THIRD EDITION

SOUND SYSTEMS: DESIGN AND OPTIMIZATION

MODERN TECHNIQUES AND TOOLS FOR SOUND SYSTEM DESIGN AND ALIGNMENT

BOB McCARTHY

Sound Systems: Design and Optimization

Third Edition

Sound Systems: Design and Optimization provides an accessible and unique perspective on the behavior of sound systems in the practical world. The third edition reflects current trends in the audio field thereby providing readers with the newest methodologies and techniques.

In this greatly expanded new edition, you'll find clearer explanations, a more streamlined organization, increased coverage of current technologies and comprehensive case studies of the author's award-winning work in the field.

As the only book devoted exclusively to modern tools and techniques in this emerging field, *Sound Systems: Design and Optimization* provides the specialized guidance needed to perfect your design skills.

This book helps you:

- Improve your design and optimization decisions by understanding how audiences perceive reinforced sound.
- Use modern analyzers and prediction programs to select speaker placement, equalization, delay and level settings based on how loudspeakers interact in the space.
- Define speaker array configurations and design strategies that maximize the potential for spatial uniformity.
- Gain a comprehensive understanding of the tools and techniques required to generate a design that will create a successful transmission/reception model.

Bob McCarthy is the Director of System Optimization at Meyer Sound and President of Alignment & Design, Inc. As a developer of FFT analysis systems, he pioneered the methods for tuning modern speaker systems that have since become standard practice in the industry. He is the foremost educator in the field of sound system optimization and has conducted training courses worldwide for over thirty years. Bob received the USITT Distinguished Achiever in Sound Design in 2014. His clients have included esteemed companies such as Cirque du Soleil and Walt Disney Entertainment, as well as many of the world's best sound designers, including Jonathan Deans, Tony Meola, Andrew Bruce and Tom Clark.

This page intentionally left blank

Sound Systems: Design and Optimization

Modern Techniques and Tools for Sound System Design and Alignment

Third Edition

Bob McCarthy

Focal Press
Taylor & Francis Group
NEW YORK AND LONDON

Third edition published 2016
by Focal Press
711 Third Avenue, New York, NY 10017

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.

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

First published XXX by XX
Second edition published XXX by XX

Library of Congress Cataloging-in-Publication Data
McCarthy, Bob.
Sound systems : design and optimization : modern techniques and tools for sound system design and alignment / by Bob McCarthy. — Third edition.
pages cm
Includes bibliographical references and index.
ISBN 978-0-415-73099-0 (hardback) — ISBN 978-0-415-73101-0 (pbk.) — ISBN 978-1-315-84984-3 (ebk) 1. Sound—Recording and reproducing. I. Title.
TK7881.4.M42 2016
621.389'3—dc23
2015016842

ISBN: 978-0-415-73101-0 (pbk)
ISBN: 978-0-415-73099-0 (hbk)
ISBN: 978-1-315-84984-3 (ebk)

Typeset in Giovanni
By Apex CoVantage, LLC

Dedication

To the woman who knew me back when this journey began, and is still there through three editions, the love of my life,
Merridith.

In memoriam

During the course of writing the first edition our field lost one of its most well-loved and respected pioneers, Don Pearson a.k.a. Dr Don. I had the good fortune to work with Don and receive his wisdom. He was there when it all began and is still missed. More recently we lost Tom Young, and Mike Shannon who contributed so much to the industry and to me personally. And finally, I lost my brother Chris who was an inspiration to all who knew him.

CONTENTS

PART II • Design

PART III • Optimization

This page intentionally left blank

PREFACE

This book is about a journey. On the one hand, the subject is the journey of sound as it travels through a sound system, then through the air, and inevitably to a listener. It is also a personal journey, my own quest to understand the complicated nature of this sound transmission. The body of this text will detail the strictly technical side of things. First, however, I offer you some of the personal side.

I was supposed to build buildings. Unbeknownst to me at the time, this calling was derailed on February 9, 1964 by the appearance of the Beatles on *The Ed Sullivan Show*. Like so many of my generation, this landmark event brought popular music and an electric guitar into my life. I became a great enthusiast of live concerts, which I regularly attended throughout my youth at any chance presented. For years, it remained my expectation that I would enter the family construction business. This vision ended on a racetrack in Des Moines, Iowa on June 16, 1974. The experience of hearing the massive sound system at this Grateful Dead concert set my life in a new direction. On that day I made the decision that I was going to work in live concert sound. I wanted to help create this type of experience for others. I would be a mix engineer and my dream was to one day operate the mix console for big shows. I set my sights on preparing for such a career while at Indiana University. This was no simple matter because there was no such thing as a degree in audio. I soon discovered the Independent Learning Program. Under the auspices of that department, I assembled a mix of relevant courses from different disciplines and graduated with a college-level degree in my self-created program of audio engineering.

FIGURE 0.1
Ticket stub from the June 16, 1974 Grateful Dead concert in Des Moines, Iowa that led to my life of crime

By 1980, I had a few years of touring experience under my belt and had moved to San Francisco. There I forged relationships with John Meyer, Alexander Yuill-Thornton II (Thorny) and Don Pearson. These would become the key relationships in my professional development. Each of us was destined to stake our reputations on the same piece of equipment: the dual-channel FFT analyzer.

I would like to say that I have been involved in live concert measurement with the dual-channel FFT analyzer from day one, but this is not the case. It was day two. John Meyer began the process on a Saturday night in May of 1984. John took the analyzer, an analog delay line and some gator clips to a Rush concert in Phoenix, Arizona, where he performed the first measurements of a concert sound system using music as the source with audience in place. I was not destined to become involved in the project until the following Monday morning.

From that day forward, I have never been involved in a concert or a sound system installation without the use of a dual-channel FFT analyzer. I haven't mixed a show since that day, resetting my vision to the task of helping mix engineers to practice their art. For Don, John, Thorny and many others, the idea of setting up a system without the presence of the FFT analyzer was unthinkable. Seeing a sound system response in

FIGURE 0.2
The author with the prototype SIM analyzer with the Grateful Dead in July 1984 at the Greek Theater in Berkeley, California (Clayton Call photo)

high resolution, complete with phase, coherence and impulse response, is a bell that cannot not be un-rung. We saw its importance and its practical implications from the very beginning and knew the day would come when this would be standard practice. Our excitement was palpable, with each concert resulting in an exponential growth in knowledge. We introduced it to everyone who had an open mind to listen. The first product to come from the FFT analysis process was a parametric equalizer. A fortuitous coincidence of timing resulted in my having etched the circuit boards for the equalizer on my back porch over the weekend that John was in Phoenix with Rush. This side project (a bass guitar preamp) for my friend Rob Wenig was already six months late, and was destined to be even later. The EQ was immediately pressed into service when John nearly fell over as he saw that it could create the complementary response (in both amplitude and phase) to what he had measured in Phoenix. The CP-10 was born into more controversy than one might imagine. Equalization has always been an emotional "hot button" but the proposition that the equalizer was capable of counteracting the summation properties of the speaker/room interaction was radical enough that we obtained the support of Stanford's Dr Julius Smith to make sure that the theory would hold up.

Don Pearson was the first outside of our company to apply the concepts of in-concert analysis in the field. Don was owner of Ultrasound and was touring as the system engineer for the Grateful Dead. Don and the band immediately saw the benefit and, lacking patience to wait for the development of what would become the Meyer Sound SIM System, obtained their own FFT analyzer and never looked back. Soon thereafter, under the guidance of San Francisco Opera sound designer Roger Gans, we became involved with arena-scale performances for Luciano Pavarotti. We figured it was a matter of months before these techniques would become standard operating procedure throughout the industry. We had no idea it would take closer to twenty years! The journey, like that of sound transmission, was far more complex than we ever expected. There were powerful forces lined up against us in various forms: the massive general resistance of the audio community to sound analyzers and the powerful political forces advocating for alternate measurement platforms, to name a few.

In general, the live sound community was massively opposed to what they conceptualized as an analyzer dictating policy to the creative forces involved in the music side of the experience. Most live concert systems of the day lacked complexity beyond piles of speakers with left and right channels. This meant that the process of alignment consisted of little more than equalization. Because all of the system calibration was being carried out at a single location, the mix position, the scientific and artistic positions were weighing in on the exact same question at the same point in space. Endless adversarial debate about what was the "correct" equalization ensued because the tonal balancing of a sound system is, and always has been, an artistic endeavor. It was an absurd construct. Which is better—by ear or by analyzer?

FIGURE 0.3
November 1984 photo of Luciano Pavarotti, Roger Gans, the author (back row), Drew Serb, Alexander Yuill-Thornton II and James Locke (front row) (Drew Serb photo)

This gave way to a more challenging and interesting direction for us: the quest beyond the mix position. Moving the mic out into the space left us with a terrible dilemma: The new positions revealed conclusively that the one-size-fits-all version of system equalization was utter fantasy. The precision tuning of parametric filters carried out with great care for the mix position had no justification at other locations. The interaction of the miscellaneous parts of the speaker system created a highly variable response throughout the room. Our focus shifted from finding a perfect EQ to the quest for uniformity over the space.

This would require the subdivision of the sound system into defined and separately adjustable subsystems, each with individual level, equalization and delay capability. The subsystems were then combined into a unified whole. The rock and roll community was resistant to the idea, primarily because it involved turning some of the speakers down in level. The SPL Preservation Society staunchly opposed anything that might detract from the maximum power capability. Uniformity by subdivision was not worth pursuing if it cost power (pretty much nothing else was either). Without subdivision, the analysis was pretty much stuck at the mix position. If we are not going to change anything, why bother to look further?

There were other genres that were open to the idea. The process required the movement of a microphone around the room and a systematic approach to deconstructing and reconstructing the sound system. We began developing this methodology with the Pavarotti tours. Pavarotti was using approximately ten subsystems, which were individually measured and equalized and then merged together as a whole. Our process had to be honed to take on even more complexity when we moved into the musical theater world with Andrew Bruce, Abe Jacob, Tony Meola, Tom Clark and other such sound designers. Our emphasis changed from providing a scientifically derived tonal response to maximizing consistency of sound throughout the listening space, leaving the tonal character in the hands of the mix engineer. Our tenure as the "EQ police" was over as our emphasis changed from tonal quality to tonal equality. The process was thus transformed into *optimization*, emphasizing spatial uniformity while encompassing equalization, level setting, delay setting, speaker positioning and a host of verifications on the system. A clear line was drawn between the artistic and the scientific sectors.

In the early days, people assessed the success of a system tuning by punching out the filters of the equalizer. Now, with our more sophisticated process, we could no longer re-enact before-and-after scenarios. To hear the "before" sound might require repositioning the speakers, finding the polarity reversals, setting new splay angles, resetting level and time delays, and finally a series of equalizations for the different subsystems. Lastly, the role of the optimization engineer became clear: to ensure that the audience area receives the same sound as the mix position.

In 1987, we introduced the Source Independent Measurement system (SIM). This was the first multichannel FFT analysis system designed specifically for sound system optimization (up to sixty-four channels). It consisted of an analyzer, multiple mics and switchers to access banks of equalizers and delays. All of this was under computer control, which also kept a library of data that could be recalled for comparison of up to sixteen different positions or scenarios. It thereby became possible to monitor the sound system from multiple locations and clearly see the interactions between subsystems. It was also possible to make multiple microphone measurements during a performance and to see the effects of the audience presence throughout the space.

This is not to say we were on Easy Street at this point. It was a dizzying task to manage the assembly of traces that characterized a frequency response, which had to be measured in seven separate linear frequency sections. A single data set to fully characterize one location at a point in time was an assembly of sixty-three traces, of which only two could be seen at any one time on the tiny 4-inch screen. Comparison of one mic position to another had to be done on a trace-by-trace basis (up to sixty-three operations). It was like trying to draw a landscape while looking through a periscope.

The multichannel measurement system opened the door to system subdivision. This approach broke the pop music sound barrier with Japanese sensation Yuming Matsutoya under the guidance of Akio Kawada, Akira Masu and Hiro Tomioka. In the arenas across Japan we proved that the techniques we had developed for musical theater and Pavarotti (level tapering, zoned equalization and a carefully combined subsystem) were equally applicable to high-power rock music in a touring application.

The introduction of the measurement system as a product was followed by the first training seminar in 1987. A seminal moment came from an unexpected direction during this first seminar as I explained the process of system subdivision and mic placement for system optimization. Dave Robb, a very experienced engineer, challenged my mic placement as "arbitrary." In my mind, the selection was anything but arbitrary. However, I could not, at that moment, bring forth any objective criteria with which to refute that assertion. Since that humiliating moment, my quest has been to find a defensible methodology for every decision made in the process of sound system optimization. It is simply not enough to know *that* something works; we must know *why* it works. Those optimization methodologies and an accompanying set of methods for sound system design are the foundation of this book.

I knew nothing of sound system design when this quest began in 1984. Almost everything I have learned about the design of sound systems comes from the process of their optimization. The process of deconstructing and reconstructing other people's designs gave me the unique ability/perspective to see the aspects that were universally good, bad or ugly. I am very fortunate to have been exposed to all different types of designs, utilizing many different makes and models of speakers, with all types of program materials and scales. My approach has been to search for the common solutions to these seemingly different situations and to distill them into a repeatable strategy to bring forward to the next application.

Beginning with that very first class, with little interruption, I have been optimizing sound systems and teaching anybody who wanted to attend my seminars everything I was learning. Thorny, meanwhile, had moved on and founded a company whose principal focus was sound system optimization services using the dual-channel FFT systems. Optimization as a distinct specialty had begun to emerge.

The introduction of SIA-SMAART in 1995 resulted from the collaboration of Thorny and Sam Berkow with important contributions by Jamie Anderson and others in later years. This low-cost alternative brought the dual-channel FFT analyzer into the mainstream and made it available to audio professionals at every level. Even so, it took years before our 1984 vision of the FFT analyzer, as standard front-of-house equipment, would become reality. Unquestionably, that time has arrived. The paradigm has reversed to the point where tuning a system without scientific instrumentation would be looked at with as much surprise as was the reverse in the old days.

Since those early days we have steadily marched forward with better tools—better sound systems, better sound design tools and better analyzers. The challenge, however, has never changed. It is unlikely that it will change, because the real challenge falls mostly in the spatial distribution properties of acoustical physics. The speakers we use to fill the room are vastly improved and signal-processing capability is beyond anything we dreamed of in those early days. Prediction software is now readily available to illustrate the interaction of speakers, and we have affordable and fast analyzers to provide the on-site data.

And yet we are fighting the very same battle that we have always fought: the creation of a uniform sonic experience for audience members seated everywhere in the venue. It is an utterly insurmountable challenge. It cannot be achieved. There is no perfect system configuration or tuning. The best we can hope for is to approach uniformity. I believe it is far better to be coldly realistic about our prospects. We will have to make decisions that we know will degrade some areas in order to benefit others. We want them to be informed decisions, not arbitrary ones.

This book follows the full transmission path from the console to the listener. That path has gone through remarkable changes along its entire electronic voyage. But once the waveform is transformed into its acoustic form it enters the very same world that Jean-Baptiste Joseph Fourier found in the eighteenth century and Harry Olson found in the 1940s. Digital, schmigital. Once it leaves the speaker, the waveform is pure analog and at the mercy of the laws of acoustical physics. These unchanging aspects of sound transmission are the focus of 90 per cent of this book.

Let's take a moment to preview the challenges we face. The primary player is the interaction of speakers with other speakers, and with the room. These interactions are extremely complex on the one hand, and yet can be distilled down to two dominant relationships: relative level and relative phase. The combination of two related sound sources will create a unique spatial distribution of additions and subtractions over the space. The challenge is the fact that each frequency combines differently, creating a unique layout. Typical sound systems have a frequency range of 30 to 18,000 Hz, which spans a 600:1 ratio of wavelengths. A single room, from the perspective of spatial distribution over frequency, is like a 600-story skyscraper with a different floor plan at every level. Our job is to find the combination of speakers and room geometry that creates the highest degree of uniformity for those 600 floor plans. Every speaker element and surface will factor into the spatial distribution. Each element plays a part in proportion to the energy it brings to the equation at a point in the space. The combined level will depend upon the relationship between the individual phase responses at each location at each frequency. How do we see these floor plans? With an acoustic prediction program we can view the layout of each floor, and compare them and see the differences. This is the viewpoint of a single frequency range analyzed over the entire space. With an acoustic analyzer we get a different view. We see a single spot on each floor from the foundation to the rooftop through a piece of pipe as big around as our finger. This is the viewpoint of a single point in space analyzed over the entire frequency range.

This is a daunting task. But it is comprehensible. This book will provide you with the information required to obtain the X-ray vision it takes to see through the 600-story building from top to bottom, and it can be done without calculus, integral math or differential equations. We let the analyzer and the prediction program do the heavy lifting. Our focus is on how to read X-rays, not on how to build an X-ray machine.

The key to understanding the subject, and a persistent theme of this book, is sound source identity. Every speaker element, no matter how big or small, plays an individual role, and that solitary identity is never lost. Solutions are enacted locally on an element-by-element basis. We must learn to recognize the individual parties to every combination, because therein lie the solutions to their complex interaction.

This is not a mystery novel, so there is no need to hide the conclusion until the last pages. The key to spatial uniformity is control of the overlap of the multiple elements. Where two elements combine they must be in phase to obtain spatial uniformity. If the elements cannot maintain an in-phase relationship, then they must decrease the overlap and subdivide the space so that one element takes a dominant role in a given area. There are two principal mechanisms to create isolation: angular separation and displacement. These can be used separately or in combination and can be further aided by independent control of level to subdivide the room. This is analogous to raising children: If they don't play well together, separate them. The interaction of speakers to the room is similar to the interaction of speakers with other speakers. Those surfaces that return energy back toward our speakers will be the greatest concern. The strength of the inward reflections will be inversely proportional to our spatial uniformity.

There is no single design for a single space. There are alternate approaches and each involves tradeoffs. There are, however, certain design directions that keep open the possibility of spatial uniformity and others that render such hopes statistically impossible. A major thrust of the text will be devoted to defining the speaker configurations and design strategies that maximize the potential for spatial uniformity.

Once designed and installed, the system must be optimized. If the design has kept the door open for spatial uniformity, our task will be to navigate the system through that door. The key to optimization is the knowledge of the decisive locations in the battle for spatial uniformity. The interactions of speakers and rooms follow a consistent set of spatial progressions. The layering of these effects over each other provides the ultimate challenge, but there is nothing random about this family of interactions. It is logical and learnable. Our measurement mics are the information portals to decipher the variations between the hundreds of floor plans and make informed decisions. Our time and resources are limited. We can only discern the meaning of the measured data if we know where we are in the context of the interaction progressions.

We have often seen the work of archeologists where a complete rendering of a dinosaur is created from a small sampling of bone fragments. Their conclusions are based entirely on contextual clues gathered from the knowledge of the standard progressions of animal anatomy. If such progressions were random, there would be nothing short of a 100 per cent fossil record that could provide answers. From a statistical point of view, even with hundreds of mic positions, we will never be able to view more than a few tiny fragments of our speaker system's anatomy in the room. We must make every measurement location count toward the collection of the data we need to see the big picture. This requires advance knowledge of the progression milestones so that we can view a response in the context of what is expected at the given location. As we shall see, there is almost nothing that can be concluded from a single location. The verification of spatial uniformity rests on the comparison of multiple locations.

This book is about defined speakers in defined array configurations, with defined optimization strategies, measured at defined locations. This book is not intended to be a duplication of the general audio resource texts. Such books are available in abundance and it is not my intention to encompass the width and breadth of the complete audio picture. My hope is to provide a unique perspective that has not been told before, in a manner that is accessible to the audio professionals interested in a deeper understanding of the behavior of sound systems in the practical world.

There are a few points that I wish to address before we begin. The most notable is the fact that the physical realities of loudspeaker construction, manufacture and installation are largely absent. Loudspeakers are described primarily in terms of acoustic performance properties, rather than the physical nature of what horn shape or transducers were used to achieve it. This is also true of electronic devices. Everything is weightless, colorless and odorless here. The common transmission characteristics are the focus, not the unique features of one model or another.

The second item concerns the approach to particular types of program material such as popular music, musical theater or religious services, and their respective venues such as arenas, concert halls, showrooms or houses of worship. The focus here is the shape of the sound coverage, the scale of which can be adjusted to fit the size of the venue at the appropriate sound level for the given program material. It is the venue and the program material taken together that create an application. The laws of physics are no different for any of these applications, and the program material and venues are so interchangeable that attempts to characterize them in this way would require endless iterations. After all, the modern-day house of worship is just as likely to feature popular music in an arena setting as it is to have speech and chant in a reverberant cathedral of stone.

The third notable aspect is that there are a substantial number of unique terminologies found here and, in some cases, modification of standard terminologies that have been in general use. In most cases the conceptual framework is unique and no current standard expressions were found. The very young field of sound system optimization has yet to develop consistent methods or a lexicon of expressions for the processes shown here. In the case of some of these terms, most notably the word "crossover," there are compelling reasons to modify the existing usage, which will be revealed in the body of the text.

The book is divided into three parts. The first part, "Sound systems," explores the behavior of sound transmission systems, human hearing reception and speaker interaction. The goal of this part is a comprehensive understanding of the path the signal will take, the hazards it will encounter along the way and how the end product will be perceived upon arrival at its destination. The second part, "Design," applies the properties of the first part to the creation of a sound system design. The goals are comprehensive understanding of the tools and techniques required to generate a design that will create a successful transmission/reception model. The final part, "Optimization," concerns the measurement of the designed and installed system, its verification and calibration in the space.

ABOUT THE THIRD EDITION

From the viewpoint of my publisher, Focal Press, this is indeed the third edition of *Sound Systems: Design and Optimization*. From my perspective it feels more like the thirtieth edition, because I have been writing about these same subjects for thirty+ years. You might think I would have figured this subject out by now but I can assure you I am still learning. This field of work continues to evolve as we get better tools and techniques, which is exactly what I find most interesting about it. Study in this field is a moving target as new technology opens doors and removes obstacles and excuses. The more I learn about this, the more I realize how much I have to learn.

Adding new areas is the easy part of creating a new edition. It's what to do with the previous material that presents a challenge. There are two ways to approach the old material: innocent until proven guilty, or the opposite. The former approach leaves things in unless they are conclusively out of date or irrelevant. The latter approach throws the old material out unless it can prove it is still current practice and up to date.

I studied the later editions of several other authors and noticed a troubling trend. Although new information was added in later editions, a lot of old information remained in place. Seeing vacuum tube circuits from the 1960s in a current-day pro audio text was a tipping point for me. The decision was made to trim out the old to make room for the new. If it's the way we do things now, it's in. If we've moved on, it's out.

The surprise for me was how much we have moved forward in this time, which meant entire chapters were bulldozed and rebuilt. One of the hardest decisions was to let go of the perspective sidebars that colored the previous editions with the wisdom and insight of so many of my friends and colleagues. The bottom line is that there is simply too much new information to be added. I take comfort in knowing that there are many other places where those voices can be heard and that optimization is now firmly ensconced as part of the audio landscape.

There have been no updated laws of physics and our audio analyzers still compute things the same way as they did in 1991. But today's analyzers are faster, easier and able to multitask, which means we can get much more done in a short time. We can tune methodically, and methods are what this book is about. We have far better loudspeakers, processors and steadily better rooms to work in. All this leads to the primary goal of this third edition: current methodologies and techniques for sound system design and optimization.

This page intentionally left blank

ACKNOWLEDGEMENTS

The development of this book spans more than thirty years in the field of sound system optimization. Were it not for the discoveries and support of John and Helen Meyer, I would have never become involved in this field. They have committed substantial resources to this effort, which have directly helped the ongoing research and development leading up to this writing. In addition, I would like to acknowledge the contribution of every client who gave me the opportunity to perform my experiments on their sound systems. Each of these experiences yielded an education that could not be duplicated elsewhere. In particular I would like to thank David Andrews, Peter Ballenger, Nick Baybak, Mark Belkie, Mike Brown, Andrew Bruce, John Cardenale, Tom Clark, Mike Cooper, Jonathan Deans, François Desjardin, Steve Devine, Martin Van Dijk, Steve Dubuc, Duncan Edwards, Aurellia Faustina, T.C. Furlong, Roger Gans, Scott Gledhill, Michael Hamilton, Andrew Hope, Abe Jacob, Akio Kawada, Andrew Keister, Tony Meola, Ben Moore, Philip Murphy, Kevin Owens, Frank Pimiskern, Bill Platt, Marvic Ramos, Harley Richardson, Paul Schmitz, David Sarabiman, Pete Savel, Rod Sintow, Bob Snelgrove, David Starck, Benny Suherman, Leo Tanzil and Geoff Zink, all of whom have given me multiple opportunities through the years to refine the methods described here. Special thanks to Mr O at the National Theatre of Korea, Mr Song at Dongseo University, Mr Lee and others at the LG Art Center.

I have also learned much from other engineers who share my passion for this field of work and constantly challenge me with new ideas and techniques. This list would be endless and includes but is not limited to Brian Bolly, Michael Creason, Ales Dravinec, Josh Evans, Peter Grubb, Glenn Hatch, Luke Jenks, Miguel Lourtie, Karoly Molnar, John Monitto, Matt Salerno, John Scandrett and Robert Scovill.

I would also like to thank Meyer Sound for sponsoring my seminars and Gavin Canaan and others for organizing them. I am grateful to everyone who has attended my seminars, as the real-time feedback in that context provides a constant intellectual challenge and stimulation for me. My fellow instructors in this field have contributed much collaborative effort through discussion and the sharing of ideas. Notable among these are Jamie Anderson, Oscar Barrientos, Timo Beckman, Sam Berkow, Harry Brill, Richard Bugg, Steve Bush, Jim Cousins, Pepe Ferrer, Michael Hack, Mauricio Ramirez, Arthur Skudrow, Hiro Tomioka, Merlijn Van Veen and Jim Woods.

Thanks to Daniel Lundberg for his creation of the uncoupled array calculator based on data from my previous edition. This tool is a mainstay of my design process.

A huge majority of the knowledge, data (and graphics) in this book comes from two sources: Meyer Sound's SIM3 Audio Analyzer and their MAPP Online™ platform. I would have nothing to write about without these tools. My gratitude goes to everyone who contributed to their creation, including (but not limited to) John and Helen Meyer, Perrin Meyer, Dr Roger Schwenke, Fred Weed, Todd Meier, Mark Schmeider, Paul Kohut and the late Jim Isom.

The following figures contain data from my earlier publications at Meyer Sound and their permission is gratefully acknowledged: Figures 3.9, 4.24, 4.27, 13.12 to 13.16. The data presented in Figures 1.1, 1.4, 1.11 and 12.11 were created using the calculations by Mauricio Ramirez. The 3-D wraparound graphics (Figure 12.8) were adapted from the animations created by Greg Linhares. Merlijn Van Veen contributed a mix of core calculations, data and graphics that went into the construction of Figures 1.12, 2.2, 2.5 to 2.12, 3.30, 3.37, 3.42, 4.8, 8.11, 8.14 and 14.10. John Huntington contributed some of the photographs used in the section break pages, specifically Section 1 (panels 1, 5 and 7) and Section 3 (panel 4).

Thanks go to all of the people who aided in the process of bringing this edition to physical reality such as my editor Megan Ball and Mary LaMacchia at Focal Press. Additional thanks go to Margo Crouppen for her support throughout the entire publishing process.

I received proofing and technical help for my previous editions from Jamie Anderson, Sam Berkow, David Clark, John Huntington, Mauricio Ramirez and Alexander (Thorny) Yuill-Thornton. Harry Brill Jr., Richard Bugg, Philip Duncan, John Huntington, Jeff Koftinoff and Mauricio Ramirez contributed to the third edition.

Merlijn Van Veen deserves special recognition for his enormous contributions to this edition. He hung with me at every step of the way, pushing me to clarify language, methodology and backing up my calculations. He also contributed greatly to the graphics, providing both material support and advice on how to best convey the information. I learned much from Merlijn in the course of this writing and his mark has clearly been left in these pages.

Finally, there is no one who comes close to contributing as much support as my wife Merridith. She has read every page and seen every drawing of every edition and been absolutely tireless in her efforts. Each edition has been an endurance test lasting over a year and she has stuck with me through each, swayed by my promises that this would be the last one (fooled her again). This book would not have been possible without her support as my agent, manager, copy editor, proofreader and cheerleader.

PART I

Sound systems

This page intentionally left blank

CHAPTER 1

Foundation

> *foundation* n. solid ground or base on which a building rests; groundwork, underlying principle; body or ground upon which other parts are overlaid.
> **Concise Oxford Dictionary**

We begin with the establishment of a firm foundation upon which to build the structure for the study of sound system design and optimization. Here we standardize definitions and terminology for usage throughout this book. If this is not your first day in audio you will already understand many of the concepts in this chapter, because much of this is the universal foundation material found in books, the Internet and the tribal knowledge passed down from elders on the road. I have, however, selectively edited the list of fundamentals to those concepts pertinent to modern-day design and optimization. We won't cover the Doppler effect, underwater acoustics, industrial noise suppression and any other areas that we can't put to immediate practical use.

The next section will read somewhat like a glossary, which is traditionally placed at the rear of the book, and the last place you would normally read. These are, however, the first concepts we need to establish and keep in mind throughout. The foundation we lay here will ease the building process as we progress upward in complexity.

We begin with the foundations of this book.

Sound

Sound is a vibration or mechanical wave that is an oscillation of pressure (a vibration back and forth) transmitted through some medium (such as air), composed of frequencies within the range of hearing.

System

A system is a set of interacting or interdependent components forming an integrated whole. A sound system consists of a connected collection of components whose purpose is to receive, process and transmit audio signals.

The basic components consist of microphones, signal processing, amplifiers, speakers, interconnection cabling and digital networking.

Design

Design is the creative process of planning the construction of an object or system. We design sound systems in rooms by selecting the components, their function, placement and signal path.

Optimization

Optimization is a scientific process whose goal is to achieve the best result when given a variety of options. In our case, the goal is the maximization of sound system performance in conformance with the design intent. And do we ever have a variety of options! The primary metric for optimization is uniformity of response over the space.

1.1 UNIVERSAL AUDIO PROPERTIES

Let's define the universal properties within our limited field of study: the acoustical and analog electrical behavior of sound and its mathematical renderings in digital form.

1.1.1 Audio

Audio is a stream of data beginning and/or ending as sound. The audible version connects directly to our ears through the air. Audio can also exist in an encrypted form that cannot be heard until decoded. The monumental breakthrough of Edison's phonograph was encoding audio into a groove on a lacquer cylinder for playback through a mechanical decoder (a diaphragm attached to a moving needle). Encoded audio exists in many forms: magnetic flux (in tape, transformers, microphones or loudspeakers), electronic signal in a wire and even as a digital numerical sequence.

Audio stream oscillations can be rendered as a sequential set of amplitude values over time. Analog audio renderings are a continuous function (i.e. the amplitude and time values are infinitely divisible). Digital audio renderings are finitely divisible (i.e. a single amplitude value is returned for each block of time).

1.1.2 Frequency (f) and time (T)

Frequency (f or Hz) is the number of oscillations completed in one second (the reciprocal of the time period). Period (T) is the time interval to complete one cycle. Frequency is cycles/second and time is seconds/cycle ($f = 1/T$ and $T = 1/f$). Either term describes an oscillation, the choice being for convenience only. Fluency in the translation of time and frequency is essential for design and optimization (Fig. 1.1). Period formulas for T are computed in seconds, but in practice we almost always use milliseconds (1 ms = 0.001 of a second).

1.1.3 Cycle

A cycle is a completed oscillation, a round trip that returns to the starting state of equilibrium. The distinction between cycle and period is simply units. A period is measured in time (usually ms) and a cycle is measured in completed trips. A cycle at 250 Hz has a period of 4 ms. One cycle @125 Hz (or two cycles @250 Hz) have 8 ms periods. We often subdivide the cycle by fractions or degrees of phase, with 360° representing a complete cycle. It is common to use the term "cycle" when dealing with the phase response, e.g. 1 ms @250 Hz, which is ¼ cycle (90°).

1.1.4 Oscillation

Oscillation is the back and forth process of energy transfer through a medium. This may be mechanical (e.g. a shaking floor), acoustical (e.g. sound in the air) or electromagnetic (e.g. an electronic audio signal). The oscillating matter's movement is limited by the medium and returns to equilibrium upon completion. Energy transfer occurs *through* the medium.

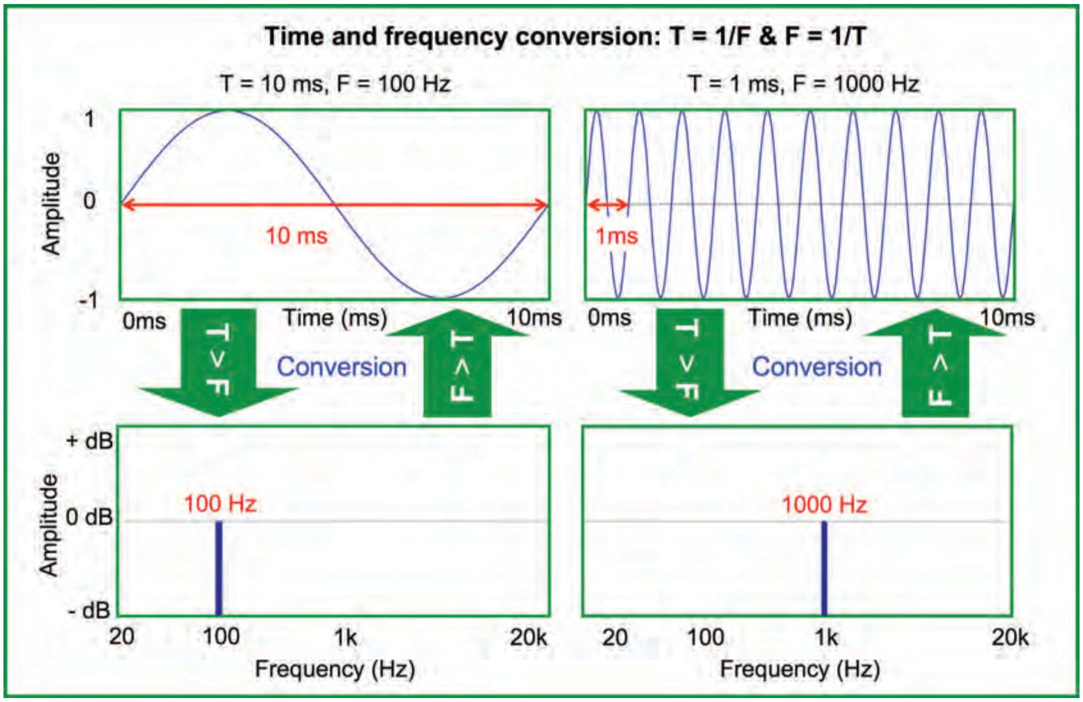


FIGURE 1.1
Relationship of time and frequency

1.1.5 Amplitude (magnitude)

Amplitude is the quantitative measure of oscillating energy, the extent of mechanical displacement (m, cm, etc.), acoustical pressure change (SPL), electrical voltage change (V), magnetic flux (B) and others. Amplitude values can be expressed linearly (e.g. volts) or logarithmically as a ratio (the dB scale). Amplitude is one of the more straightforward aspects of audio: bigger is bigger. Amplitude (black T-shirt) and magnitude (lab coat) are interchangeable terms. We will introduce audio amplitude in various forms next and then cover the scaling details (such as the dB scale) in section 1.2. It may be necessary to bounce between those sections if you are completely unfamiliar with these scales.

1.1.5.1 DC POLARITY (ABSOLUTE POLARITY)

DC (direct current) polarity is the signal's directional component (positive or negative) relative to equilibrium. Electrical: +/- voltage, acoustical +/- pressure (pressurization/rarefaction), etc. This is applicable in strict terms to DC signals only, because AC signals have both positive and negative values. A 9 V battery illustrates the electrical version. Connecting the battery to a speaker illustrates the acoustical version, because it only moves in one direction.

1.1.5.2 ABSOLUTE AMPLITUDE

Absolute amplitude is the energy level relative to equilibrium (audio silence). Electrical audio silence is 0 VAC, whether DC is present or not. Only AC can make audio. DC moves a speaker but unfortunately the only sound it can make is the speaker burning. Acoustic systems are referenced to changes above or below the ambient air pressure (air's equivalent for DC). Absolute amplitude values cannot be less than zero, because we can't have less movement than equilibrium. A "-" sign in front of an amplitude value indicates negative polarity. Relative amplitude values are more common in audio than absolute ones.

1.1.5.3 RELATIVE TO A FIXED REFERENCE

Audio levels change on a moment-to-moment basis. Therefore most amplitude measurements are relative to a reference (either fixed or movable). Examples of fixed references include 1 V (electrical) or the threshold of human hearing (acoustical) (Fig. 1.2). The reference level can be expressed in various units and scales, as long as we agree on the value. An amplitude value of 2 volts can be expressed as 1 volt above the 1 V volt reference (linear difference) or twice the reference level (linear multiple). Many reference standards for audio are specified in decibel values (dB), which show amplitude changes in a relative log scale (like our hearing).

One volt, 0 dBV and +2.21 dBu are the same amount of voltage, expressed in different units or scales. A musical passage with varying level over time can be tracked against the fixed reference, e.g. a certain song reaches a maximum level of 8 volts (+18 dBV, +20.21 dBu) and an acoustical level of 114 dB SPL (114 dB above the threshold of hearing).

1.1.5.4 RELATIVE TO A VARIABLE REFERENCE (AMPLITUDE TRANSFER FUNCTION)

We can monitor the amplitude of constantly changing signals in second-cousin form, i.e. relative to a relative (Fig. 1.3). We compare signal entering and exiting a device, such as music going through a processor. The relative/relative measurement (the 2-channel output/input comparison) is termed "transfer function measurement," the primary form of analysis used in system optimization. Frequency response amplitude traces in this book are relative amplitude (transfer function) unless specified otherwise.

Let's return to the above example. The music level is changing, but the output and input waveforms track consistently as long as the processor gain remains stable. If output and inputs are level matched (a 1:1 ratio), the device has a transfer function voltage gain of unity (0 dB). If the voltage consistently doubles, its transfer function gain is 2× (+6 dB). We can span the electronic and acoustic domains by comparing the processor output with the sound level in the room. This reveals a voltage/SPL tracking relationship, such as +0 dBV (1 V) creates 96 dB SPL (and +6 dBV (2 V) creates 102 dB SPL etc.).

The beauty of transfer function measurement is its ability to characterize a device (or series of devices) with random input material across multiple media, so long as the waveforms at both ends are correlated. This will be covered extensively in Chapter 12.

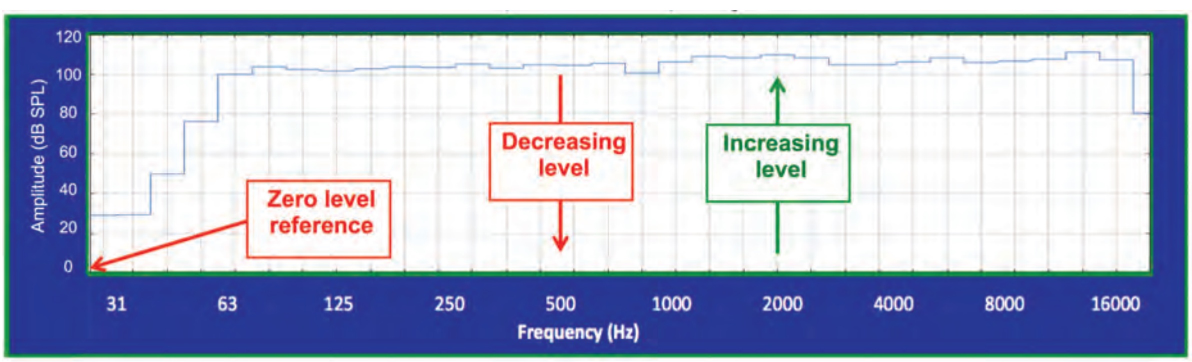


FIGURE 1.2
Absolute amplitude vs. frequency. Amplitude is referenced to 0 dB SPL at the bottom of the vertical scale.

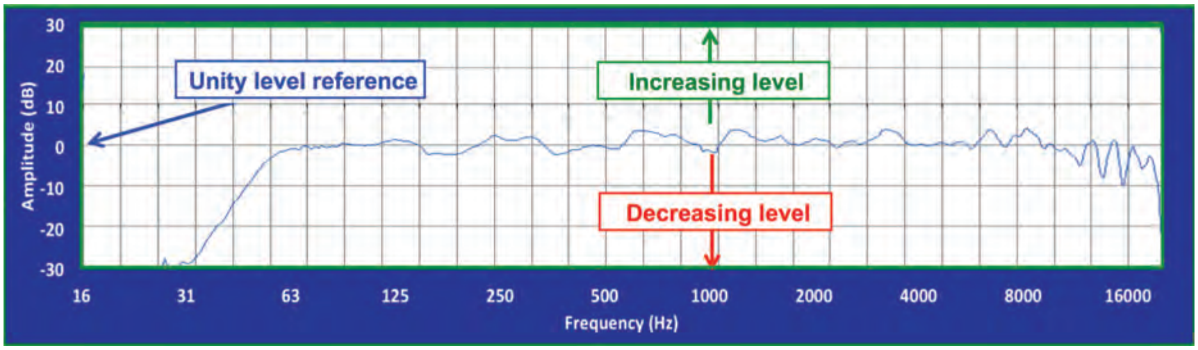


FIGURE 1.3
Transfer function amplitude vs. frequency. Amplitude is referenced to unity gain at the center of the vertical scale.

1.1.5.5 PEAK (PK) AND PEAK-TO-PEAK (PK–PK)

The peak (pk) amplitude value is the signal's maximum extent above *or* below equilibrium whereas peak-to-peak (pk–pk) is the span between above *and* below values. Any device in the transmission path must be capable of tracking the full extent of the pk–pk amplitude. Failure results in a form of harmonic distortion known as "clipping" (because the tops of the peaks are flattened). The waveform seen on an oscilloscope or digital audio workstation is a representation of the pk–pk values.

1.1.5.6 RMS (ROOT MEAN SQUARED)

The rms value (root-mean-squared) is the waveform's "average-ish" amplitude. The rms calculation makes AC (+ and -) equivalent to DC (+ or -). For example a 9 V_{DC} battery and 9 V_{RMS} generator supply the same power. We use rms instead of simple averaging because audio signals move both above and below equilibrium. A sine wave (such as the AC line voltage) averages to zero because it's equally positive and negative. Sticking your fingers in a wall socket provides a shocking illustration of the difference between "average" and rms. Kids, don't try this at home.

The rms value is calculated in three steps: (s) squaring the waveform, which enlarges and "absolutes," making all values positive, (m) finding the squared waveform's mean value and (r) taking the square root to rescale it back to normal size. Its proper full name is "the root of the mean of the square." Note that rms values are strictly a mathematical rendering. We never hear the rms signal (it would be less recognizable as audio than an MP-3 file). The waveforms we transmit, transduce, render digitally and hear are peak–peak.

1.1.5.7 CREST FACTOR

Crest factor is the ratio between the actual amplitude traced by the waveform (the peak or crest) and the heat-load-simulating rms value (Fig. 1.4). For a DC signal (such as a battery) the difference is nothing: crest factor of 1 (0 dB). The simplest audio signal, the sine wave, has a crest factor of 1.414 (3 dB). Complex signals can have vastly higher peak/average ratios (and higher crest factors). Pink noise (see section 1.1.8.3) is approximately 4:1 (12 dB), whereas transient signals such as drums can have 20 to 40 dB.

1.1.5.8 HEADROOM

Headroom is the remaining peak–peak amplitude capability before overload at a given moment, the reserved dynamic range, and our insurance policy against overload. Every electronic or electromagnetic device has its

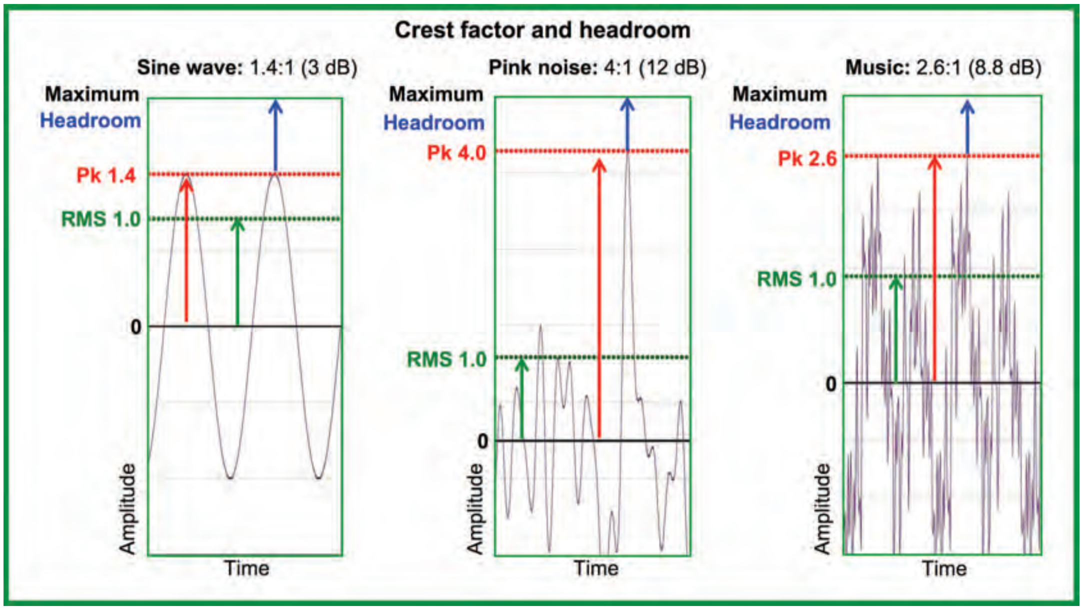


FIGURE 1.4
Crest factor examples with different signals

amplitude upper limit. Linear audio transmission requires the entire extent of the peak–peak signal to pass through without reaching the device's upper limit (no clipping, limiting or compression). Headroom is the remainder between the device's limits and the signal's positive or negative peak. This has historically had a mysterious quality in part because of slow metering ballistics that fall short of tracking the peak–peak transient values. This leaves engineers concerned (rightfully) about potential clipping even when meters indicate remaining dynamic range. An oscilloscope demystifies headroom/clipping because it displays the peak–peak waveform. Digital headroom is the remaining upper bits in the rendering of the pk–pk waveform.

1.1.6 Phase

Phase is the radial clock that charts our progress through a cycle. A completed cycle is 360°, a half-cycle is 180° and so on. The phase value is calculated in reference to a specific frequency. There is not a limit to the phase value, i.e. we can go beyond 360°. The phasor radial positions of 0°, 360° and 720° are equivalent but the phase delay is not, revealing that things have fallen one and two cycles behind respectively. Two race cars with matched radial positions will cross the finish line together, but there is a million dollar difference between 0° and 360°. This makes a difference to sound systems as well, as we will see.

1.1.6.1 ABSOLUTE PHASE

Absolute phase is the value at a given moment relative to a stationary time reference, typically the internal clock of an analyzer. Yes, you read that correctly. Absolute phase is *relative* to a reference, such as the start of the measurement period, which becomes the 0 ms time and 0° phase reference. We don't need to see the absolute phase numbers even though our analyzers compute them. It's like having a wristwatch with only a second hand, which won't help us get to the gig on time. Our analyzers show *relative* phase (a comparison between two channels of internal absolute phase calculations). Note that the term "absolute phase" is often misapplied for the concept of absolute polarity (section 1.1.7.1).

1.1.6.2 RELATIVE (PHASE TRANSFER FUNCTION)

As stated above, relative phase is the difference between two absolute phase values (Fig. 1.5). Relative phase (in degrees) is the only version of phase response shown in this book, so we can proceed to shorten the relative phase phrase to simply "phase." A series of phase values over frequency taken together create a phase slope that can be translated to phase delay (section 1.1.6.6). Phase is a radial function that can set our heads spinning when it comes to reading the 2-D charts. The response often contains "wraparound," which is how we display phase response that exceed the limits of the 360° vertical scale (see section 1.3.4).

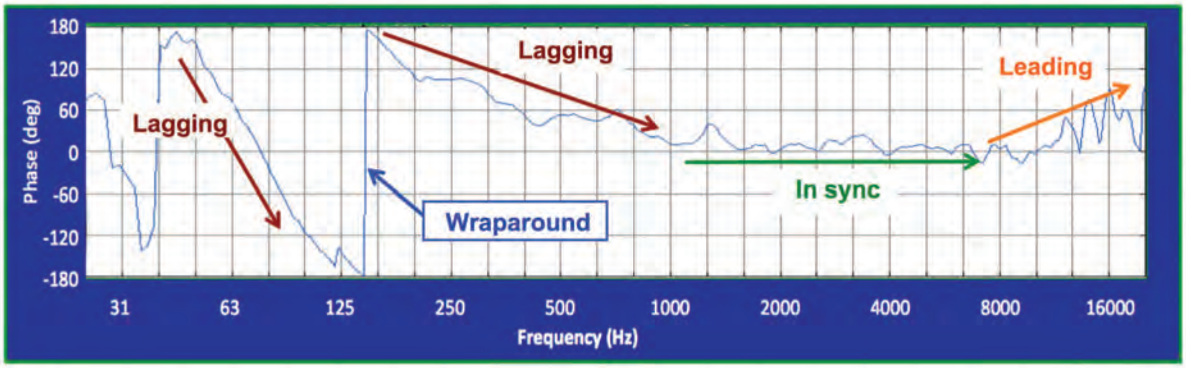


FIGURE 1.5
Transfer function phase vs. frequency. Phase is referenced to unity time (0°) at the center of the vertical scale.

1.1.6.3 PHASE SHIFT

The "shift" in question here is phase change over frequency (Fig. 1.6). This can be stable and constant (such as phase shift caused by a filter) or unstable and variable (such as wind). Our practical concern is frequency-dependent delay (i.e. different frequencies shifted by different amounts). A system with such phase shift has a temporally stretched transient response (often termed "time smearing"). A translation example: The rise and fall of a drum hit would be rounded and expanded because parts of the transient are behind others. The secondary concern regarding phase shift is compatibility with other devices that share common signals and sum together (either in the air or inside our gear). An example is the combination of two different speaker models, with unmatched phase shift characteristics. When summed together we get a phase shift between the phase shifts, which we term the phase offset (the next topic).

1.1.6.4 PHASE OFFSET

Phase offset is the favored term here for phase differences between two measured systems (Fig. 1.7) and a famous audio sorority (Δ-Phi). A known phase offset (in degrees over a frequency span) can be converted to time offset (or vice versa). Phase offset is put to practical use when correlated sources are summed. With known phase and level offsets we can precisely predict the summed response at a given frequency. Phase offset requires a frequency specification and is therefore preferred for frequency-dependent time differences, such as the crossover between an HF and LF driver.

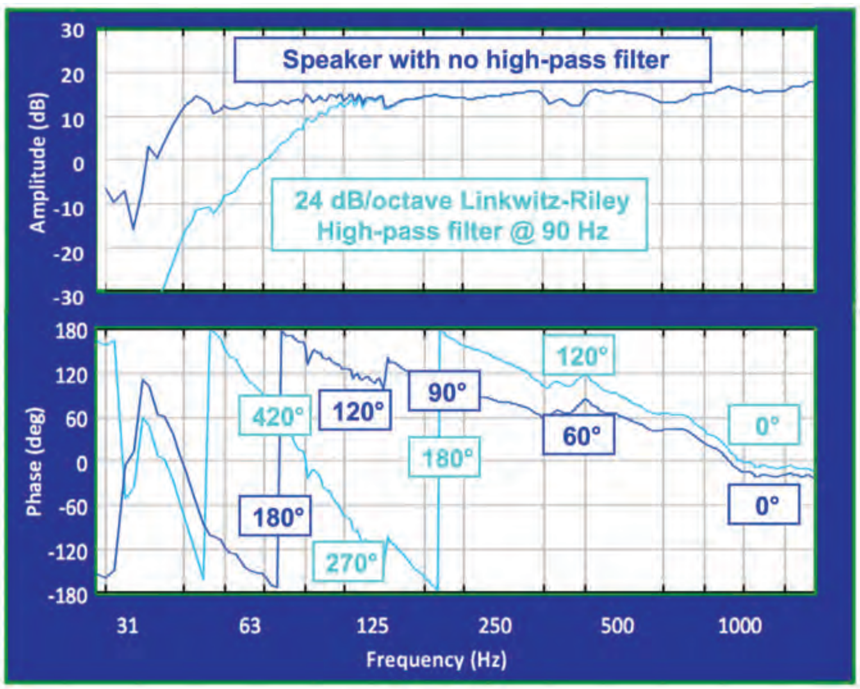


FIGURE 1.6
Example of phase shift added by a 24 dB/octave Linkwitz-Riley filter @90 Hz. The filter only affects the amplitude response below 100 Hz but the phase is shifted over a much wider range.

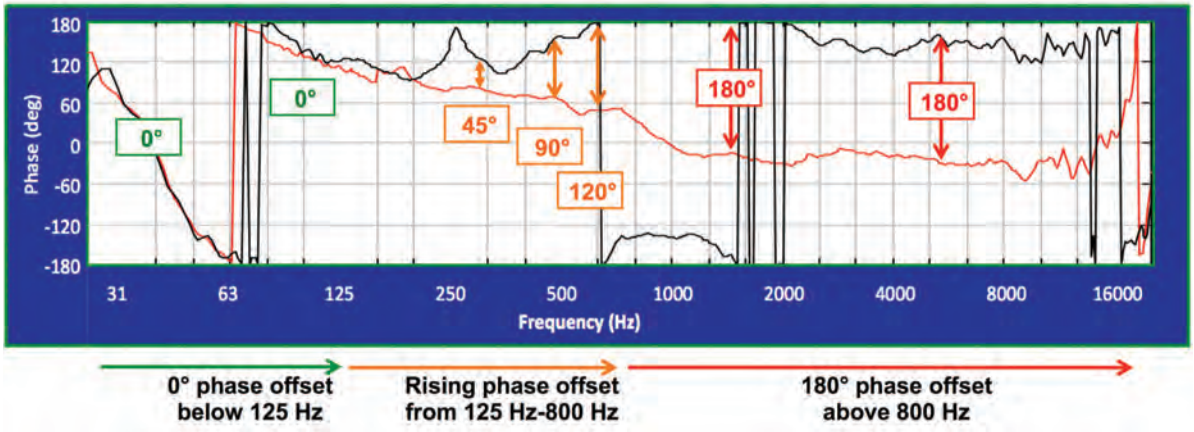


FIGURE 1.7
Phase offset example showing two different speaker models. The LF ranges match but the HF ranges are offset by 180°.

We don't use the terms "phase offset" and "phase shift" interchangeably here. The distinction is drawn as follows: Phase shift occurs inside a single device or system. Phase offset is *between* devices or systems. A filter creates phase shift. Moving two speakers apart creates phase offset.

1.1.6.5 TIME OFFSET

Time offset (in ms) is a frequency-independent measure for propagation paths (Fig. 1.8). Latency in the signal path or different arrival times between a main and delay speaker are practical examples of time offset. Time offset is our preferred term to describe frequency-independent time differences. Frequency-dependent time offsets are better described by phase delay (below). Time offset (in ms) can be translated to phase offset (in degrees for a given frequency), and vice versa. A fixed-time offset at all frequencies creates an increasing amount of phase offset as frequency rises.

We are very concerned with time offsets within the signal path, particularly in the analog electronic and digital paths where even small amounts are very audible. Strategies for managing time offset between speakers play a big part in system optimization.

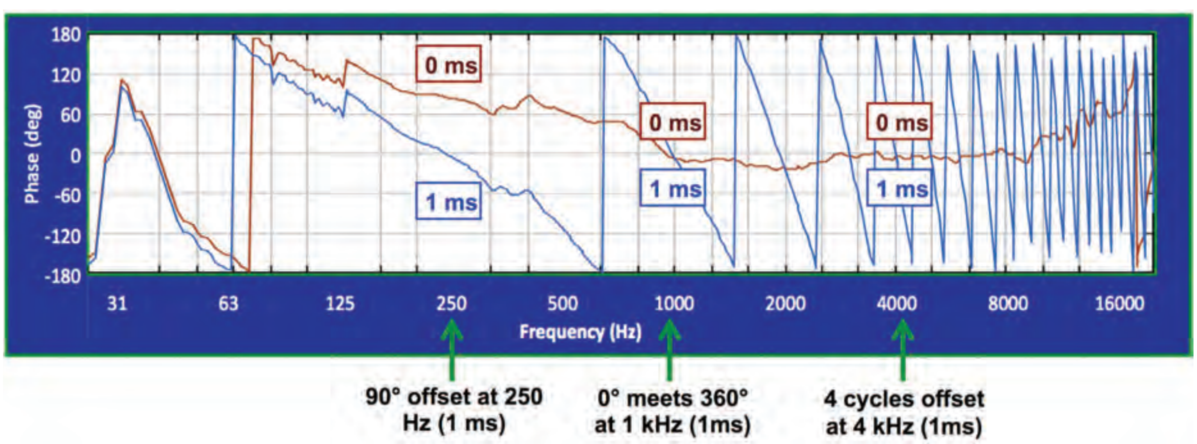


FIGURE 1.8
Time offset example showing 1.0 ms between two otherwise matched speakers. The phase offsets are 90° @250 Hz, 360° @1 kHz and 1440° @4 kHz.

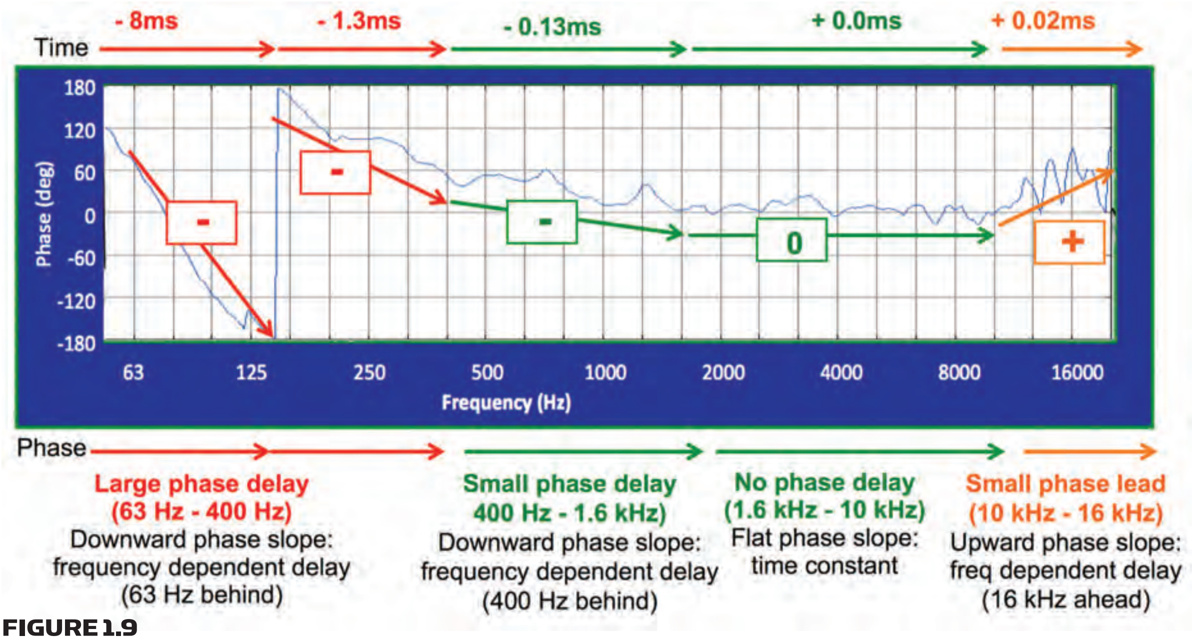


FIGURE 1.9
Example of frequency-dependent phase delay. The LF range has a downward phase slope, indicating phase delay behind the HF region.

1.1.6.6 PHASE DELAY

Phase delay, the time-translated version of phase shift, is a metric used to describe systems with frequency-dependent delay (Fig. 1.9). Phase delay is computed by finding the phase shift over a range of frequencies (the phase slope). The phase value at two frequencies must be known to make the calculation. Phase delay is mostly interchangeable with the term "group delay," an unimportant distinction for optimization decisions.

Any frequency band-limited device will exhibit some phase delay, i.e. some frequencies are transmitted later than others. This includes every loudspeaker not known to marketing departments. Unless extraordinary measures are performed, real-world loudspeakers exhibit increasing phase delay in the LF range. Speaker models have substantially different amounts of phase delay. This can cause compatibility issues when combined, because a single time offset value cannot synchronize the two speakers over the full spectrum. Phase delay is often used to characterize and remedy phase offsets between speakers during optimization.

1.1.7 Polarity

Polarity is a binary term representing the positive vs. negative amplitude orientation of the waveform. Systems with "normal" polarity (+) proceed first in a positive direction followed by negative and back to equilibrium (whereas reverse polarity systems do the opposite). Loudspeakers with a positive voltage applied show normal polarity as a forward movement (pressurization) and negative polarity as a rearward movement (rarefaction).

Polarity has a storied history in professional audio. The foremost concerns JBL founder James B. Lansing, who in 1946 chose to reverse the polarity of his speakers to ensure incompatibility with those of his former employer Altec Lansing. The polarity war lasted over forty years. We all lost.

Our standard line level connector also lacked a polarity standard (the XLR connector was pin 3 hot in the USA and pin 2 hot for Europe and Japan). We had actually entered the digital audio age before an analog standard of pin 2 hot was established worldwide.

The term polarity is often mistakenly substituted for "phase" because a polarity reversal creates a unique case of phase shift, where all frequencies are changed by 180°. Polarity, however, has no time offset or phase delay and is frequency independent. Phase shift (described above) is frequency dependent and is caused by delay in the signal.

It is interesting to note that we have more terms for polarity than options: +/-, normal or reversed, non-inverting or inverting, right or wrong.

1.1.7.1 ABSOLUTE POLARITY

The "absolute" here refers to a system's net polarity orientation to the original source, often referred to by audiophiles as "absolute phase." A positive absolute polarity value is claimed if the waveform at the end of the chain matches the polarity of the original. Let's use an example to follow absolute polarity from start to finish. A kick drum moves forward (+). The positive pressure wave moves a microphone diaphragm inward (+). The console and signal processing maintain normal polarity to the power amplifier inputs (+). The amplifier output swings positive voltage on the hot terminal (+) and the speaker moves forward creating a positive pressure wave (+). Success! But some big assumptions have been made. Perfect mics, electronics and speakers that have flat phase responses. Pssst! They don't. There is frequency-dependent delay in the speakers, enough to make parts of them 180° different than others. Polarity is suddenly ambiguous. High-pass filters, low-pass filters, reflex box tuning and LF driver radiation properties all add up to frequency-dependent phase delay. Who's right? The phase response of the waveform at the end of the chain does not match the original, which means the polarity can no longer be called absolute.

I will not enter the fray as to whether humans can detect absolute polarity. Yet one thing is clear: A perfectly flat amplitude and phase transmission system moves us closer to reproducing the original drum sound, possibly to the point where polarity *is* absolute enough for us to make some conclusions. Nonetheless it is gospel among audiophiles that this is *hugely* important. Here is the best part about that. Audiophiles listen to recordings. Recordings can be assembled from hundreds of tracks recorded at different times, from different studios, transposed across different media, manipulated from here to kingdom come in digital audio workstations, sent to a mastering lab across the country, transcribed to a lacquer mother, a metal stamper and then pressed to vinyl, played by a turntable cartridge and finally through speakers that have drivers on the sides and rear to give that awesome surroundoscopic envelopment and create stereo everywhere. Can you please tell me what exactly is the absolute reference for my polarity?

1.1.7.2 RELATIVE POLARITY

Relative polarity is the one that counts: the polarity relationship between audio devices (Fig. 1.10). There is no disputing the fact that summing devices with different relative polarities causes cancellations and other unexpected outcomes. Relative polarity errors are best prevented by ensuring all devices are matched (and normal).

1.1.7.3 POLARITY AND PHASE DELAY COMBINED

Systems that exhibit frequency-dependent phase delay (e.g. speakers and most filters) can have their phase response further modified by a polarity change. The resulting phase values over frequency are a combination

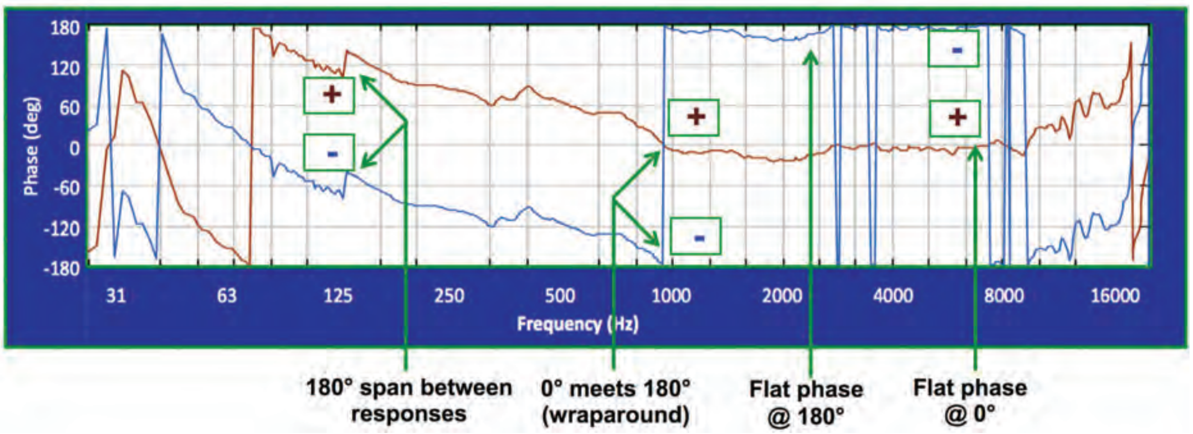


FIGURE 1.10
Example of polarity and reverse polarity over frequency

of these two modifiers (phase delay is frequency dependent whereas polarity reversal changes all frequencies by 180°). This finds practical use in acoustical crossover filter and delay settings, and steering cardioid subwoofer arrays.

1.1.8 Waveform

The waveform is the cargo transmitted through our delivery network: the sound system. It's the shape of amplitude over time, the fingerprint of an audio signal in all its simplicity or complexity. A sine wave is waveform simplicity. "In-a-Gadda-da-Vida" is a 17-minute complex waveform with many frequencies and three chords.

The waveform audio stream is a continuous function, a chronological series of amplitude values. Complex waveforms result from combinations of sine waves at various frequencies with unique amplitude and phase values. We can mathematically construct the waveform once we know the amplitude and phase values for each frequency. Conversely, we can deconstruct a known waveform into its component frequencies and their respective amplitude and phase values.

We are all familiar with the "synthesizer," a musical device that creates complex waveforms by combining individual oscillators at selectable relative levels and phase. We had an early MOOG synthesizer at Indiana University (SN# 0005) that required patch cables to mix oscillators together to build a sound. This was raw waveform synthesis. Its inverse (the deconstruction of a complex signal into its sine components) is the principle of the Fourier Transform, the heart of the audio analyzer used in optimization.

1.1.8.1 SINE WAVE

A sine wave, the simplest waveform, is characterized as a single frequency with amplitude and phase values. A "pure" sine wave is a theoretical construct with no frequency content other than the fundamental. Generated sine waves have some unwanted harmonic content (harmonics are linear multiples of the fundamental frequency) but can be made pure enough for use as a test signal to detect harmonic distortion in our devices.

The sine wave is the fundamental building block of complex audio signals, combinations of sine waves of different frequencies with individual amplitude and phase values.

The steady state of the sine wave makes it suitable for level calibration in electronic systems (analog or digital) whose flat frequency response allows for a single frequency to speak for its full operating range. Speakers cannot be characterized as a whole by a single sine wave because their responses are highly frequency dependent and readings can be strongly influenced by reflections.

1.1.8.2 COMBINING UNMATCHED FREQUENCIES

Multiple frequencies coexist in the waveform when mixed together (Fig. 1.11). The amplitude and phase characteristics of the individual frequencies are superpositioned over each other in the waveform. The amplitude rises in portions of the waveform where signals are momentarily phase matched, thereby increasing the crest factor beyond the 3 dB of the individual sine wave.

It is possible to separate out the individual frequencies with filters or by computation (e.g. the Fourier Transform). The relative phase of mixed frequencies affects the combined waveform amplitude but does not affect that of the individual frequencies. In other words, 10 kHz cannot add to, or cancel, 1 kHz regardless of relative phase. The combined waveform, however, differs with relative phase.

1.1.8.3 WHITE AND PINK NOISE

There are a few special waveforms of recurring interest to us. White noise is arguably the most natural sound in the world: all frequencies, all the time, with random phase and statistically even level. White noise is the product

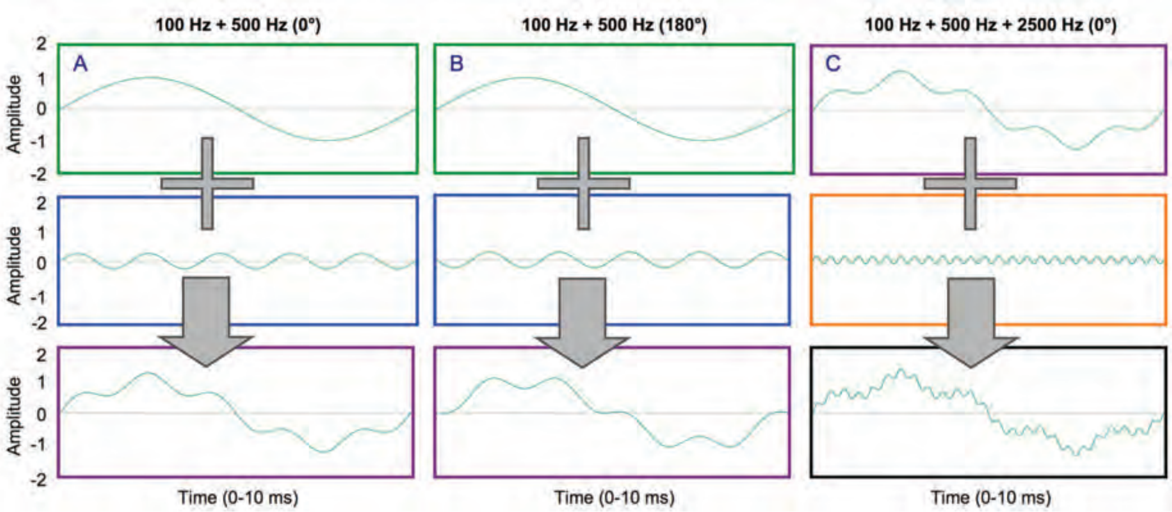


FIGURE 1.11
Waveform mixing: the combination of signals with unmatched frequencies

of random molecular movements in our electronics, well known as the noise floor and the sound of radio mic disaster. The energy is spread evenly over the linear frequency range, so half of the humanly audible energy is below 10 kHz and half above. White noise is perceived as spectrally tilted toward the HF (and called "hiss" because our ears respond on a log basis over frequency).

Pink noise, the most common audio spectrum test signal, is doctored white noise, filtered to sound even to our ears. High frequencies are attenuated at 3 dB/octave. This "logs" the linear noise and reallocates the energy to equal parts/octave, ⅓ octave, etc.

1.1.8.4 IMPULSE

We just met random noise: all frequencies, equal level, continuous with random phase. The impulse is another special waveform: all frequencies, equal level, one cycle, in phase. Think of an impulse as all frequencies at the starting line of the racetrack. The starter pistol goes off and everybody runs one lap and stops. In fact, a starter pistol is an acoustic impulse generator used by acousticians for analysis. Listening to an impulse in a room reveals the timing and location of reflection paths. The impulse generator that won't get you in trouble with airport security is two hands clapping.

1.1.8.5 COMBINING MATCHED FREQUENCIES

Signals with matched frequencies merge to create a new waveform differing only in amplitude from the original signals that comprise it (Fig. 1.12). The resulting waveform depends upon its relative amplitude and phase characteristics, and may be greater, lesser or equal to the individual contributors. The combination of equal level-matched frequency signals will generally be additive when the phase offset is < ±120° and subtractive when between ±120° and ±180°. It is not possible, *post facto*, to separate out the original signals with filters or computation. The combined waveform of two amplitude- and phase-matched signals is indistinguishable from that of a single signal at twice the level. Conversely, the combined waveform of two amplitude-matched signals with 180° phase offset is indistinguishable from unplugged.

1.1.8.6 COMBINING UNCORRELATED WAVEFORMS (MIXING)

Unmatched (uncorrelated) complex waveforms merge together to create a new waveform with a random association to each of the original parts. The combination of unmatched complex waveforms (multiple frequencies) must be evaluated on a moment-to-moment basis. The obvious example of uncorrelated signals

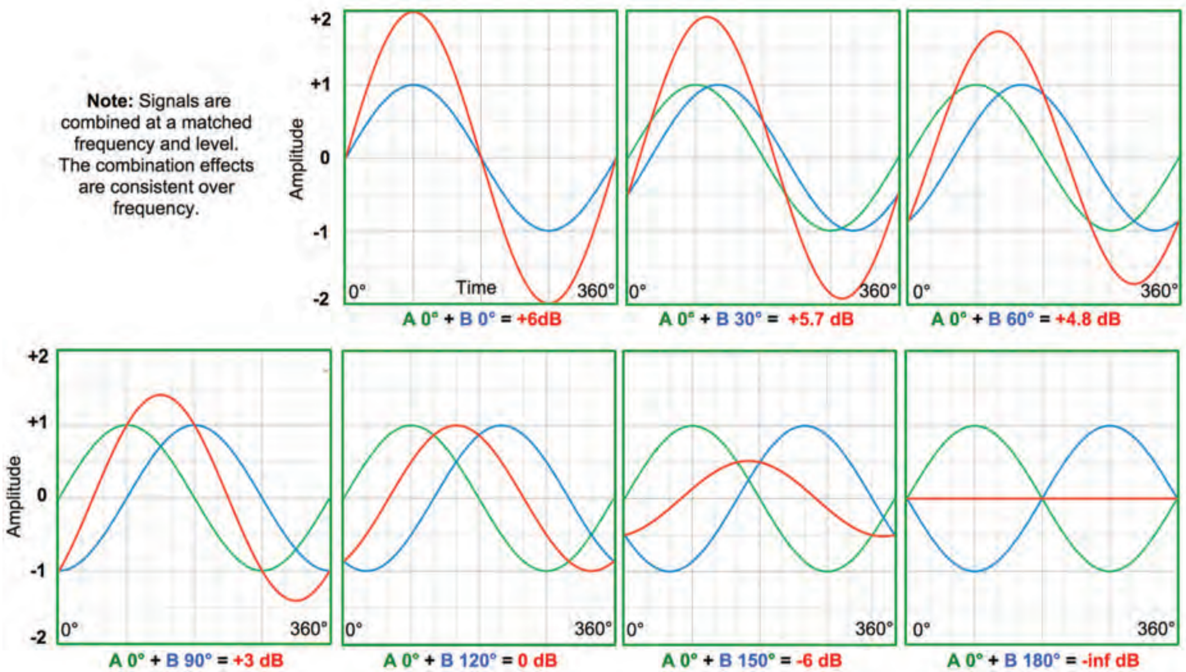


FIGURE 1.12
Combining signals with matched frequencies (correlated summation): The effects of relative phase on the combined responses are shown

is two different music streams: frequencies are in phase one moment and out of phase the next. These could be different songs, different instruments playing the same song or even violins playing the same parts in a symphony. In all cases the relationship between the signals is unstable and therefore incapable of consistent addition or subtraction at a given frequency. Combining unmatched waveforms is the essence of "mixing," and thus is separate from the combinations of matched waveforms (correlated summation), a primary concern of system optimization.

1.1.8.7 COMBINING CORRELATED WAVEFORMS (SUMMATION)

The combination of matched (correlated) complex waveforms also creates a new waveform, but with an orderly and predictable relationship to the original parts. Phase-matched signals create a combined waveform similar in shape to the individuals but with higher amplitude. If there is time offset between the signals, the new waveform will be modified in a stable and predictable way, with alternating additions and subtractions to the amplitude response over frequency (a.k.a. "comb filtering").

An example is copies of the same music stream combined in a signal processor. With no time offset, the combination will be full-range addition. Comb filtering is created by time offset between the signals. The interaction of speakers carrying the same signal in a room is more complex because the interaction must be evaluated on a frequency-by frequency and location-by-location basis (covered in depth in Chapter 4).

1.1.8.8 ANALOG FORM

Oscillation is a continuous function. We cannot get from Point A to Point B without passing through all the points in between. That is the essence of analog audio: the motion of a tuning fork, string, phonograph needle, speaker cone and more. A song, in analog form, is a series of movements in the waveform over time, always going one way or the other (+/-, in/out, etc.). If the movement stops, so does the song. If we can trace this continuous movement on one device and transfer it to another audio device, we will recognize it as the same song, even if one version came from magnetic flux (a cassette tape) and another came from the mechanical motion of a needle in a groove. Analog audio is like a continuous drawing exercise, a transcription that never lifts the pencil from the paper.

Transferring a waveform between electrical, magnetic and acoustic transmission mediums is like redrawing that pencil sketch with a different medium such as paint, stone or whatever. Notice that the term "medium" has much the same meaning in both fields.

1.1.8.9 DIGITAL FORM

Digital audio is a non-continuous function. We can only get from Point 0 to Point 1 without evaluating any points in between. Listening to 0s and 1s sounds pretty boring, even for people at raves. Digital waveforms are copies of analog waveforms, but the operation differs from the transduction process between analog mediums discussed above. Analog-to-digital converters slice the continuous signal into tiny bits (ba-da-boom), each of which represent the best fit for the momentary amplitude value. The faithfulness of the digital rendering depends on how finely we slice both amplitude and phase (time). Amplitude resolution is defined by the number of bits (24-bit is the current standard), and temporal resolution (a.k.a. sample rate) is typically 48 kHz (approximately .02 ms).

It's as if we have a photograph of our original pencil drawing. If you look close enough at the photograph you can see that there are no continuous lines, just lots of little dots. That is the essence of digital. Audio pixels. The beauty of it is that once we have the digital copy we can send it around the world without changing it. As long as we are very careful, that is. This is not even hypothetically true of an analog signal because all audio mediums have some form of degradation (distortion, frequency response variation, phase shift, etc.). Bear in mind that we can never hear digital audio. We only get that pleasure after the waveform is converted back to analog, because at the end of the day we have to use the analog air medium to get to our analog ears. This time the conversion requires our scribe to pick up the pencil again and draw one continuous connecting line between every dot in our digital photograph of the original line drawing.

1.1.9 Medium

Analog audio waveforms propagate through a medium. Within this book air molecules will be the acoustical medium and electronic charge and magnetic flux serve as the electromagnetic medium. Each medium has unique properties such as transmission speed, propagation characteristics, loss rate, frequency response, dynamic range and more. Digital audio (between analog conversions) is transmitted over a medium (not through it).

1.1.9.1 PROPAGATION SPEED

Propagation through a medium is a chain reaction. Each energy transfer takes time (an incremental latency) so the more media we go through, the longer it takes. Propagation speed is constant over frequency but variable by medium. Electromagnetic propagation is so fast we are mostly able to consider it to be instantaneous. Acoustic propagation speed is related to the medium's molecular density (higher density yields higher speeds). Sound propagates through a metal bar (very dense) faster than water (medium density), which is faster than air (low density). Sound propagation is the same speed, however, for heavy metal bars and air shows (ba-da-boom).

1.1.9.2 WAVELENGTH (λ)

An audio frequency has physical size once it exists within a transmission medium. The wavelength (λ) is the transmission speed/frequency, or transmission speed × period (T). Wavelength is inversely proportional to frequency (becomes smaller as frequency rises).

Audible wavelengths in air range in size from the largest intermodal-shipping container to the width of a child's finger, a 1000:1 range. Why should we care about wavelength? After all, no acoustical analyzers show this, and no knobs on our console can adjust it. In practice, we can be blissfully ignorant of wavelength, as long as we use only a single loudspeaker in a reflection-free environment. Good luck getting that gig. Wavelength is a decisive parameter in the acoustic summation of speaker arrays and rooms. Once we can visualize wavelength, we can move a speaker and know what will happen.

Transmission mediums				
Medium	Transmission	Speed (m/s)	Wavelength @ 1kHz	Transducers
Air	Pressure change	342	0.342	Microphone, loudspeaker
Water	Pressure change	1484	1.484	Hydrophone, loudspeaker
Iron	Mechanical vibration	5102	5.102	Accelerometer, vibrator
Electrical	Electrical charge	200k+	200+	Transformer, mic, speaker
Magnetic	Magnetic flux	200k+	200+	Transformer, mic, speaker

FIGURE 1.13
Basic properties of audio transmission mediums

1.1.9.3 TRANSDUCTION, TRANSDUCERS AND SENSITIVITY

Transduction is the process of waveform conversion between media (Fig. 1.13). Transducers are media converters. Examples include acoustic to electromagnetic (microphones), and vice versa (speakers). The amplitude values and wavelength in one media (e.g. pressure for acoustical) are scaled and converted to another media (e.g. voltage for electromagnetic). The scaling key for transduction is termed "sensitivity," which inherently carries units from both sides of the conversion. Microphone sensitivity links output voltage to input pressure (SPL), with the standard units being mv/pascal. Speaker sensitivity relates acoustic output pressure to the input power drive. The standard form is dB SPL@1 meter with 1-watt input.

1.2 AUDIO SCALES

This book is full of charts and graphs, all of which have scales. The sooner we define them the easier it will be to put them to use.

1.2.1 Linear amplitude

Amplitude is all about size. There are a million ways to scale it (or should I say there are 120 dB ways). Linear level units are seldom used in audio even though they correspond directly to electrical and acoustic pressure changes in the physical world (our level perception is logarithmic). Many engineers go their entire careers without thinking of linear sound level. Can we do a rock concert with a sound system that can only reach 20 pascals (120 dB SPL)? My mix console clips at 10 V_{RMS}. Is this normal? (Yes, that's +20 dBV.) I know a guy who sent +42 dBV @60 Hz as a test tone to some people he didn't like. You might recognize the linear version of that: 120 V_{RMS} @60 Hz. Voltage is found in many areas outside the audio path, so it helps to have bilingual fluency between linear and log.

Let's count in linear. Incremental changes from 1 volt to 2, 3 and 4 volts are sequential linear changes of +1 volt. If we started at 101 V and continued this linear trend we would see 101 V, 102, 103 and 104 V.

Now let's count the same voltage sequence in log (approximately): 1 V to 2 V (+6 dB), 2 V to 3 V (+4 dB), 3 V to 4 V (+2 dB). The total run from 1 V to 4 V is 12 dB. By contrast, the entire 4-volt sequence starting at 101 V would not even total 0.5 dB. Linear amplitude scales include volts (electrical), microbars or pascals (acoustical), mechanical movement (excursion), magnetic flux and more.

1.2.2 Log amplitude (20 $\log_{10}$ decibel scale)

The log scale characterizes amplitude as a ratio (dB) relative to a reference (fixed or variable) (Fig. 1.14). Examples include dBV (electrical) and dB SPL (acoustical). Successive doublings (+6 dB) of sound pressure are perceived as equal increments of change (a log scale perception). Therefore acoustic levels (and the electronics that drive them) are best characterized as log. Successive linear doublings of 1 microbar (to 2, 4 and 8 microbars) would be successive changes of approximately 6 dB (94, 100, 106 and 112 dB SPL).

Decibel scale ratio reference							
20*log₁₀ (Output / Input)				10*log₁₀ (Output / Input)			
Pressure (SPL), voltage (V), current (I)				Power (P)			
Gain +		Loss -		Gain +		Loss -	
Log (dB)	Ratio (Out/In)	Log (dB)	Ratio (Out/In)	Log (dB)	Ratio (Out/In)	Log (dB)	Ratio (Out/In)
0.0	1.00	0.0	1.00	0.0	1.00	0.0	1.00
1.0	1.12	-1.0	0.89	0.5	1.12	-0.5	0.89
2.0	1.26	-2.0	0.79	1.0	1.26	-1.0	0.79
3.0	1.41	-3.0	0.71	1.5	1.41	-1.5	0.71
4.0	1.59	-4.0	0.63	2.0	1.59	-2.0	0.63
5.0	1.78	-5.0	0.56	2.5	1.78	-2.5	0.56
6.0	2.00	-6.0	0.50	3.0	2.00	-3.0	0.50
7.0	2.24	-7.0	0.45	3.5	2.24	-3.5	0.45
8.0	2.51	-8.0	0.40	4.0	2.51	-4.0	0.40
9.0	2.82	-9.0	0.35	4.5	2.82	-4.5	0.35
10	3.16	-10	0.32	5.0	3.16	-5.0	0.32
12	4.00	-12	0.25	6.0	4.00	-6.0	0.25
14	5.00	-14	0.20	7.0	5.00	-7.0	0.20
15	5.63	-15	0.18	7.5	5.66	-7.5	0.18
18	8.00	-18	0.13	9.0	8.00	-9.0	0.13
20	10	-20	0.10	10	10	-10	0.10
26	20	-26	0.05	13	20	-13	0.05
32	40	-32	0.025	16	40	-16	0.025
38	80	-38	0.013	19	80	-19	0.013
40	100	-40	0.010	20	100	-20	0.010
60	1,000	-60	0.001	30	1,000	-30	0.001
80	10,000	-80	0.0001	40	10,000	-40	0.0001
100	100,000	-100	0.00001	50	100,000	-50	0.00001

FIGURE 1.14
Decibel scale reference table showing ratio conversions for the 20 log and 10 log scales

Audio system voltage level reference							
Audio signal level			Mic level	Line level	Speaker level		
Voltage (Volts)	dBV	dBu	250 Ω (Watts)	1k Ω (Watts)	16 Ω (Watts)	8 Ω (Watts)	4 Ω (Watts)
89	39	41.2			512	1024	2048
63	36	38.2			256	512	1024
45	33	35.2			128	256	512
32	30	32.2			64	128	256
22	27	29.2			32	64	128
16	24	26.2		250 m	16	32	64
11	21	23.2		125 m	8	16	32
8.0	18	20.2		63 m	4	8	16
5.6	15	17.2		32 m	2	4	8
4.0	12	14.2		16 m	1	2	4
2.8	9	11.2		8 m	500 m	1	2
2.0	6	8.2		4 m	250 m	500 m	1
1.4	3	5.2		2 m	125 m	250 m	500 m
1.0	0	2.2		1 m	63 m	125 m	250 m
707 m	-3	-0.8			31 m	63 m	125 m
500 m	-6	-3.8			16 m	31 m	63 m
356 m	-9	-6.8			8 m	16 m	31 m
250 m	-12	-9.8			4 m	8 m	16 m
178 m	-15	-12.8			2 m	4 m	8 m
125 m	-18	-15.8			1 m	2 m	4 m
89 m	-21	-18.8				1 m	2 m
63 m	-24	-21.8					1 m
45 m	-27	-24.8					
32 m	-30	-27.8	250 m				
22 m	-33	-30.8	125 m				
16 m	-36	-33.8	63 m				
11 m	-39	-36.8	32 m				
8.0 m	-42	-39.8	16 m				
5.6 m	-45	-42.8	8 m				
4.0 m	-48	-45.8	4 m				
2.8 m	-51	-48.8	2 m				
2.0 m	-54	-51.8	1 m				
1.4 m	-57	-54.8					
1.0 m	-60	-57.8					
707 µ	-63	-60.8					
500 µ	-66	-63.8					
356 µ	-69	-66.8					
250 µ	-72	-69.8					
178 µ	-75	-72.8					
125 µ	-78	-75.8					
89 µ	-81	-78.8					
63 µ	-84	-81.8					
45 µ	-87	-84.8					
32 µ	-90	-87.8					
22 µ	-93	-90.8					
16 µ	-96	-93.8					
11 µ	-99	-96.8					
8.0 µ	-102	-99.8					

FIGURE 1.15
Analog electronic operational levels: mic, line and speaker levels

1.2.2.1 ELECTRONIC DECIBEL (DBV AND DBU)

One of the most heavily enforced standards in professional audio is to never have just one standard (Fig. 1.15). The dB scale for voltage has at least twenty. Only two are used enough any more to be worth memorizing: dBV (1 volt standard) and dBu (0.775 volt standard). The difference between them is a constant 2.21 dB (0 dBV = +2.21 dBu and 0 dBu = -2.21 dBV). There is an extensive history regarding dB voltage standards going back to the telephone, which you can read somewhere else when you have trouble sleeping.

The dB scale is favored because our audio signals are in a constant state of change. We can't control the music or the musicians. We are constantly riding current levels relative to each other, to a moment ago or to the legal level allowed before the police shut us down. The dB scale is complicated but is easier than linear when trying to monitor relative levels that have a 1,000,000:1 ratio.

We should at least know the maximum voltage level for our equipment, which is usually around 10 volts (i.e. +20 dBV, +22.21 dBu). The noise floor should be in the -100 dBV range. In between we find the "nominal" value of 0 dBV (or dBu), the placement target for the audio mainstream. This leaves 20 dB of headroom above and the noise 100 dB below.

$$Level(dBV)=20\times log10 Level\ 1/1\ V$$

$$Level(dBu)=20\times log10 Level\ 1/0.775\ V$$

1.2.2.2 ACOUSTIC DECIBEL (dB SPL)

The common term for acoustic level is **dB SPL** (sound pressure level), the measure of pressure variation above and below the ambient air pressure.

$$Level(dBSPL)=20\times log10\ P/0.0002$$

where P is the RMS pressure in microbars (dynes/square centimeter).

The reference standard is 0 dB SPL, the threshold of the average person's hearing (Fig. 1.16). The limit of audibility approaches the noise level of the air medium, i.e. the level where the molecular motion creates its own random noise. It is comforting to know we aren't missing out on anything. The threshold of pain is around 3 million times louder at 130 dB SPL. The threshold of audio insanity has reached 170 dB SPL by car stereo fanatics.

The following values are equivalent expressions for the threshold of hearing: 0 dB SPL, 0.0002 dynes/cm^2, 0.0002 microbars and 20 micropascals (μPa). dB SPL is log and all the others are linear. For most optimization work we need only deal with the log form.

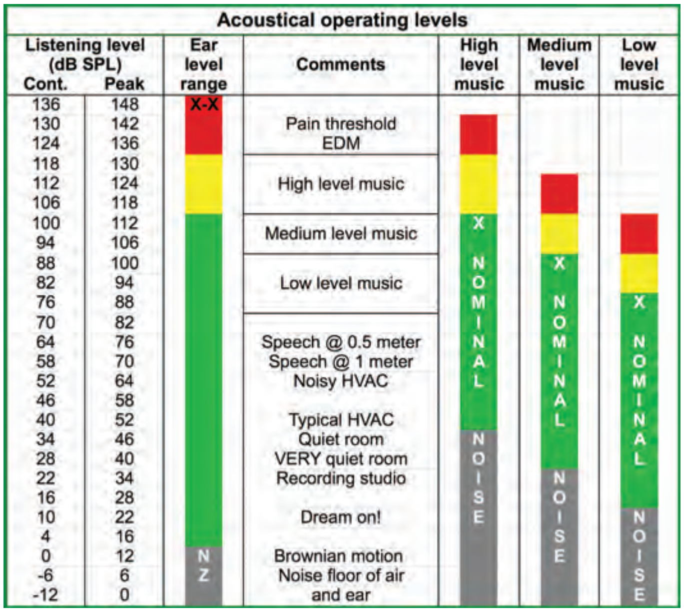


FIGURE 1.16
Acoustical operational levels: quiet, loud and insane

dB SPL subunits

dB SPL has average and peak levels in a manner similar to the voltage units. The SPL values differ, however, in that there may be a time constant involved in the calculation.

- **dB SPL peak**: The highest level reached over a measured period is the peak (dB SPL$_{pk}$).
- **dB SPL continuous (fast)**: the average SPL over a time integration of 250 ms. The fast integration time mimics our hearing system's perception of SPL/loudness to relatively short bursts. It takes about 100 ms for the ear to fully integrate the sound level.
- **dB SPL continuous (slow)**: An extension of the integration time to 1 second that is more representative of heat load and long-term exposure.
- **dB SPL LE (long term)**: This is the average SPL over a very long period of time, typically minutes. This setting is used to monitor levels for outdoor concert venues that have neighbors complaining about the noise. An excessive LE reading can cost the band a lot of money.

dB SPL weighting

There are filtered versions of the SPL scale (Fig. 1.17) that seek to compensate for the level-dependent frequency response variations in human hearing perception (the "equal loudness curves" described in section 5.1).

- **dB SPL Z ("Z" weighting)**: This is a recent trend to designate the unweighted response. Easier to spell.
- **dB SPL A ("A" weighting)**: Corresponds to the ear's response at low levels. LF range is virtually nonexistent. Applicable for noise floor measurements. Often used as a maximum SPL specification for voice transmission. Subwoofers on or off goes undetected with A-weighted readings.
- **dB SPL B ("B" weighting)**: Corresponds to the ear's response at intermediate levels. LF range is rolled off but not as much as A weighting (-10 dB @60 Hz). Applicable for measurements using music program material.
- **dB SPL C ("C" weighting)**: Corresponds to the response of the ear at high levels. Close to a flat response. Applicable for maximum-level measurements. Used as a specification for full-range music system transmission levels. Subwoofers on or off will have a noticeable effect when C weighting is used.

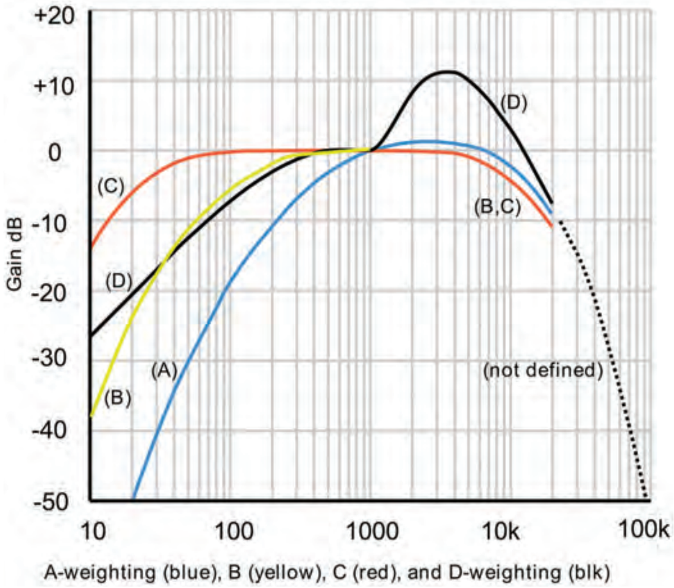


FIGURE 1.17
Frequency response curves for ABC and D weighting. Note: D weighting is shown here but not typically used for our applications (graphic by Lindosland @ en.wikipedia, public domain, thank you).

1.2.3 Power (10 $\log_{10}$ decibel scale)

Power is derived from a combination of parameters (e.g. voltage and current or pressure and surface area). The 10 $\log_{10}$ formula is the log conversion of power ratios (also Fig. 1.14). It's rarely used in system optimization because the analyzers monitor singular parameters such as voltage and SPL rather than power (in watts). It's better to prioritize our limited memory space for the 20 $\log_{10}$ formula and leave the 10 $\log_{10}$ for Google.

1.2.4 Phase

The standard scale for phase spans from 0° to 360°. The most common display places 0° at center and ±180° on the upper and lower extremes, but other options are available. Phase scaling is circular and therefore very different from amplitude. When we get more amplitude we simply expand the scale. We don't go up to 10 volts and then start over at 0 if we go higher. Phase is different because it tops out at 360°. When phase shift exceeds 360° it wraps around as an alias value within the 360° limits (e.g. 370° reappears as 10°). This is similar to an automobile race where cars on the lead lap are indistinguishable from those a lap behind. We only know who will win the race from other evidence, such as watching the entire race, not just the last lap. For audio phase our evidence will be similarly provided, but in this case it's about watching the phase values over the whole frequency response rather than a single frequency.

Phase serves us poorly as a unit of radial measure. Radians are the more "calculation-friendly" unit for radial angle found inside formulas that require accurate phasor positioning. Radians are rarely used by audio engineers for system optimization (or in a sentence for that matter). A complete cycle (360°) has a value of 2 π radians so beware of radian cancellation if two sources fall π radians out of sync (180°).

We learn to memorize the 360° phase scale even though it seems an arbitrary division for a circular function. It is a vestige of a merciful rounding error by ancient Egyptian mathematicians. Let's be grateful it's not 365° with a leap-degree every four cycles.

1.2.5 Linear frequency axis

Linear frequency scaling shows equal spacing by bandwidth (unequal spacing in octaves). The linear scale is an annoying, reality-based construction that corresponds to how frequency and phase interact in the physical world. The spacing between 1 kHz, 2 kHz, 3 kHz and 4 kHz (consecutive bandwidths of 1 kHz) is shown as equal spacing. Phase, the harmonic series and comb filter spacing all follow the linear frequency axis. The frequency resolution of the Fourier Transform (the math engine of our analyzer) is linear. For example, a set of 100 Hz wide filters with 100 Hz spacing is termed "100 Hz resolution."

1.2.6 Log frequency axis

A log frequency scale shows equal spacing in *percentage* bandwidth (octaves), and unequal spacing in bandwidth. This corresponds closely to our perception of frequency spacing. The equal linear spacing between 1 kHz, 2 kHz, 3 kHz and 4 kHz are percentage bandwidths of 1 octave, ½ and ⅓ octave respectively. A true log frequency response is made of log spacing of log filters. For example, a set of ⅓ octave filters at ⅓ octave spacing is termed "⅓ octave resolution."

1.2.7 Quasi-log frequency axis

The quasi-log frequency scale is a log/linear hybrid (log spacing of linear bandwidths). Stretching a single linear response over the full audio range is not practical for optimization because there is not enough data in the lows and/or too much in the highs. Instead a series of (typically) eight octave-wide linear sections are spliced together to make a quasi-log display. This is implemented in almost every modern analyzer used for optimization. (Full details are in Chapter 12.) Each octave of the quasi-log frequency response is derived from log spacing of the linear resolution (e.g. 1/48 octave spacing of 48 data points). This is termed "48 points/octave resolution."

1.2.8 Time

Tick, tick, tick. It seems strange to have to write that time is linear (evenly spaced increments) and not log (proportionally spaced increments). This is only mentioned because the frequency response effects of time offsets are entirely linear but are perceived by our log brains. So let's put this to rest: There is no log time.

1.3 CHARTS AND GRAPHS

Let's put the scales together to make the charts and graphs we use for design and optimization. Fluency in reading these graphs is a mandatory skill in this field. The graphs are 2-D, with an *x*-axis and *y*-axis. We generally find frequency and time on the *x*-axis and amplitude, phase and coherence on the *y*-axis.

1.3.1 Amplitude (*y*) vs. time (*x*)

Amplitude vs. time is a peak–peak waveform tracing commonly seen on oscilloscopes or digital audio editors (Fig. 1.18). The amplitude scale can be linear or log but time is only linear.

1.3.2 Amplitude (*y*) vs. frequency (*x*)

Absolute level over frequency is used to check the noise floor, the incoming spectrum, harmonic distortion, maximum output and more. The *y*-axis shows level against a fixed standard. We can observe the individual channels used to make transfer function computations (next item).

1.3.3 Relative amplitude (*y*) vs. frequency (*x*)

Relative level over a quasi-log frequency scale is the most common graph in system optimization (Fig. 1.19). This transfer function response is used for level setting, crossover alignment, equalization, speaker positioning and more. The *y*-axis scales unity level to the center and shows gain above and loss below (in dB). The quasi-log frequency scale is preferred because of its constant high resolution (24 to 48 points/octave) and close match to human hearing perception. An alternative option, the linear frequency scale can help identify time-related mechanisms (such as phase, comb filtering, reflections, etc.).

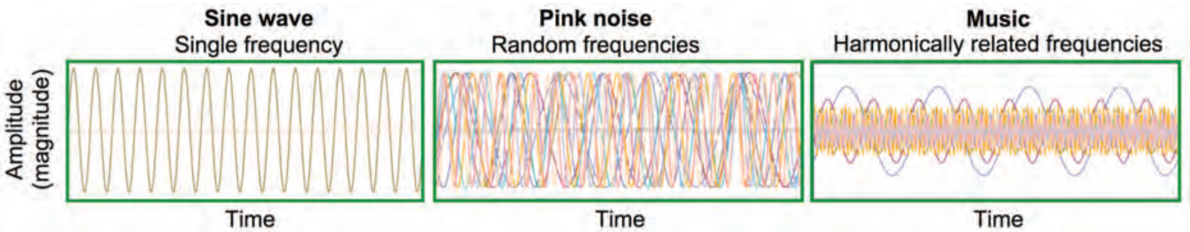


FIGURE 1.18
Amplitude vs. time plots for various waveforms

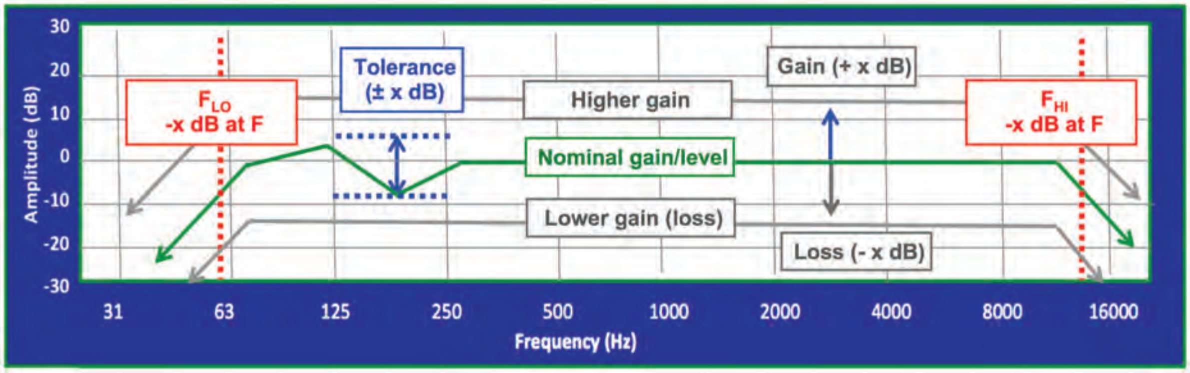


FIGURE 1.19
Introduction to the relative amplitude vs. frequency (quasi-log) display

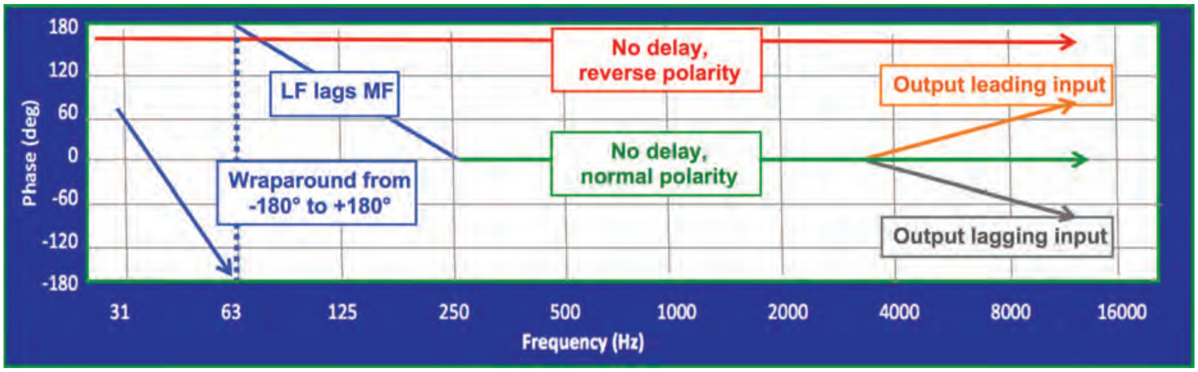


FIGURE 1.20
Introduction to the relative phase vs. frequency (quasi-log) display

1.3.4 Relative phase (y) vs. frequency (x)

This is our standard phase display (Fig. 1.20). The y-axis shows a 360° span. The vertical center is typically 0° but can be normalized around any phase value. A flat phase response (horizontal line) indicates zero phase shift and zero time offset over the frequency range shown. Variations from flat phase response indicate some time offset, either full band (i.e. latency) or frequency–dependent (phase delay). A downward slope (left to right) indicates positive delay whereas an upward slope indicates negative delay.

A constant phase rotation at linear frequency intervals indicates latency (frequency-independent delay). A quasi-log display shows the slope steepening with frequency (a linear function in a log display). Comb filtering appears as increasingly narrowing spacing of peaks and dips as frequency rises (again, linear function, log display). Filters create frequency-dependent delay. The phase response of a filter with a given Q will maintain the same shape (slope) as frequency rises.

Note: Don't freak out if you don't understand phase yet. This section is intended to provide just enough info to help read phase traces as we go along. If phase were easy, this book would be five pages long.

Relative phase on a linear frequency scale is less popular, but more intuitive than log. The phase slope over a linear frequency scale clearly reveals the relationship of phase and time (a linear mechanism on a linear scale). A flat phase response indicates no time difference at any frequency, just as the log display. Latency creates a constant phase slope as frequency rises. Comb filtering appears as consistently spaced peaks and dips as frequency rises. The term "comb filtering" comes from its linear frequency scale appearance.

1.3.5 Impulse response: relative amplitude (y) vs. relative time (x)

The impulse response is a favorite of the modern analyzer (Fig. 1.21). We can find delay offsets between speakers with extreme accuracy in seconds. Follow the dancing peak, read the number in ms. Done! Magic! And it seems like magic to most of us, even more so when we stop to think about what this computation is.

The FFT analyzer impulse response is a mathematical construction of the picture we would see on a hypothetical oscilloscope (amplitude vs. time) with a hypothetical single pulse. In practice we get relative amplitude vs. relative time. The full story on what's under the hood will have to wait until section 12.12. For now we will focus on how to read it.

Relative level (y-axis) is not like our amplitude over frequency. The vertical center is silence, not unity gain. Unity gain (normal polarity) is shown as an upward vertical peak at a value of 1 (0 dB) and positive gain is a bigger peak (and loss is smaller). A downward peak indicates polarity inversion.

Time (x-axis) is relative, a comparison of output–input arrival times. A centered impulse indicates synchronicity. The peak moves rightward when the output is late and vice versa. Yes, it is possible to have the output before the input in our measurements, because we can delay signals inside the analyzer.

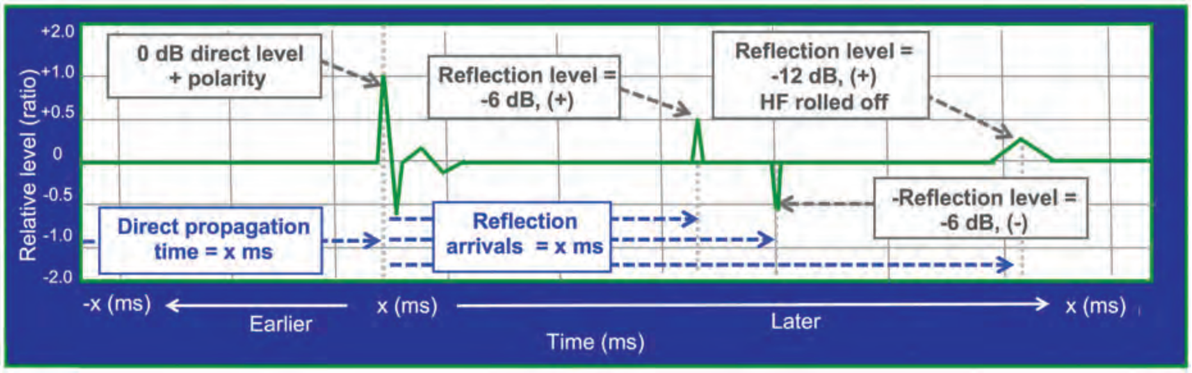


FIGURE 1.21
Introduction to the impulse response display

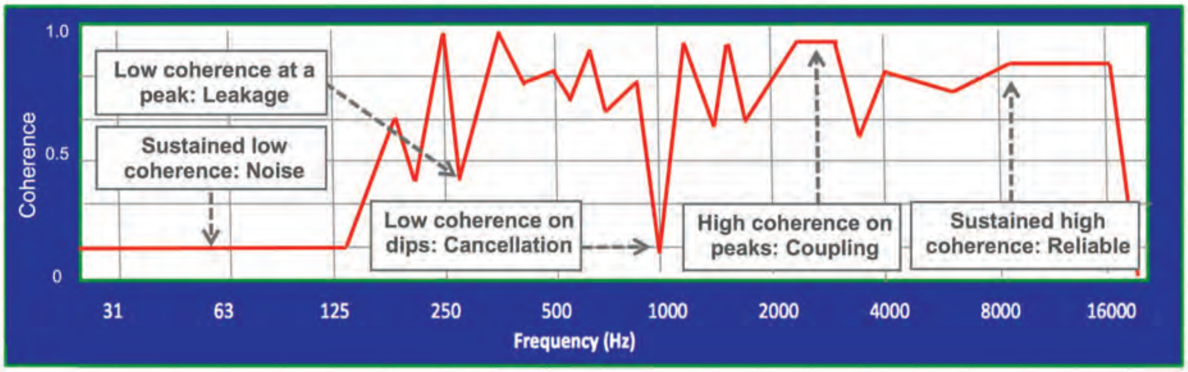


FIGURE 1.22
Introduction to the coherence vs. frequency display

A perfectly flat frequency response (amplitude and phase) makes a featureless impulse shape (straight single line, up and down). The pulse will have ringing, overhang and various other distortions to its shape if the measured device's response is *not* flat. Reflections appear as secondary impulses on the display. Like phase, this is enough information to enable us to read the displays going forward.

1.3.6 Coherence vs. frequency

The coherence function is a data quality index that indicates how closely the output signal relates to the input (Fig. 1.22). Amplitude and phase data are deemed reliable when coherence is high and unreliable when low. Coherence alerts one to take the wooden shipping cover off the speaker front instead of boosting the HF EQ. Yes, a true story.

Coherence is derived from averaging dual-channel frequency responses and is indicative of data stability. A value from 0 to 1 is awarded based on how closely the individual samples match the averaged value in amplitude and phase. Details are in section 12.11.

1.4 ANALOG ELECTRONIC AUDIO FOUNDATION

1.4.1 Voltage (*E* or *V*)

Voltage, electrical pressure, can be characterized linearly (in volts) or logarithmically in dB (dBV, dBu, etc.). The electronic waveform is a tracing of voltage vs. time. Voltage is analogous to acoustical pressure.

1.4.2 Current (*I*)

Current is the density of signal flow through an electronic circuit. As current increases, the quantity of electron flow rises. Analog audio transmission between electronic devices (except amplifiers to speakers) requires minimal current flow.

1.4.3 Resistance (*R*)

Resistance restricts current flow in an electronic circuit. As resistance rises, the current flow (for a given voltage) falls. Resistance is frequency independent (impedance is not).

1.4.4 Impedance (*Z*)

Impedance is the frequency-dependent resistive property of a circuit, a combination of resistance, capacitance and inductance. Impedance ratings are incomplete without a specified frequency. Output/input impedance ratios play a critical role in the interconnection of analog electronic devices, amplifiers and speakers, determining the maximum quantity of devices, cable loss and upper and lower frequency range limits.

An 8 Ω speaker (nominal impedance) illustrates the difference between impedance and resistance. The DC resistance is 6 Ω. The lowest impedance in its operating range is around 8 Ω. Impedance rises (and output level falls) at frequencies above its operating range.

1.4.4.1 CAPACITANCE (CAPACITIVE REACTANCE)

Capacitors pass signal across conductive, but unconnected, parallel plates. DC cannot flow across the gap in the plate. An AC signal, however, can flow across the plate (resistance falls as frequency rises). Capacitors approximate an open circuit (maximum impedance) to DC and short circuit (minimum impedance) to AC. Capacitance in series rolls off the LF response (resistance is inversely proportional to frequency). Parallel capacitance (such as between wires of an audio cable) rolls off the HF via a shunt path to ground.

1.4.4.2 INDUCTANCE (INDUCTIVE REACTANCE)

Inductor coils resist voltage changes in the signal. An unchanging signal (DC) passes freely but the inductor becomes increasingly resistant as frequency rises (the rate of electrical change increases). The inductor's response is the opposite of a capacitor: maximum impedance to AC signals and the minimum impedance to DC. Series inductance increasingly rolls off the HF response. Parallel inductance shunts the LF to ground.

1.4.5 Power (P)

Electrical power (in watts) is the combined product of voltage, current and impedance. Ohm's law expresses the relationship of these three parameters (Fig. 1.23). Most electronic transmission involves negligible power (low voltage, low current and high impedance). Speaker-level transmission requires substantial power (high voltage, high current and very low impedance). Acoustical power, also in watts, is produced by pressure, surface area and acoustic impedance (inertance). Our ears (and microphones) are pressure sensors, not power sensors, which characterize sound level by pressure only (SPL).

1.4.6 Operating levels

Signal levels are divided into three categories by voltage and impedance. A lucky waveform can experience all three (return to Fig. 1.15) if beginning at mic level, going through a "pre" amp to line level and rising to speaker level in a power amplifier.

1.4.6.1 MIC LEVEL

Mic-level signals are typically generated by small passive devices such as microphone coils, guitar pickups, phonograph cartridges, etc. Mic level is a matter of necessity, not choice. There are few advantages and many disadvantages to operating in the microvolt range. Signals are vulnerable to induced noise and other complications

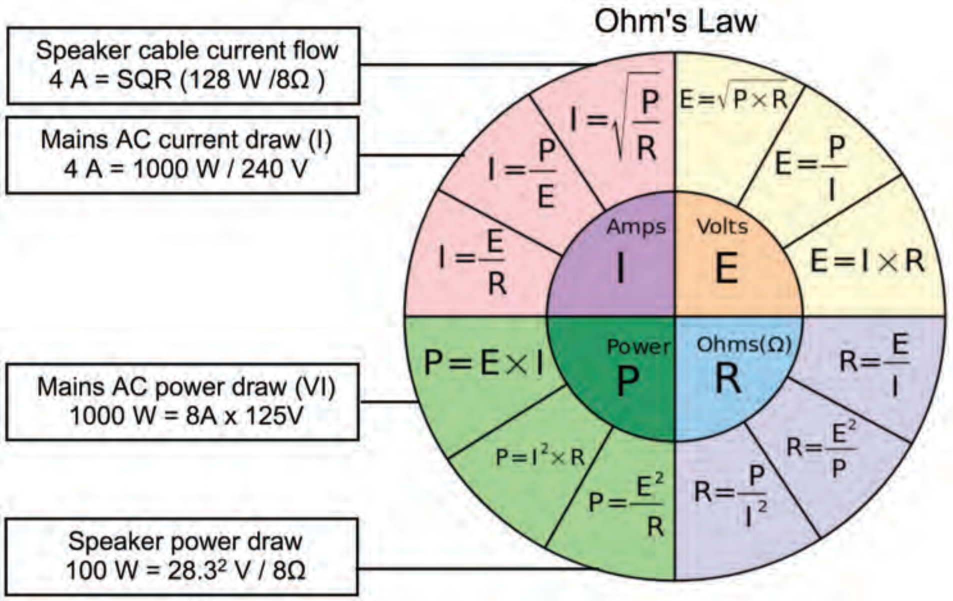


FIGURE 1.23
The Ohm's law pie chart with some examples applicable to audio systems (chart authored by Matt Rider, http://commons.wikimedia.org/wiki/File:Ohm's_Law_Pie_chart.svg)

relating to extremely low voltage and current flow (e.g. jumping connector and relay gaps). The winning strategy is to preamplify mic-level signals to line level as soon as possible. The worst-case scenario, unbalanced, high-impedance mic level (e.g. a guitar pickup), requires the shortest possible preamp path. Mic-level sources generate signals in the µV to 100 mV range with nominal source impedances of a few hundred ohms (microphones) and a few k Ω (pickups).

1.4.6.2 LINE LEVEL

Active devices, such as consoles, processor, instrument direct outs, playback equipment and more, usually generate line-level signals. This is the standard operating range, with nominal levels in the 1 V range, and maximum levels over 10 V. Balanced line-level low-impedance outputs (150 Ω typical) driving high-impedance inputs (10 kΩ typical) should have good noise immunity and minimal loss.

1.4.6.3 SPEAKER LEVEL

Power amplifiers are the exclusive generators of speaker-level signals. There is no "nominal" speaker level. Instead, speaker level ratings refer to the maximum power capability (watts). Power ratings require two known parameters: voltage and impedance. For example, a 100-watt amplifier with an 8 Ω speaker load can generate 28.3 volts, whereas a 400-watt amp will reach 56.6 volts.

1.4.7 Analog electronic audio metrics

This is an overview of some relevant standard specifications for professional-grade analog audio devices (Fig. 1.24).

1.4.7.1 FREQUENCY RESPONSE/RANGE (±dB/Hz)

There are two main categories for frequency response: range and deviation. Range is the spectral area between the limits, typically the half-power points (-3 dB points) at the LF and HF extremes. Response deviations are given as ±×dB for the area within the optimal range. Electronic device deviations are normally <±1 dB.

1.4.7.2 MAXIMUM LEVEL (dBV, dBu)

Any electronic device has a maximum voltage limit: the clipping point, typically specified at 1 kHz. This parameter helps ensure compatibility between interconnected devices to ensure full dynamic swing through the signal chain. It is preferable to have all devices clip at around the same level, lest the voltage swing become limited by the weakest link.

1.4.7.3 NOISE FLOOR (dB)

Let's measure a device unplugged (or with its input shorted). What's left is the device's residual self-generated noise floor. This is typically given as the worst case across the spectrum and falls into two categories: hum (harmonic multiples of the line frequency) and noise (white noise). Noise floor measurements are usually shown as "A"-weighted values.

1.4.7.4 DYNAMIC RANGE (dB)

Dynamic range is the span between the maximum-level capability and noise floor. Devices with a single gain stage have straightforward dynamic range specifications. Devices with multiple gain stages can vary their dynamic range by offsetting the internal settings and thereby decrease the maximum capability and/or increase the noise floor. In other words, cranking up the input while cranking down the output is likely to change the dynamic range through the device.

1.4.7.5 POLARITY (+/-)

Every device has one of two polarities: right (normal) or wrong (inverted). This should be a non-issue for balanced-input/balanced-output line-level devices. Polarity must be specified for any device with unbalanced inputs (e.g. DJ mixers) or outputs (e.g. power amplifiers).

1.4.7.6 VOLTAGE GAIN (SENSITIVITY)

A device passing audio has voltage gain, even if it's loss. Sound strange? The term "voltage gain" connotes the output/input voltage ratio. It's positive gain when the output is larger than the input and negative gain when less. Unity (0 dB) is the expected default gain in line-level devices. Power amplifiers require positive voltage gain, specified directly in dB or as sensitivity, i.e. the input voltage level required to reach full output power.

1.4.7.7 MAXIMUM POWER RATING (WATTS)

This is the maximum amount of power a device output can drive to another device input (the load). This specification is only used for power amplifiers and the speakers they drive.

1.4.7.8 TOTAL HARMONIC DISTORTION (THD)

Harmonic distortion is the addition of uninvited frequencies to the original waveform. Harmonics are linear multiples of the original transmitted frequency. Every transmission system adds distortion; the question is how much, and the answer is given in percentage of total harmonic distortion (%THD). Analog electronic devices normally exhibit a fairly consistent percentage THD over level and frequency. Therefore the specification is normally given at 1 V_{RMS} (line level) or rated output power (amplifiers) at 1 kHz.

1.4.7.9 INTERMODULATION DISTORTION (IMD)

Audio signals are generally (and hopefully) more complex than a simple sine wave. Therefore our system must remain linear while reproducing complex waveforms as well as simple sinusoids. Intermodulation distortion has the potential to arise when two or more sine tones are simultaneously reproduced. Where harmonic distortion generates spurious frequencies by multiplication, intermodulation does so by addition and subtraction. IMD products are difference tones related to the linear spacing between the signals. For example, a mix of 60 Hz and 1 kHz would potentially show IMD products at 940 Hz and 1060 Hz. In this case 60 Hz is, in essence, modulating 1 kHz, hence the name. Loudspeaker motion must track complex waveforms so IMD testing is one way to separate the men from the boys (or the under-seat FX generators from real speakers).

Analog electronic metrics				
Category	**Measurement**	**Typical**	**Results**	**Applications**
Range/tolerance	Amplitude vs. frequency	±3 dB points relative to nominal, ±1 dB	F_{LOW} to F_{HI} ± x dB	Sets the operating range for design
Maximum input/output	Sine wave input at maximum output	Clip point of either input or output	x dBV, dBu or volts for the input and output	Provides design info regarding interconnection and cable runs
Noise	Residual noise with no input (or shorted input)	White noise and hum	x dB (A weighted)	Provides performance criteria for design
Phase	Phase shift or phase delay over frequency	< ± 60° within operating frequency range	± x degrees (or ms) over a specified frequency range	Provides performance criteria for design
Distortion	THD, IMD, TIM	< 0.05 % at nominal level	x % @ nominal level (e.g. 1v) at 1 kHz or other freq	Provides performance criteria for design
Impedance	Input and output	Input > 5kΩ, Output < 250Ω	Input = x Ω, Output = x Ω	Provides design info regarding interconnection, cable runs, and parallel options

FIGURE 1.24
Standard metrics for the evaluation of analog electronic devices

The Society of Motion Picture and Television Engineers (SMPTE) standard for IMD testing is 60 Hz and 7 kHz mixed at a 4:1 ratio (-12 dB).

1.5 DIGITAL AUDIO FOUNDATION

Digital audio is the numerical rendering of an analog waveform constructed from a series of evenly spaced, end-to-end time records. The data are transmitted in non-continuous packets and reassembled for further processing, re-transmission or conversion to analog.

1.5.1 Numerical value (analogous to amplitude)

The waveform is traced as a series of numerical values, i.e. bytes (the more modern term is "octets" because bytes most commonly are 8-bit words), spaced evenly in time (the sample rate). The number can be referred to a fixed-point standard or floating-point. Fixed-point values can be linear or log (in dB), relative to the full-scale reference.

1.5.1.1 FIXED-POINT

"Fixed-point" describes the process of carrying the amplitude value as an integer, limited by the possible bit combinations. Increments are $2n^{th}$ power, with n being the number of bits. Sixteen-bit audio uses a sign bit (+/-) and 15-signal level bits for a total of 65,536 possible values (from 0000000000000000 to 1111111111111111). Twenty-four-bit has 16,777,216 iterations. Fixed-point topology has an absolute upper limit, known as "full-scale digital," and a lower limit, set by the "bit depth."

1.5.1.2 FLOATING-POINT

Floating-point math renders the amplitude value as an integer of fixed length and an exponent and direction (+/-) vector. The multiplication capability allows numbers to exceed the fixed-point topology's simple binary bit limits. Floating-point systems can carry "out of scale" numbers for internal processing functions, but must revert to "in scale" at the point of D/A conversion or transmission to a fixed-point audio device. Floating-point numbers have virtually unlimited size, but not unlimited precision. The precision is limited by the resolution of the mantissa (the numbers that precede the exponent, e.g. 1.3×10^6 vs. 1.33×10^6 vs. $1.3333333333333333333333 \times 10^6$).

Hey kids, what's the highest number in the universe? The fixed-point child answers "10 gazillion." The floating-point kid answers "$10^{GAZILLION}$."

1.5.2 Current, resistance, impedance and power

Digital audio can be transmitted in electrical form but we won't be using an ohm meter to check continuity. There is voltage and current flow between digital devices, but they relate to the transmission of bits, not the waveform. In essence, an electronic digital audio system has two electrical states (e.g. 0 or +5 V_{DC}), in contrast with the continuous 120 dB range of analog. Digital audio connections are typically singular, point-to-point and made through a cable with a characteristic impedance of 75 Ω or 100 Ω. The key is to connect compatible devices with an impedance-matched cable under its specified maximum length.

1.5.3 Digital audio metrics

Here is an overview of the established parameters for digital audio devices (Fig. 1.25).

1.5.3.1 SAMPLE RATE

Sample rate sets the frequency range upper limit. The highest usable frequency must not exceed half the sample rate (the Nyquist frequency) to prevent calculation errors. Therefore steep filters are employed around the Nyquist frequency to prevent aliasing errors. A 48 kHz sample rate yields a frequency range up to 24 kHz and a 96 kHz sample rate yields 48 kHz of audio bandwidth.

1.5.3.2 FULL-SCALE DIGITAL

The maximum numerical value for the device at the conversion point or transmission to another device is termed "full-scale digital." The maximum value for a 24-bit signal is 16,777,215. Full-scale digital is equivalent to the analog clip point.

1.5.3.3 LEAST SIGNIFICANT BIT

Semiconductors and other analog electronic components constantly emit some level of molecular level noise. Not so with the 1s and 0s of digital. The digital floor is the sound of the lowest bit's inability to make a firm decision, a rounding error known as quantization noise. The least significant bit (LSB) has the difficult task of determining whether to round the tiniest signals up or down. Quantization noise is the sound of flip-flopping, which can be more disturbing to listeners than old-fashioned analog hiss. Adding low-level white noise to the signal, a process known as "dithering," can mask audible quantization noise. Today's 20-bit and 24-bit systems have moved the quantization noise under the analog noise floor, which reduces the need for dithering.

1.5.3.4 BIT DEPTH

Bit depth is the analog for dynamic range. In fixed-point digital this is the level difference between the maximum capability (full scale) and the noise floor (the LSB). We approximate dynamic range as bit number × 6 dB. Each time we add a bit to the resolution we are able to slice the finest increment in half, which is equivalent to 6 dB in terms of amplitude. For example, 16-bit audio can reach a dynamic range of 96 dB (16 × 6 dB), which is less than a well-designed analog circuit. By contrast, a 24-bit system has a potential of 144 dB, a very high bar for an analog circuit to reach.

1.5.3.5 dB FULL SCALE (dBFS)

Recall that sensitivity is the conversion factor for acoustic/electric transduction. dBFS plays this role between analog and digital. The process has three stages: voltage to number to voltage (dBV to full scale to dBV). dBFS is the answer to the question: "How many volts equals all bits set to 1?" Let's set an example A/D converter to full scale at 14.14V^{PK} (10 V_{RMS} for a sine wave). This is +20 dBV = 0 dBFS, a sensible choice that matches the digital maximum to the typical analog maximum. It's easy to get confused about this because our analog side thinks of 0 dB as "nominal" (with 20 dB of headroom remaining for peaks) whereas the digital world sees 0 dBFS as maximum with 20 dB of legroom below this to reach the "nominal" value (see Fig. 3.29 for reference). Confusion on this concept can cost a pile of dynamic range. We can keep it simple by following the industry standard for dBFS, which is . . . just kidding. There are too many to count.

Digital audio metrics				
Category	Measurement	Typical	Results	Applications
Sample frequency	Number of samples taken for A/D conversion	48 kHz, 96 kHz	HF limits of 22 kHz and 44 kHz respectively	Higher resolution more closely approximates the analog signal but also requires more data transfer speed
Bit depth	Number of bits sets dynamic range	20-24 bit	Seconds, octave bands 125 Hz to 8 kHz	Bit depth sets the digital dynamic range limit. Ideally this will equal or exceed the analog limits
dB FS	Sensitivity conversion factor for voltage to full scale digital	No standard	x dBV, dBu or Volts = Full scale digital	Standardized dB FS values enables maximization of dynamic range of both analog and digital systems
Format	The standard packet for audio file transmission	AES3 or multichannel compatible versions	Standard AES3 64-bit audio packet	Industry wide

FIGURE 1.25
Standard metrics for the evaluation of digital audio devices

RMS and peak

Recall that the digital numerical sequence traces the waveform's peak–peak structure. There is no such thing as an rms audio waveform, either analog or digital: ONLY peak (negative and positive). Rms values are mathematical calculations for heat dissipation or integrations of perceived loudness. Are we worried about cooking the numbers? That's the accountant's job. Loudness doesn't matter in the numbers game. It is simply a matter of fitting the waveform within the limits of the counter (the digital pipe). Our concern is headroom, the remainder between the positive or negative peak and the full-scale number. Now how do we relate them?

dBFS meets the sound system

The key is to make sure we match the reference points within our terminology. We have to link nominal with nominal and/or maximum with maximum. 0 dBV is nominal line level and +20 dBV is maximum analog. 0 dBFS is maximum digital and -20 dBFS is "nominal." The mistake would be linking this 0 dB analog to this 0 dB digital because they are nominal and maximum respectively. We can link maximum to maximum with settings of +20 dBV = 0 dBFS, more typically expressed as 1 V_{RMS} (0 dBV) = -20 dBFS.

We can reasonably expect to find systems in the range of -16 to -20 dBFS.

1.6 ACOUSTICAL FOUNDATION

We now enter the world of acoustics. Because we are only covering overwater acoustics we can boil this down to two things: air and surfaces. Air, the medium, is the easy one, with relatively few parameters in play. The surface part is the challenge: floors, ceilings, walls, I-beams, movie screens, balcony fronts and even our paying patrons. Anything that the sound hits, bends around, reflects off or pushes its way through before it has decayed below our hearing threshold goes in the category of acoustic surfaces.

1.6.1 Sound propagation

Sound moves through the air medium as pressure variations propagating spherically from the source at a speed of approximately 1 meter every 3 ms (1.1 ft/ms). The surface area increases and sound pressure level falls as the wave propagates outward. There is no distance limit to the propagation but the frictional and pressure losses eventually decrease the pressure variations to the point of inaudibility and noise floor of the air itself (Fig. 1.26).

1.6.1.1 ACOUSTIC AMPLITUDE (SPL)

Sound pressure level (SPL) is analogous to voltage in the electronic waveform. The standard free-field loss rate is 6 dB/doubling of distance from the source.

1.6.1.2 ACOUSTIC CURRENT (SURFACE AREA)

The waveform stretches over an increasingly large area as sound propagates outward. Acoustic power remains constant (neglecting frictional heat loss) so pressure falls as surface area expands.

1.6.1.3 ACOUSTIC IMPEDANCE (INERTANCE)

Acoustic impedance is the static pressure resisting the transmission of our waveform through the elastic medium of air. Speakers must generate actual power to overcome this resistance (unlike microphones, which can just sense the pressure). Air impedance varies slightly over temperature and altitude, but not enough to make us design mountaintop sound systems different from valley sound systems. The acoustic impedance in outer space is infinitely low (a vacuum), where it's easy for speakers to move but hard to make noise.

1.6.1.4 ACOUSTIC POWER (WATTS)

Acoustic power is analogous to electric power derived from the acoustic properties of pressure, surface area and inertance (voltage, current and impedance). The acoustic power generated by a loudspeaker dissipates gradually by friction heat loss but otherwise remains nearly constant as it propagates spherically outward. It's easy to miss this because our ears sense pressure (which falls rapidly over distance), not power. A stun grenade has almost as much far-field acoustic power as near-field, but proximity makes a big difference in how we experience it (175 dB SPL peak @1 m).

The difference in acoustic power between senders and receivers is found in the surface area (acoustic current), not the SPL. A speaker generating 94 dB SPL at 10 m fills a surface area of 1256 sq/m. By contrast the mic senses the same SPL over the surface area of its diaphragm.

1.6.2 Direct sound

The term "direct sound" refers to the first arrival, free and clear of any surface reflection paths. It's a straight path unless bent by wind, refraction or diffraction around a relatively small object. The direct sound path is the most important trajectory in sound system design due to its strong effect on perceived uniformity. Analyzers used for system optimization are highly focused on the direct sound response and early reflections, with varying degrees of inclusion of the late reflections.

Acoustical propagation		
Parameter	**Properties**	**Notes**
Spherical propagation	Propagates at equal speed and loss rate in all directions	Directional speakers have an initial pressure imbalance over surface area, which holds over distance
Speed	Set by elasticity of air, approximately 344 m/sec	Changes slightly with temperature and ambient pressure
Loss rate	Pressure falls as surface area expands. 6 dB loss (pressure) for each doubling of distance	Changes in the HF range with humidity. Affected by boundaries and reflections
Power	Power required to overcome the inertance of the medium (acoustic resistance)	Remains constant over distance except for frictional losses. Pressure loss is traded for expanded surface area
Dynamic limits	Reduced linearity at extreme high pressures. Brownian noise (air molecule noise floor). Cannot transmit in a vacuum.	The negative side of the waveform at 194 dB SPL is -1 atmosphere. Therefore essentially a vacuum for the negtive side of the waveform, causing acoustical clipping
Reflection	Sound bounces off large surfaces (relative to wavelength)	Angle of incidence = angle of reflection (like a mirror)
Refraction	Propagation path bends when passing through medium at different speeds	Air temperature affects speed so sound bends when passing through cold and warm air gradients
Diffraction	Large wavelengths bend around small objects or through openings	Otherwise we could hear our neighbor's television.

FIGURE 1.26
Parameters of acoustic propagation

1.6.3 Reflected sound

Reflections are the direct sound's children. They are undeniably related, always late and interrupt speech. Although there is only a single direct sound path, the number of reflected sound paths can be almost infinite. Acousticians and sound engineers categorize reflected sound into "early" and "late" differently. The acoustician's focus is on the long-term statistical reflection model, the room's unitary decay characteristic. They love reflections because they would not have a job without them. The sound engineer sees the direct sound distribution as primary and hopes the reflections don't do too much damage.

1.6.3.1 EARLY REFLECTIONS

Early reflections arrive close behind the direct sound. This small but critical minority of the overall reflections has strong effects on tonal perception and acoustic gain. Their paths are analyzed individually more than collectively (especially by sound engineers).

1.6.3.2 LATE REFLECTIONS

Late reflections are the tail end of the sound. There's lots of gray area around the division between early and late reflections (see section 5.3). For now let's simply say that early reflections fuse closely enough with the direct signal to be perceived as adding loudness. Late reflections only stretch time, which is characterized statistically. Ideally the late reflection decay character matches the venue's program material.

1.6.4 Room acoustical metrics

The parameters of room acoustics are well established. These metrics describe the room "unplugged," i.e. without a sound system. The numbers are useful to room designers, especially for this endangered species: the room without a sound system. Nonetheless, this is the bottom line for sound system design and optimization: These numbers have very little bearing on our decision-making process. Knowing "strength" or "bass ratio" or whether the decay is double sloped changes nothing for us. We still point the speakers at the audience, not the walls. More on this topic as we go (Fig. 1.27).

1.6.4.1 REVERB TIME (RT 60)

Reverb time is the interval for a signal to decay 60 dB after cessation. This is the oldest and most common singular room characterization metric. The measurement was conducted historically by exciting the room with a loud impulse such as a balloon or acoustic (starter pistol) and charting the decay response. This can now be done computationally by a variety of methods and analyzers.

1.6.4.2 REVERB TIME (T30, T20, T15)

We can measure a shorter sample of time and then extrapolate the results to the RT60 time frame. T30 cuts the time in half, T20 to ⅓ and so on. The assumption is that the decay trend of the first 15, 20 or 30 dB would continue with a longer measurement. The shortened decay measurements can increase confidence that we are above the room's noise floor, which can be a challenge for full 60 dB decays.

1.6.4.3 EARLY DECAY TIME (EDT)

Early decay time (EDT) measurements use a different starting point (the first reflection) than the RT series (direct sound arrival). As the name implies, this is a statistical analysis of the early reflections. The published value is derived from 15 dB of decay and extrapolated to the 60 dB equivalent. EDT highlights the different perspectives of acousticians and sound engineers. An EDT measurement characterizes a room without direct sound and a free-field speaker measurement characterizes the direct sound without a room.

1.6.4.4 NOISE FLOOR (NOISE CRITERIA)

The noise floor for rooms is generally specified with an "NC" rating. This stands for noise criteria, which relates to its perceived loudness. NC ratings are one of the few acoustical levels you'll see specified without dB in the answer,

Acoustical metrics				
Category	**Measurement**	**Inclusive**	**Results**	**Applications**
RT 60	60 dB of measured decay	Direct sound to 60 dB decay	Seconds, octave bands, 125 Hz to 8 kHz	Optimal times are application dependent
T 10	Initial 10 dB decay extrapolated to the 60 dB decay rate	-5 dB to -15 dB decay		
T 20	Initial 20 dB decay (to 60 dB rate)	-5 dB to -25 dB decay		
T 30	Initial 30 dB decay (to 60 dB rate)	-5 dB to -35 dB decay		
Early decay time (EDT)	10 dB decay from first reflection (60 dB rate)	0 dB to -10 dB decay		Favors early arrivals, so this is one of the more relevant acoustical numbers when sound systems are used.
Clarity (C50)	Ratio of early to late sound energy	Before and after 50 ms	0 dB is equal. + x dB indicates early is greater than late	Used for vocal intelligibility analysis. Positive clarity (in dB) aids intelligibility
Clarity (C80)	Ratio of early to late sound energy	Before and after 80 ms		Used for musical applications. -2 and below are favored for musical support

FIGURE 1.27
Standard metrics for the evaluation of room acoustics

e.g. a quiet room has an NC 20 rating. The NC ratings are derived from octave bands from 63 to 8 kHz and can be integrated to a single value that can be described as "A overweighted" at lower levels, meaning it is even less sensitive to the LF range. The weighting is level dependent (like the equal loudness contours). An NC rating of 25 corresponds to approximately 35 dB A. The primary applications are noise control (heating, ventilation, and air conditioning (HVAC), isolation, etc.).

1.6.5 Speaker acoustical metrics

There are also well-established parameters to describe the free-field acoustical behavior of loudspeakers (exclusive of the room) (Fig. 1.28). Only the most basic are discussed here as this will be covered in detail in section 2.7.

1.6.5.1 FREQUENCY RESPONSE/RANGE

There are two main categories for frequency response: operating range and tolerance. Range limits are less standardized for speakers than simple electronics, with some manufacturers using 3 dB, 4 dB, 6 dB and even 10 dB down. The most logical limit choice is -6 dB because it's the lowest value that can still combine to a unity crossover.

Tolerance is given as ± × dB within the operational range. Speaker tolerances are often < ±4 dB, but this spec is not standardized and may have unspecified frequency resolution (⅓ octave, octave, etc.).

1.6.5.2 MAXIMUM SPL

Maximum SPL specifications separate speakers by operating level. A single frequency was sufficient for evaluating the maximum level on electronic equipment. But what can 1 kHz tell us about our subwoofer, or any speaker for that matter? A speaker's maximum level needs to be evaluated on a case-by-case (frequency-by-frequency) basis. It is obvious that a single maximum SPL number cannot accurately characterize a speaker, but that doesn't mean that you won't see it on spec sheets.

SPL ratings are standardized around a 1 m distance (even if actually measured at 2 m or 4 m). The standard input signal is maximum-level continuous pink noise. From here it gets sketchy. How much distortion is allowable? Compression? Limiting? No standard. Do we know the speaker will survive another 3 seconds? In the end we get some numbers: maximum continuous, peak, "music," "program" and more. Various spectral weighting curves "A" and "C" or "Z" tailor the SPL specification over frequency.

It is important to understand that a meter reading of 120 dB SPL does not mean a speaker is generating 120 dB SPL at all frequencies. The 120 dB value is the integration of all frequencies (unless otherwise specified) and no conclusion can be made for a particular frequency range. Maximum SPL for a particular frequency range can be measured by limiting the input drive to the desired spectrum, a practice known as banded SPL measurements. The same data cannot be attained by analysis with post-process band limiting. Energy is spread over the full band when a full-range signal is applied to a device. Band-limited measurements show lower maximum levels for a given band if the device is reproducing frequencies outside of the measured band.

1.6.5.3 COVERAGE ANGLE

Coverage angle is the radial spread between the -6 dB points (referenced at an equal distance to the on axis point), normally given for the vertical and horizontal planes. No known speaker can maintain a constant coverage angle over frequency, and yet the determination of "nominal" coverage angle is not standardized. The 1 kHz coverage angle is not applicable as the nominal value for a speaker. Instead we use the average coverage angle over the upper several octaves (devices with a fairly constant coverage angle in the HF range) or the narrowest range (wavelength-proportional coverage angle). This spec helps match the speaker to the coverage target shape and for array building.

1.6.5.4 BEAMWIDTH

Beamwidth is coverage angle over frequency. Coverage angle, as a single number to describe a speaker, can be insufficient for many applications. Beamwidth plots allow us to compare the frequency dependent coverage shape of different speakers (which may have the same "nominal" coverage angle). This spec greatly aids the process of matching speakers to the desired coverage shape and array building.

1.6.5.5 SENSITIVITY (1 WATT/1M)

Sensitivity is the electrical and acoustical conversion factor, given as the speaker's SPL at 1 m with 1 watt drive level. Sensitivity ratings help users pair generic speakers and amplifiers. Industry trends show increasing preference for manufacturer-selected pairings or self-powered speakers, leaving sensitivity concerns to the manufacturers.

Loudspeaker metrics

Category	Measurement	Inclusive	Results	Applications
Range/tolerance	Amplitude vs. frequency	Typically -6 dB points relative to nominal	F_{LOW} to F_{HI} ± x dB	Application dependent. e.g. Fills can have reduced LF range, higher tolerance
Maximum SPL	Full range level at maximum output at 1m	Max level with < 3 dB of limiting (Not standardized)	Seconds, octave bands, 125 Hz to 8 kHz	Provides power scale for designing with the speaker
Coverage angle	Amplitude vs. angle	0 dB to -6 dB in the HF range	x degrees over a nominal range	Provides coverage data for speaker selection (solo and array applications)
Beamwidth	Coverage angle vs. frequency	0 dB to -6 dB in octave or 1/3rd octave bands	Graph of coverage angle vs. frequency	Provides coverage data particularly relevant for array applications
Phase	Phase shift or phase delay over frequency	No industry standard	± x degrees (or ms) over a specified frequency range	Provides performance and compatibility data for speakers
Sensitivity	Speaker efficiency	Calculated on a per driver basis	x dB (1 watt, 1 meter)	Provides info for pairing speakers with amplifiers
Rated power	Power capability of the speaker	Maximum level before fire for each driver	x watts (peak, continuous, burst, long term)	Provides info for pairing speakers with amplifiers
THD	Distortion over level	Calculated on a per driver basis	x % THD at F_X at x dB SPL	Provides info for whether a speaker sounds like s#&%.

FIGURE 1.28
Standard metrics for the evaluation of loudspeakers

1.6.5.6 MAXIMUM POWER RATING (WATTS)

This is the maximum power the speaker can dissipate before being sent to the smoking section, with various iterations such as continuous, music, FTC, maximum and marketing watts. One could assume that an 800 watt speaker could be run at full level with an 800 watt amp, but your mileage may vary (a lot). This spec is intended to help match generic speakers and amps.

1.6.5.7 TOTAL HARMONIC DISTORTION (THD)

Harmonic distortion level in speakers is highly variable over level and frequency, which makes this a much more difficult specification to monitor. Many manufacturers provide little or no data on this, and if you measure their speakers you will know why.

1.6.6 Combined loudspeaker/room acoustical metrics

And finally we have metrics that describe the performance of the speakers in the room. These include frequency response, direct/reverberant ratio, intelligibility and others. These are intended as final measures of installed system/room performance (Fig. 1.29).

1.6.6.1 FREQUENCY RESPONSE

This is the combined frequency response of the speaker (direct sound) and the room (reflected sound). Analyzers differ in their approach to reflected response inclusion. Single-spectrum linear analyzers use a single time record. Reflection inclusion is the same for all frequencies within the specified time period and includes more wavelengths of reflected sound as frequency rises.

Quasi-log FFT analyzers use a series of different length time records that vary with frequency (LF are longest). Reflection inclusion is proportional to wavelength, i.e. the response includes the direct sound and a given number a wavelengths beyond. This is consistent with our hearing system's perceived tonal fusion zone.

1.6.6.2 DIRECT/REVERBERANT RATIO

Direct to reverberant ratio is exactly what the name connotes. Higher ratios correlate to greater clarity and increased signal/noise. Results vary greatly over frequency and therefore are poorly suited to being represented as a single number value. The HF will normally have a much higher ratio than the LF. Published D/R specifications standardize around the 1 kHz band. There are various versions C50, C7 and C35 but all incorporate only this single range. An additional limitation to D/R ratio is that it incorporates very few sound system performance parameters. Distortion and gross mismanagement of the optimization settings are just a few things that could slip under this metric's radar.

1.6.6.3 CRITICAL DISTANCE

Critical distance links a location in the room to direct-to-reverberant ratio. It is the distance from the loudspeaker where the D/R reaches unity. Closer locations are presumed to have more direct (and equivalent reverb) and more distant locations the opposite. Some engineers find value in staking out this position but I can't offer any guidance on how to find it or what to do with it. The reason is that D/R ratio is a number with 1000 faces (20 kHz/20 Hz) and therefore critical distance is in 1000 places. The distance that puts 1 kHz in critical condition leaves 10 kHz with a scratch and 100 Hz with a toe tag. One-size-fits-all numbers can't speak for our full-range systems and one-place-fits-all locations are just a version of the same thing. I have never been able to place a mic in a room and say, "As you can see, we are now at the critical distance."

1.6.6.4 COHERENCE

Coherence is the data quality metric used for system optimization. It is a standard feature of all the dual-channel FFT analysis systems. It is covered in detail (section 12.10) but we will touch on it briefly here. Coherence is able to discern signal/noise ratio vs. frequency (typically 48 point/octave). This allows us to see the degrading effects of reflections, air loss and wind on S/N ratio in high resolution. It also detects distortion, compression and combing between elements on the speaker side. A coherence factor of 50% indicates 0 dB S/N ratio at that location at that frequency.

1.6.6.5 INTELLIGIBILITY

Intelligibility is a complex metric that seeks to correlate system performance to its intelligible speech transmission capability. This is a book of its own (or books since there is so much debate about the various methods). There is widespread agreement on one item: More work needs to be done to find a metric that reliably correlates to intelligibility while incorporating both the room and the sound system. It's a tremendously complex task, but the fairest grade we can give to these systems at this time is "incomplete."

%Alcons (percentage articulation loss of consonants)

D/R ratio and EDT data are combined to create a quality index based on the percentage of lost consonants. Acceptable loss rates are expected to be <10% for general PA applications. Only a ⅓ octave band centered at 2 kHz is included in the data so its results are not valid for James Earl Jones.

Speech Transmission Index (STI)

STI uses a dedicated test signal that mimics speech. Scores are on a scale of 0 to 1. Human speech results from fundamental frequencies that are modulated by our vocal cords and mouth movements. The STI test signal is modulated to mimic this response. The transmission index scores the system by how much of the modulation depth is lost in transit.

Rapid Speech Transmission Index (RASTI)

RASTI is based on the STI principles but packaged as a commercially available test system. RASTI is simpler to operate and interpret than STI but at the cost of reduced spectral representation (two octave bands centered at 500 Hz and 2 kHz). RASTI is economical and has enjoyed widespread usage. Nonetheless it has serious shortcomings for those of us who design systems that span beyond the telephone range. Distortion, equalization, timing errors, compression and limiting can all affect (or not affect) the readings in unexpected ways. For example, limiting reduces the score, but clipping does not. Let's book Spinal Tap and turn it up to 11!

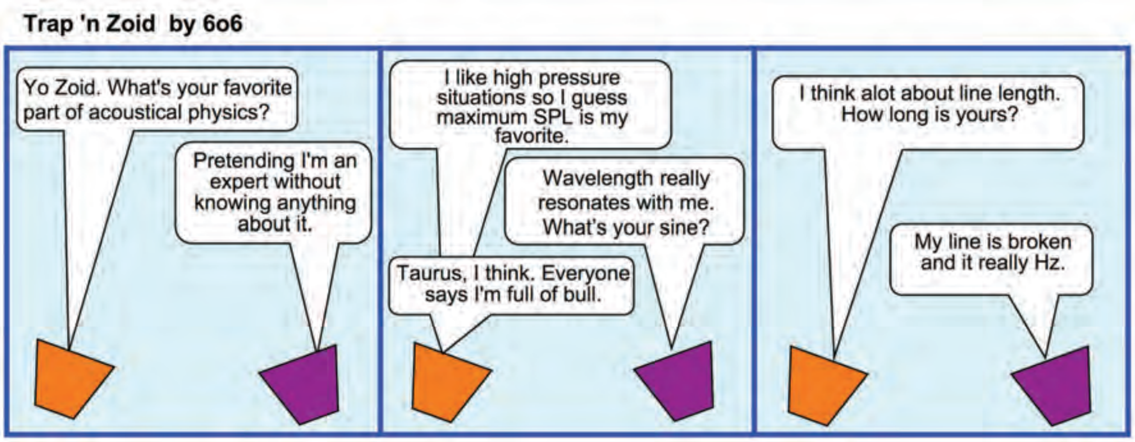
Loudspeaker/room combination metrics

Category	Measurement	Typical	Results	Applications
Maximum SPL	Maximum level for the whole system	Too loud	x dB SPL at a given position, typically FOH	Match power scale to program material. Evaluate need for delays, fills etc.
Frequency response	Spectral response	Many specs say ± 2 dB though unrealistic	Seconds, octave bands, 125 Hz to 8 kHz	Overall curve is artistically determined. Used to evaluate acoustic treatment, speaker aiming, needs for fills etc.
Level uniformity	Relative level over the seating area	Many specs say ± 2 dB though unrealistic	x dB over and under the target	Evaluation of level distribution, need for fills and speaker aiming
Direct/Reverb ratio	Various methods. Most common in the modern era is coherence	Venue dependent	> 50%-80% in favorable situations.	Used to evaluate acoustic treatment, speaker aiming, needs for fills etc.
Intelligibility	STI		Rating system of 1 to 5	Evaluation of speech quality for the combination of room and speakers
% ALCons	Modulation transfer function (loss of modulation depth)	Loss of 10 % or less is considered acceptable for PA applications	x % ALCons	Evaluation of speech quality for the combination of room and speakers

FIGURE 1.29
Standard metrics for the evaluation of loudspeaker/room combination

SII

SII, as the name connotes, is 2nd-generation STI. This is currently under standards development and has the promise of superior correlation to our experience than the previous generations. It's not perfect but has addressed many of the shortcomings of the previous methods. SII has an expanded bandwidth (150 Hz to 8.5 kHz) and higher frequency resolution than previous methods. The effects of reverberation, noise and distortion are also included. Cancel Spinal Tap.



FURTHER READING

http://en.wikipedia.org/wiki/A-weighting.
Jones, R., Speech Intelligibility Papers: www.meyersound.com/support/papers/speech.
http://en.wikipedia.org/wiki/Intelligibility_(communication).
http://en.wikipedia.org/wiki/Speed_of_sound.

CHAPTER 2

Classification

What are sound systems made of? How do we classify the "system" into component parts for discussion? The simplest distillation might be signal sources, processors, transmitters and the connections between them. Now let's slice it a little finer. Sources include signal emitters such as the voice, musical instruments, recorded playback, etc. The human voice will tell you straight from the source that they don't want to be categorized as part of the sound system. They are the source for our sources: microphones and direct interfaces that give us custody of the signal. These in turn become sources for the next stage, the mix console, which in turn becomes the source for the system's signal processing. The signal processing includes the filters, delays, level distribution and more that prepare the signal for transmission. The transmission system consists of the amplifiers and loudspeakers to put the signal into the air toward the listeners. We can slice it finer yet, which is precisely the focus of this chapter, an inventory of the system components.

***class** n. **1**. division according to quality.*
***2**. group of persons or things having some characteristic in common.*
***classify** v.t. arrange in classes; assign to a class.*
Concise Oxford Dictionary

2.1 MICROPHONES

Microphones are acoustic to electronic waveform media converters, i.e. transducers. The various microphone types differ in the intermediate stages of the conversion process. There are two main branches of the microphone family tree: dynamic (moving coil in a static magnetic field) and condenser (moving charged capacitive plate in an electric field). In both cases the diaphragm tracks the waveform's acoustic response but the transduction path to electricity differs. From there the mics differentiate into directional types, diaphragm sizes, materials and physical arrangements.

Most books focus on the primary application of microphones: recording and reinforcing sound sources. They are at the beginning of the signal path, serving as our original waveform sources. This book, by contrast, focuses primarily on the measurement microphones at the end of the signal path serving as surrogate ears for acoustical analysis. Nonetheless, we will briefly cover the microphones in general.

2.1.1 Microphone types

2.1.1.1 DYNAMIC

Dynamic microphones are the mechanical inverse of a standard loudspeaker (Fig. 2.1). The diaphragm has an attached coil, which floats inside a static magnetic field. The acoustic waveform creates a tracking movement in the diaphragm, which moves the coil in the gap between the opposing magnetic fields. The current induced in the coil is a transcription of the acoustic waveform. Voltage fluctuations in the coil are the electronic waveform we carry forward through the signal path. Dynamic mics, like their namesake the dynamo, are electrical power generators (albeit microscopic amounts) and therefore do not require external power.

A variety of factors converge on the dynamic mic to make life difficult in the VHF range: the mechanics of moving a coil on a diaphragm, reactive losses in the wire, the extremely low levels generated by the mic and more. This renders the dynamic mic unsuitable for applications requiring VHF extension such as cymbals and acoustic measurement.

Except for people saying "Check, 1,2" into an SM58 at front of house (FOH), I have never seen anyone use a dynamic mic as the test reference mic for system optimization (and hope I never will). Therefore this will be the end of our discussion regarding dynamic mics that are classified as velocity microphones.

2.1.1.2 CONDENSER

Condenser microphones have a variety of subsets and names including capacitance, electret and more (also Fig. 2.1). The basic scheme is the same: The waveform is transcribed by variable capacitance. The diaphragm must be a conductive metal (or coated with one) because it will act as one of the charged parallel plates of a capacitance circuit. The second plate is fixed in position and charged with a DC voltage. When no acoustic signal is present the DC voltage becomes the ambient-level reference point. When the diaphragm moves, the spacing between the plates changes, resulting in a variable capacitance between them. The variations are a transcription of the waveform and create an audio signal (AC) that can later be separated from the DC polarizing voltage. Condenser mics require an external DC power supply, commonly known as "phantom power" (48 volts DC) to charge the plates.

Condenser mics have higher cost and lower durability than dynamic mics, but the extended HF range and minimal coloration weigh in their favor. Condensers have the mechanical advantage of moving only a lightweight diaphragm, with no coil attached. Second, the phantom power allows for on-board circuitry to minimize reactive losses and maximize signal/noise ratio. For this reason the condenser mic is the preferred choice for applications that require VHF extension such as cymbals and acoustic measurement.

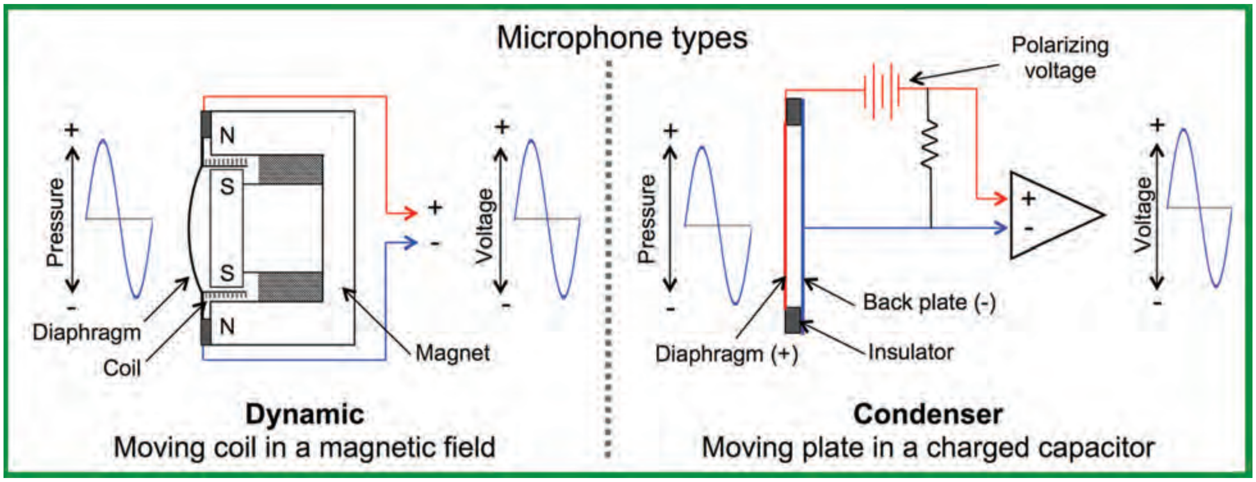


FIGURE 2.1
Microphone types

2.1.2 Microphone coverage patterns

Coverage pattern is an independent variable in microphone construction. In other words, dynamic or condenser mics can be configured in any of the basic patterns described below. The term "coverage pattern" refers to the mic's directional response, i.e. the differences between the on-axis response and those at other angles of incidence. In gross terms, there are mics that are largely indifferent to direction (omnidirectional) and those that have strong orientation in particular directions (Fig. 2.2).

2.1.2.1 OMNIDIRECTIONAL

An omnidirectional microphone is one whose frequency response is indifferent (or nearly so) to its angular orientation to a sound source. The mics are sealed so no sound reaches the rear of the diaphragm. Therefore there is no cancelling counterforce to the diaphragm movement. A pressure change from any direction can move the diaphragm, which is why these are known as "pressure-operated" microphones.

You might think an ideal omni mic would have no difference in its frequency response over angle, but such a thing is not possible or even desirable. A mic can only stay fully omnidirectional if its diaphragm is infinitely small. As diaphragm size rises, the frequency range of omnidirectionality falls. There is another side effect of diaphragm size: noise. As diaphragm size falls, the mic's self-noise rises. It is wise to consider the real-world tradeoff between best directionality and noise floor for a given application.

Omni mics have limited use in sound reinforcement because they have high leakage and low gain before feedback. They are the standard measurement mic for system optimization. All acoustic frequency response traces in this book come from omnidirectional condenser mics. There are more details in Chapter 12.

2.1.2.2 UNIDIRECTIONAL

A unidirectional microphone has substantial response variations over angle. Directional mics have a path for sound to reach both the front and rear of the diaphragm. The output signal is the pressure difference between the front and rear of the diaphragm. These are termed "pressure-gradient" microphones.

Bi-directional microphone (figure 8)

A mic whose diaphragm is equally open on both front and back is termed "bi-directional." Sound sources traveling toward the front or back of the mic move the diaphragm and create an electrical waveform. The frequency response may be similar in front and back but polarity is reversed for signals from the rear. Sources at the side of the mic are rejected because there is no pressure difference between the front and back of the diaphragm. The result is a "figure 8" pattern. This is the most common configuration for ribbon mics.

Cardioid, hypercardioid and supercardioid microphone

Consider the approach to the back side of the omni and bi-directional mics just discussed. The omni is sealed behind the diaphragm and the bi-directional is fully open to both front and back. Both mics cover the area in front and behind. The bi-directional rejects the sides. What would happen if we made a hybrid of those two: a mic that was open in the front and partially open in the rear? The response would be a hybrid that strongly rejects the rear and to a lesser extent the sides. This is the cardioid family of microphones.

Cardioid mics enable a portion of the sound to reach the back of the diaphragm through side and rear ports. We'll examine this in three parts. We first follow a sound wave's pressurized (+) portion arriving from behind. Sound first enters the port and later moves around to the microphone front. Two paths converge: Positive pressure from behind pushes the diaphragm outward, while positive pressure from the front pushes it inward. It's a perfect polarity reversal cancellation, right? Not quite, because perfect cancellation requires unity gain and exactly 180° phase offset. We have neither. The rear path to the diaphragm is shorter than the front. The clever remedy is adding a tiny labyrinth in the mic housing between the port and the diaphragm rear. This delays the rear arrival until it times out perfectly to maximize cancellation. However, we lose a few dB by taking the port path compared with the front.