.598\ R^2 = 3.464\ r^2$ $R = s = 1.1155r$ $r = 0.866\ s = 0.866\ R$ $s = R = 1.1155\ r$
 Regular octagon	A = area; R = radius of circumscribed circle; r = radius of inscribed circle; $A = 4.828\ s^2 = 2.828\ R^2 = 3.314\ r^2$ $R = 1.307\ s = 1.082\ r$ $r = 1.207\ s = 0.924\ R$ $s = 0.765\ R = 0.828\ r$
 Regular polygon	A = area; n = number of sides. $\alpha = 360° \div n$ $\beta = 180° - \alpha$ $A = \frac{nsr}{2} = \frac{ns}{2}\sqrt{R^2 - \frac{s^2}{4}}$ $R = \sqrt{r^2 + \frac{s^2}{4}}$; $r = \sqrt{R^2 - \frac{s^2}{4}}$; $s = 2\sqrt{R^2 - r^2}$

Figure 50-11. Equations for finding surface areas for complex shapes.

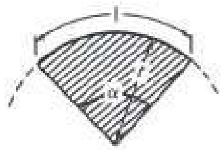
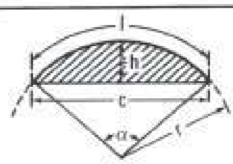
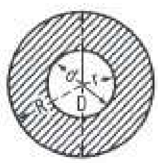
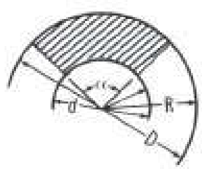
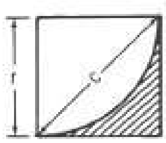
 Circle	A = area; C = circumference. $A = \pi r^2 = 3.1416 r^2 = 0.7854\ d^2$ $C = 2\pi r = 6.2832\ r = 3.1416\ d$ $r = C \div 6.2832 = \sqrt{A \div 3.1416} = 0.564\sqrt{A}$ $d = C \div 3.1416 = \sqrt{A \div 0.7854} = 1.128\sqrt{A}$ Length of arc for center-angle of 1° = 0.008727 d Length of arc for center-angle of n° = 0.008727 nd
 Circular sector	A = area; l = length of arc; α = angle, in degrees. $l = \frac{r \times \alpha \times 3.1416}{180} = 0.01745\ r\alpha = \frac{2A}{r}$ $A = \frac{1}{2} rl = 0.008727\ \alpha r^2$ $\alpha = \frac{57.296\ l}{r}$ $\quad r = \frac{2A}{l} = \frac{56.296\ l}{\alpha}$
 Circular segment	A = area; l = length of arc; α = angle, in degrees. $c = 2\sqrt{h(2r-h)}$ $\quad A = \frac{1}{2}[rl - c(r-h)]$ $r = \frac{c^2 + 4h^2}{8h}$ $\quad l = 0.01745\ r\alpha$ $h = r - \frac{1}{2}\sqrt{4r^2 - c^2}$ $\quad \alpha = \frac{57.296 l}{r}$
 Circular ring	A = area. $A = \pi(R^2 - r^2) = 3.1416\ (R^2 - r^2)$ $= 3.1416\ (R + r)(R - r)$ $= 0.7854\ (D^2 - d^2) = 0.7854\ (D + d)(D - d)$
 Circular ring sector	A = area; α = angle, in degrees. $A = \frac{\alpha\pi}{360}(R^2 - r^2) = 0.00873\alpha\ (R^2 - r^2)$ $= \frac{\alpha\pi}{4 \times 360}(D^2 - d^2) = 0.00218\ \alpha\ (D^2 - d^2)$
 Spandrel or fillet	A = area. $A = r^2 - \frac{\pi r^2}{4} = 0.215\ r^2$ $= 0.1075\ c^2$

Figure 50-12. Equations for finding surface areas for complex shapes.

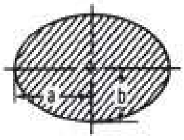
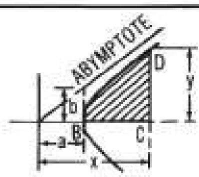
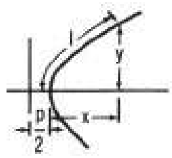
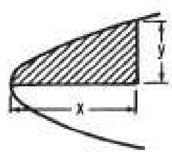
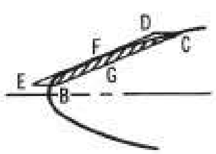
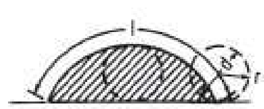
 Ellipse	A = area; P = perimeter or circumference. $A = \pi ab = 3.1416\ ab.$ An approximate formula for the perimeter is: $P = 3.1416\sqrt{2(a^2+b^2)}$
 Hyperbola	A = area BCD $A = \frac{xy}{2} - \frac{ab}{2}\ \text{hyp. log}\left(\frac{x}{a} + \frac{y}{b}\right)$
 Parabola	l = length of arc. $l = \frac{p}{2}\left[\sqrt{\frac{2x}{p}\left(1+\frac{2x}{p}\right)} + \text{hyp. log}\left(\sqrt{\frac{2x}{p}} + 1 + \frac{2x}{p}\right)\right]$ When x is small in proportion to y, the following is a close approximation: $l = y\left[1 + \frac{2}{3}\left(\frac{x}{y}\right)^2 - \frac{2}{5}\left(\frac{x}{y}\right)^4\right]$ or $l = \sqrt{y^2 + \frac{4}{3}x^2}$
 Parabola	A = area. $A = \frac{2}{3}xy$ (The area is equal to two-thirds of the rectangle which has x for its base and y for its height.)
 Segment of Parabola	A = area. Area BFC = $A = \frac{2}{3}$ area of parallelogram BCDE. If FG is the height of the segment, measured at right angles to BC, then: Area of segment BFC = $\frac{2}{3}$ BC×FG
 Cycloid	A = area; l = length of cycloid. $A = 3\pi r^2 = 9.4248\ r^2 = 2.3562\ d^2$ $= 3 \times$ area of generating circle $l = 8r = 4d$

Figure 50-13. Equations for finding surface areas for complex shapes.

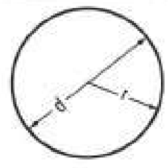
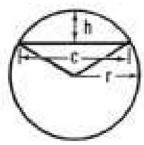
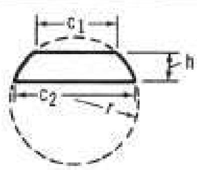
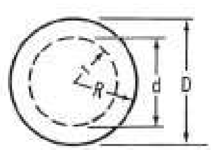
 Sphere	V = volume; A = area of surface. $V = \frac{4\pi r^2}{3} = \frac{\pi d^3}{6} = 4.1888\, r^3 = 0.5236\, d^3$ $A = 4\pi r^2 = \pi d^2 = 12.5664 r^2 = 3.1416\, d^2$ $r = \sqrt[3]{\frac{3V}{4\pi}} = 0.6204 \sqrt{V}$
 Spherical sector	V = volume; A = total area of conical and spherical surface. $V = \frac{2\pi r^2 h}{3} = 2.0944 r^2 h$ $A = 3.1416 r \left(2h + \frac{1}{2} c\right)$ $c = 2\sqrt{h(2r-h)}$
 Spherical segment	V = volume; A = area of spherical surface. $V = 3.1416\, h^2 \left(r - \frac{h}{3}\right) = 3.1416 h \left(\frac{c^2}{8} + \frac{h^2}{6}\right)$ $A = 2\pi rh = 6.2832 rh = 3.1416 \left(\frac{c^2}{4} + h^2\right)$ $c = 2\sqrt{h(2r-h)};\ r = \frac{c^2 + 4h^2}{8h}$
 Spherical zone	V = volume; A = area of spherical surface. $V = 0.5236\, h \left(\frac{3{c_1}^2}{4} + \frac{3{c_2}^2}{4} + h^2\right)$ $A = 2\pi rh = 62832\ rh$ $r = \sqrt{\frac{{c_2}^2}{4} + \left(\frac{{c_2}^2 - {c_1}^2 - 4h^2}{8h}\right)^2}$
 Spherical wedge	V = volume; A = area of spherical surface; α = center angle in degrees. $V = \frac{\alpha}{360} \times \frac{4\pi r^3}{3} = 0.0116\, \alpha r^3$ $A = \frac{\alpha}{360} \times 4\pi r^2 = 0.0349\, \alpha r^2$
 Hollow sphere	V = volume. $V = \frac{4\pi}{3} (R^3 - r^3) = 4.1888\, (R^3 r^3)$ $= \pi\ 6\ (D^3 - d^3) = 0.5236\, (D^3 - d^3)$

Figure 50-14. Equations for finding surface areas for complex shapes.

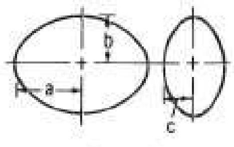
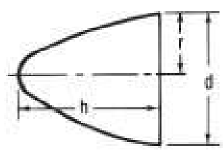
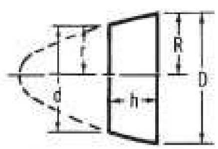
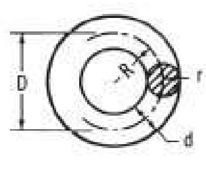
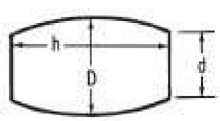
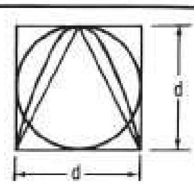
 Ellipsoid	V = volume; A = area of surface. $V = \frac{4\pi}{3}\ abc = 4.1888\ abc$ In an ellipsoid of revolution, or spheroid, where $b = c$: $V = 4.1888\ ab^2$, and $A = \frac{4\pi}{\sqrt{2}}\ b\ \sqrt{a^2+b^2}$
 Paraboloid	V = volume; $V = \frac{1}{2}\pi r^2 h = 0.3927\ d^2 h$ A = area; $A = \frac{2\pi}{3p}\left[\sqrt{\left(\frac{d^2}{4}+p^2\right)^3} - p^3\right]$ in which $p = \frac{d^2}{8h}$
 Parabolidal segment	V = volume. $V = \frac{\pi}{2}\ h\ (R^2+r^2) = 1.5708\ h\ (R^2+r^2)$ $= \frac{\pi}{8}\ h\ (D^2+d^2) = 0.3927\ h\ (D^2+d^2)$
 Torus	V = volume; A = area of surface. $V = 2\pi^2\ Rr^2 = 19.739\ Rr^2$ $= \frac{\pi^2}{4}\ Dd^2 = 2.4674\ Dd^2$ $A = 4\pi^2\ Rr = 39.478\ Rr$ $= \pi^2 Dd = 9.8696\ Dd$
 Barrel	V = approximate volume. If the sides are bent to the arc of a circle: $V = \frac{1}{12}\pi h\ (2D^2+d^2) = 0.262\ h\ (2D^2+d^2)$ If the sides are bent to the arc of a parabola: $V = 0.209\ h\ (2D^2+Dd+\frac{3}{4}d^2)$
	If d = base diameter and height of a cone, a paraboloid and a cylinder, and the diameter of a sphere, then the volumes of these bodies are to each other as below: Cone: paraboloid: sphere: cylinder = $\frac{1}{3} : \frac{1}{2} : \frac{2}{3} : 1$

Figure 50-15. Equations for finding surface areas for complex shapes.

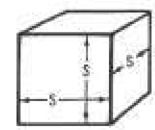
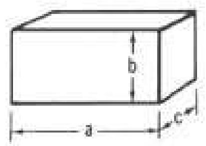
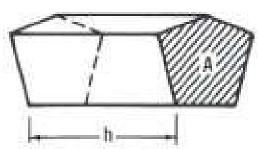
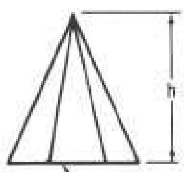
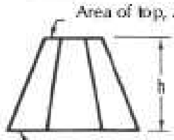
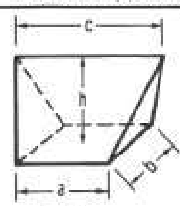
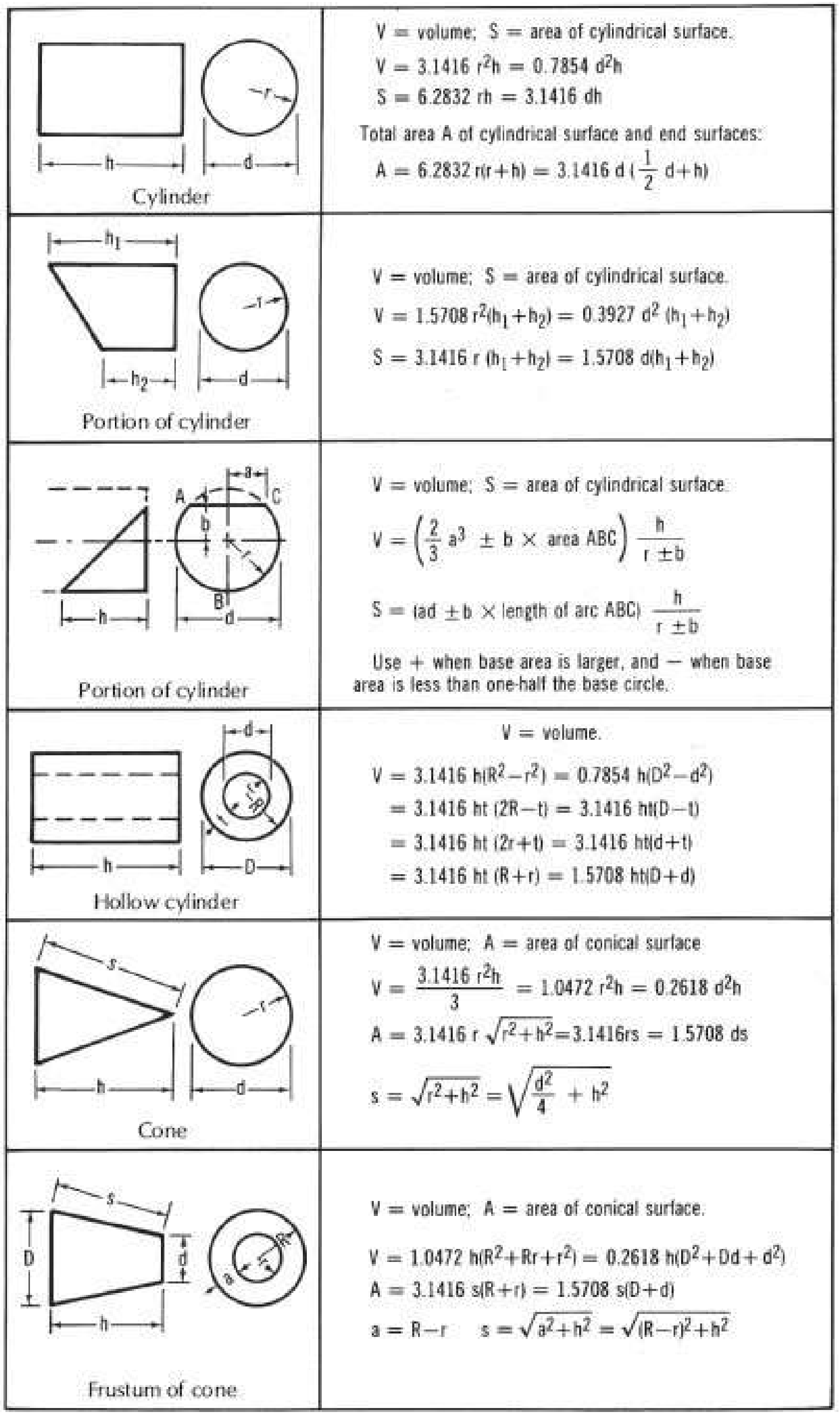
 Cube	V = volume. $V = s^3$ $s = \sqrt[3]{V}$
 Square prism	V = volume. $V = abc$ $a = \dfrac{V}{bc}$ $b = \dfrac{V}{ac}$ $c = \dfrac{V}{ab}$
 Prism	V = volume; A = area of end surface. $V = h \times A$ The area A of the end surface is found by the formulas for areas of plane figures on the preceding pages. Height h must be measured perpendicular to end surface.
 Base area Pyramid	V = volume $V = \dfrac{1}{3} h \times$ area of base. If the base is a regular polygon with n sides, and s = length of side, r = radius of inscribed circle, and R = radius of circumscribed circle, then: $V = \dfrac{nsrh}{6} = \dfrac{nsh}{6}\sqrt{R^2 - \dfrac{s^2}{4}}$
Area of top, A_1  Area of base, A_2 Frustum of pyramid	V = volume. $V = \dfrac{h}{3}(A_1 + A_2 + \sqrt{A_1 \times A_2})$
 Wedge	V = volume. $V = \dfrac{(2a+c)bh}{6}$

Figure 50-16. Equations for finding surface areas for complex shapes.



Cylinder	V = volume; S = area of cylindrical surface. $V = 3.1416\, r^2h = 0.7854\, d^2h$ $S = 6.2832\, rh = 3.1416\, dh$ Total area A of cylindrical surface and end surfaces: $A = 6.2832\, r(r+h) = 3.1416\, d\left(\frac{1}{2}d+h\right)$
Portion of cylinder	V = volume; S = area of cylindrical surface. $V = 1.5708\, r^2(h_1+h_2) = 0.3927\, d^2(h_1+h_2)$ $S = 3.1416\, r(h_1+h_2) = 1.5708\, d(h_1+h_2)$
Portion of cylinder	V = volume; S = area of cylindrical surface. $V = \left(\frac{2}{3}a^3 \pm b \times \text{area } ABC\right)\frac{h}{r \pm b}$ $S = (ad \pm b \times \text{length of arc } ABC)\frac{h}{r \pm b}$ Use + when base area is larger, and − when base area is less than one-half the base circle.
Hollow cylinder	V = volume. $V = 3.1416\, h(R^2-r^2) = 0.7854\, h(D^2-d^2)$ $= 3.1416\, ht(2R-t) = 3.1416\, ht(D-t)$ $= 3.1416\, ht(2r+t) = 3.1416\, ht(d+t)$ $= 3.1416\, ht(R+r) = 1.5708\, ht(D+d)$
Cone	V = volume; A = area of conical surface $V = \frac{3.1416\, r^2h}{3} = 1.0472\, r^2h = 0.2618\, d^2h$ $A = 3.1416\, r\sqrt{r^2+h^2} = 3.1416 rs = 1.5708\, ds$ $s = \sqrt{r^2+h^2} = \sqrt{\frac{d^2}{4}+h^2}$
Frustum of cone	V = volume; A = area of conical surface. $V = 1.0472\, h(R^2+Rr+r^2) = 0.2618\, h(D^2+Dd+d^2)$ $A = 3.1416\, s(R+r) = 1.5708\, s(D+d)$ $a = R-r \quad s = \sqrt{a^2+h^2} = \sqrt{(R-r)^2+h^2}$

Figure 50-17. Equations for finding surface areas for complex shapes.

Bibliography

Metric Units of Measure and Style Guide, US Metric Association, Inc., Sugarloaf Star Route, Boulder, CO 80302.

V. Antoine, *Guidance for Using the Metric System* (SI Version), Society for Technical Communication, 1010 Vermont Avenue, NW, Washington, DC 20005.

D. Davis, "Synergetic Audio Concepts," Vol. 5, no. 2, October 1977.

John L. Feirer, *SI Metric Handbook*, The Metric Company.

Handbook of Electronics Tables and Formulas, 6th Edition, Carmel IN: SAMS, a division of Macmillan Computer Publishing, 1986.

D. Bohn, "The Bewildering Wilderness," *S&VC*, September 2000.

Index

Symbols

Numerics

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z