

Principles of
Digital Audio

S i x t h E d i t i o n



Ken C. Pohlmann

Principles of Digital Audio

Ken C. Pohlmann

Sixth Edition



New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

Copyright © 2011 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-166347-2

MHID: 0-07-166347-9

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-166346-5,
MHID: 0-07-166346-0.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGrawHill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

*for
Leslie Hope
and her
headwind tenacity
and tailwind exuberance*

About the Author

Ken C. Pohlmann is a professor emeritus at the University of Miami in Coral Gables, Florida. He was the director of the Music Engineering program in the university's Frost School of Music from 1983 to 2007. He initiated new undergraduate and graduate courses in digital audio, advanced digital audio, Internet audio, acoustics and psychoacoustics, and studio production. In 1986, he founded the first master's degree program in Music Engineering in the United States. Mr. Pohlmann holds Bachelor of Science and Master of Science degrees in Electrical Engineering from the University of Illinois in Urbana-Champaign.

Mr. Pohlmann is the author of *Principles of Digital Audio* (McGraw-Hill); this book has appeared in six editions and has been translated into Dutch, Spanish, and Chinese. He is also author of *The Compact Disc Handbook* (A-R Editions); this book has appeared in two editions and has been translated into German. He is co-author of the most recent (fifth) edition of *The Master Handbook of Acoustics* (McGraw-Hill). He is also co-author of *Writing for New Media* (John Wiley & Sons), and editor and co-author of *Advanced Digital Audio* (Howard W. Sams). Since 1982, he has written over 2500 articles for publications including *Audio*, *Broadcast Engineering*, *dB*, *Car Stereo Review*, *CD Review*, *Edmunds.com*, *Electronics Australia*, *Guitar Player*, *Handbook for Sound Engineers*, *IEEE Spectrum*, *Journal of the Audio Engineering Society*, *Laserdisk Professional*, *McGraw-Hill Encyclopedia of Science and Technology*, *Mix*, *Mobile Entertainment*, *National Association of Broadcasters Handbook*, *NARAS Journal*, *PC*, *Road & Track Road Gear*, *Sound & Vision*, *Scientific American*, *Spektrum der Wissenschaft*, *Stereo Review*, *Video*, and *World Book Encyclopedia*. He is a contributing technical editor and columnist for *Sound & Vision*.

Mr. Pohlmann has worked extensively in the research, development, and testing of new audio technology. He serves as a consultant in the theory and application of digital audio systems and the development of sound systems for automobile manufacturers. Some of his consulting clients include: Alpine Electronics, Analog Devices, Apple Computer, Bertlesmann Music Group, Blockbuster Entertainment, BMW, Canadian Broadcasting Corporation, Cirrus Logic, DaimlerChrysler, Eclipse, Ford, Fujitsu Ten, Harman International, Hughes Electronics, Hyundai, IBM, Kia, Lexus, Lucent Technologies, Microsoft, Mitsubishi Electronics, Motorola, Nippon Columbia, Onkyo America, Philips, RealNetworks, Recording Industry Association of America, Samsung, Sensomatic, Sonopress, Sony, TDK, Time Warner, Toyota, and United Technologies.

Mr. Pohlmann has served as a consultant or expert witness for patent infringement and other issues with law firms such as Arnold & Porter; Baker & McKenzie; Barnes & Thornburg; Christie Parker & Hale; Cushman Darby & Cushman; Darby & Darby; Dewey Ballantine; Firmstone & Feil; Fish & Neave; Fish & Richardson; Greenberg, Glusker, Fields, Claman, Machtinger & Kinsella; Howrey; Hunton & Williams; Jenner & Blocker; Jones Day; Kenyon & Kenyon; Kirkland & Ellis; McDermott Will & Emery; McDonnell Boehnen Hulbert & Berghoff; Paul, Weiss, Rifkind, Wharton & Garrison; Young & Thompson; and the U.S. Justice Department Anti-Trust Division.

Mr. Pohlmann co-founded Microcomputer Arts, Inc. (1980), International Business Information Systems Inc. (1982), and U.S. Digital Disc Corporation (1985). He chaired the Audio Engineering Society's (AES) International Conference on Digital Audio in Toronto in 1989 and co-chaired the Society's International Conference on Internet Audio in Seattle in 1997. He was presented two AES Board of Governors Awards (1989 and 1998) and was named an AES Fellow in 1990 for his work as an educator and author. He was elected to the AES Board of Governors in 1991. He was presented the University of Miami Philip Frost Award for Excellence in Teaching and Scholarship in 1992. He served as AES convention papers chairman in 1984 and papers co-chairman in 1993. He was elected as the Vice President of the AES Eastern U.S. and Canada Region in 1993. He served as a Non-Board Member of the National Public Radio Distribution/Interconnection Committee (2000–2003). He served on the Board of Directors of the New World Symphony (2000–2005). Mr. Pohlmann joined the Advisory Board of SRS Labs in 2009 as the charter member.

Contents

Preface	xv
1 Sound and Numbers	1
Physics of Sound	1
Sound Pressure Level	3
Harmonics	4
Digital Basics	5
Number Systems	5
Binary Number System	6
Binary Codes	8
Weighted Binary Codes	9
Unweighted Binary Codes	10
Two's Complement	10
Boolean Algebra	13
Analog versus Digital	16
2 Fundamentals of Digital Audio	19
Discrete Time Sampling	19
The Sampling Theorem	20
Nyquist Frequency	21
Aliasing	25
Quantization	28
Signal-to-Error Ratio	29
Quantization Error	33
Other Architectures	35
Dither	36
Types of Dither	39
Summary	44
Postscript	45
3 Digital Audio Recording	47
Pulse-Code Modulation	47
Dither Generator	50
Input Lowpass Filter	50
Sample-and-Hold Circuit	53
Analog-to-Digital Converter	56
Successive Approximation A/D Converter	60
Oversampling A/D Converter	62
Record Processing	64

Channel Codes	65
Simple Codes	68
Group Codes	70
Code Applications	75
4 Digital Audio Reproduction	77
Reproduction Processing	77
Digital-to-Analog Converter	79
Weighted-Resistor Digital-to-Analog Converter	81
R-2R Ladder Digital-to-Analog Converter	83
Zero-Cross Distortion	84
High-Bit D/A Conversion	85
Output Sample-and-Hold Circuit	86
Output Lowpass Filter	89
Impulse Response	90
Digital Filters	94
FIR Oversampling Filter	97
Noise Shaping	100
Output Processing	102
Alternate Coding Architectures	102
Floating-Point Systems	103
Block Floating-Point Systems	105
Nonuniform Companding Systems	106
m-Law and A-Law Companding	106
Differential PCM Systems	107
Predictive Differential Coding	108
Delta Modulation	108
Adaptive Delta Modulation	110
Companded Predictive Delta Modulation	111
Adaptive Differential Pulse-Code Modulation	112
Timebase Correction	113
Jitter	114
Eye Pattern	115
Interface Jitter and Sampling Jitter	116
Jitter in Mechanical Storage Media	117
Jitter in Data Transmission	119
Jitter in Converters	120
5 Error Correction	125
Sources of Errors	126
Quantifying Errors	128
Objectives of Error Correction	128
Error Detection	129
Single-Bit Parity	130
ISBN	132
Cyclic Redundancy Check Code	133

Error-Correction Codes	138
Block Codes	139
Hamming Codes	142
Convolutional Codes	144
Interleaving	146
Cross-Interleaving	148
Reed–Solomon Codes	149
Cross-Interleave Reed–Solomon Code (CIRC)	154
CIRC Performance Criteria	157
Product Codes	158
Error Concealment	162
Interpolation	162
Muting	162
Duplication	164
6 Optical Disc Media	165
Optical Phenomena	165
Diffraction	168
Resolution of Optical Systems	170
Polarization	171
Design of Optical Media	174
Nonerasable Optical Media	175
Read-Only Optical Storage	175
Write-Once Optical Recording	177
Erasable Optical Media	178
Magneto-Optical Recording	179
Phase-Change Optical Recording	181
Dye-Polymer Erasable Optical Recording	182
Digital Audio for Theatrical Film	183
7 Compact Disc	187
Development	187
Overview	188
Disc Design	190
Disc Optical Specification	190
Data Encoding	193
Player Optical Design	197
Optical Pickup	197
Autofocus Design	200
Autotracking Design	201
One-Beam Pickup	202
Pickup Control	204
Player Electrical Design	205
EFM Demodulation	205
Error Detection and Correction	207
Output Processing	208
Subcode	209

Disc Manufacturing	213
Premastering	213
Disc Mastering	214
Electroforming	216
Disc Replication	216
Alternative CD Formats	218
CD-ROM	219
CD-R	224
CD-RW	231
CD-MO	233
CD-i	233
Photo CD	233
CD + G and CD + MIDI	234
CD-3	234
Video CD	235
Super Audio CD	235
Disc Design	236
DSD Modulation	238
DST Lossless Coding	242
Player Design	242
8 DVD	245
Development and Overview	245
Disc Design	248
Disc Optical Specification	248
Disc Manufacturing and Playback	251
Optical Playback	253
Data Coding	256
Reed–Solomon Product Code	257
EFMPlus Modulation	258
Universal Disc Format (UDF) Bridge	259
DVD-Video	260
DVD-Video Video Coding	262
DVD-Video Audio Coding	264
DVD-Video Playback Features	266
DVD-Video Authoring	268
DVD-Video Developer’s Summary	270
DVD-Audio	274
DVD-Audio Coding and Channel Options	275
DVD-Audio Disc Contents	280
DVD-Audio Developer’s Summary	283
Alternative DVD Formats	287
DVD-ROM	289
DVD-R and DVD + R	289
DVD-RW and DVD + RW	290
DVD-RAM	292

DVD Content Protection	293
DVD-Video Copy Protection	293
DVD-Audio Copy Protection	295
Content Protection for Recordable Media	296
Secure Digital Transmission	297
DVD Watermarking	297
HD DVD	297
9 Blu-ray	299
Development and Overview	299
Disc Capacity	301
BD-ROM Player Profiles	302
Disc Design	303
Disc Optical Specification	303
Optical Pickup Design	306
Disc Manufacturing	308
BD-ROM Specifications	310
Audio Codecs	313
Video Codecs	316
Modulation and Error Correction	318
Audio-Video Stream Format and Directory	321
Blu-ray UDF File System	324
HDMV and BD-J Application Programming Modes	325
Blu-ray 3D	326
Region Playback Code	326
Content Protection	327
Blu-ray Recordable Formats	330
AVCHD Format	332
10 Low Bit-Rate Coding: Theory and Evaluation	335
Perceptual Coding	335
Psychoacoustics	336
Physiology of the Human Ear and Critical Bands	339
Threshold of Hearing and Masking	344
Temporal Masking	347
Psychoacoustic Models	349
Spreading Function	351
Tonality	352
Rationale for Perceptual Coding	353
Perceptual Coding in Time and Frequency	355
Subband Coding	356
Transform Coding	360
Filter Banks	364
Quadrature Mirror Filters	364
Hybrid Filters	365
Polyphase Filters	366
MDCT	367

Multichannel Coding	369
Tandem Codecs	370
Spectral Band Replication	371
Perceptual Coding Performance Evaluation	373
Critical Listening	376
Listening Test Methodologies and Standards	379
Listening Test Statistical Evaluation	383
Lossless Data Compression	384
Entropy Coding	385
Audio Data Compression	386
11 Low Bit-Rate Coding: Codec Design	393
Early Codecs	393
MPEG-1 Audio Standard	394
MPEG Bitstream Format	396
MPEG-1 Layer I	398
Example of MPEG-1 Layer I Implementation	399
MPEG-1 Layer II	402
MPEG-1 Layer III (MP3)	404
MP3 Bit Allocation and Huffman Coding	407
MP3 Stereo Coding	409
MP3 Decoder Optimization	409
MPEG-1 Psychoacoustic Model 1	410
MPEG-1 Psychoacoustic Model 2	414
MPEG-2 Audio Standard	417
MPEG-2 AAC	419
AAC Main Profile	420
AAC Allocation Loops	422
AAC Temporal Noise Shaping	423
AAC Techniques and Performance	425
ATRAC Codec	425
Perceptual Audio Coding (PAC) Codec	430
AC-3 (Dolby Digital) Codec	432
AC-3 Overview	432
AC-3 Theory of Operation	434
AC-3 Exponent Strategies and Bit Allocation	435
AC-3 Multichannel Coding	438
AC-3 Bitstream and Decoder	441
AC-3 Applications and Extensions	443
DTS Codec	444
Meridian Lossless Packing	446
12 Speech Coding for Transmission	451
Speech Coding Criteria and Overview	451
Waveform Coding and Source Coding	453
Human Speech	455
Source-Filter Model	457
Channel, Formant, and Sinusoidal Codecs	458

Predictive Speech Coding	461
Linear Predictive Coding	464
Code Excited Linear Prediction	467
CELP Encoder and Decoder	468
CELP Codebooks	470
Vector Quantization	471
Examples of CELP Codecs	472
Scalable Speech Coding	472
G.729.1 and MPEG-4 Scalable Codecs	474
Bandwidth Extension	476
Echo Cancellation	479
Voice Activity Detection	480
Variable Bit Rate	480
Speech Recognition	481
Speex Codec	482
Quantifying Performance of Speech Codecs	482
Speech Coding Standards	483
13 Audio Interconnection	485
Audio Interfaces	485
SDIF-2 Interconnection	486
AES3 (AES/EBU) Professional Interface	488
AES3 Frame Structure	489
AES3 Channel Status Block	491
AES3 Implementation	494
AES10 (MADI) Multichannel Interface	495
S/PDIF Consumer Interconnection	497
Serial Copy Management System	500
High-Definition Multimedia Interface (HDMI) and DisplayPort	501
Musical Instrument Digital Interface (MIDI)	501
AES11 Digital Audio Reference Signal	503
AES18 User Data Channels	506
AES24 Control of Audio Devices	506
Sample Rate Converters	507
Fiber-Optic Cable Interconnection	509
Fiber-Optic Cable	509
Connection and Installation	513
Design Example	514
14 Personal Computer Audio	517
PC Buses and Interfaces	517
IEEE 1394 (FireWire)	518
Digital Transmission Content Protection (DTCP)	521
Universal Serial Bus (USB)	522
Sound Card and Motherboard Audio	525
Music Synthesis	526
Surround Sound Processing	527
Audio Codec '97 (AC '97)	528

High Definition Audio (HD Audio)	529
Windows DirectX API	531
MMX	532
File Formats	532
WAV and BWF	533
MP3, AIFF, QuickTime, and Other File Formats	535
Open Media Framework Interchange (OMFI)	537
Advanced Authoring Format (AAF)	538
Material eXchange Format (MXF)	539
AES31	540
Digital Audio Extraction	540
Flash Memory	543
Hard-Disk Drives	544
Magnetic Recording	544
Hard-Disk Design	545
Digital Audio Workstations	547
Audio Software Applications	548
Professional Applications	550
Audio for Video Workstations	551
15 Telecommunications and Internet Audio	553
Telephone Services	553
ISDN	555
Asymmetric Digital Subscriber Line (ADSL)	557
Cellular Telecommunications	558
Networks and File Transfers	558
Ethernet	561
Asynchronous Transfer Mode (ATM)	561
Bluetooth	562
IEEE 802.11 Wireless LAN (Wi-Fi)	566
MediaNet	567
Internet Audio	567
Voice over Internet Protocol (VoIP)	570
Digital Rights Management	571
Audio Encryption	573
Audio Watermarking	574
Audio Fingerprinting	576
Streaming Audio	578
G2 Music Codec for Streaming	579
Audio Webcasting	581
MPEG-4 Audio Standard	582
MPEG-4 Interactivity	584
MPEG-4 Audio Coding	585
MPEG-4 Versions	589
MPEG-4 Coding Tools	592
MPEG-7 Standard	594

16	Digital Radio and Television Broadcasting	597
	Satellite Communication	597
	Direct Broadcast Satellites	601
	Digital Audio Radio	601
	Digital Audio Transmission	602
	Spectral Space	604
	Data Reduction	605
	Technical Considerations	606
	Eureka 147 Wideband Digital Radio	609
	In-Band Digital Radio	613
	HD Radio	616
	HD Radio FM-IBOC	616
	HD Radio AM-IBOC	622
	Direct Satellite Radio	626
	Sirius XM Radio	626
	Digital Television (DTV)	628
	DTV and ATSC Overview	628
	Video Data Reduction	629
	MPEG-1 and MPEG-2 Video Coding	631
	MPEG-1 Video Standard	637
	MPEG-2 Video Standard	638
	ATSC Digital Television	639
	ATSC Display Formats and Specification	640
	DTV Implementation	644
17	Digital Signal Processing	647
	Fundamentals of Digital Signal Processing	647
	DSP Applications	648
	Discrete Systems	649
	Linearity and Time-Invariance	649
	Impulse Response and Convolution	649
	Complex Numbers	652
	Mathematical Transforms	654
	Unit Circle and Region of Convergence	658
	Poles and Zeros	659
	DSP Elements	661
	Digital Filters	662
	FIR Filters	663
	IIR Filters	667
	Filter Applications	669
	Sources of Errors and Digital Dither	672
	DSP Integrated Circuits	674
	Processor Architecture	674
	Fixed Point and Floating Point	676
	DSP Programming	678
	Filter Programming	679

Texas Instruments Code	680
Motorola Code	681
Analog Devices SHARC Code	684
Specialized DSP Applications	685
Digital Delay Effects	685
Digital Reverberation	687
Digital Mixing Consoles	688
Loudspeaker Correction	691
Noise Removal	691
18 Sigma-Delta Conversion and Noise Shaping	695
Sigma-Delta Conversion	695
Delta Modulation	697
Sigma-Delta Modulation	699
Analysis of a First-Order Sigma-Delta Modulator	700
Higher-Order Noise Shaping	702
Idle Tones and Limit Cycles	704
One-Bit D/A Conversion with Second-Order Noise Shaping	704
Multi-Bit D/A Conversion with Third-Order Noise Shaping	708
Multi-Bit D/A Conversion with Quasi Fourth-Order Noise Shaping	711
Sigma-Delta A/D Conversion	713
Digital Filtering and Decimation	716
Sigma-Delta A/D Converter Chip	720
Sigma-Delta D/A Converter Chip	722
Sigma-Delta A/D–D/A Converter Chip	723
Noise Shaping of Nonoversampling Quantization Error	724
Psychoacoustically Optimized Noise Shaping	726
Buried Data Technique	730
Appendix The Sampling Theorem	733
Bibliography	737
Index	783

Preface

It seems inconceivable, but the editions of this book have arrived at number six. Clearly, decades ago, I had no idea what I was getting myself into. In 1984, at an Audio Engineering Society convention in New York, I presented a lecture on digital audio theory, using material from a course I had recently inaugurated at the University of Miami. After the lecture, I was thronged by engineers seeking more information on this wondrous new topic. An astute acquisitions editor noted the response and, in true Mephistophelean style, persuaded me to author a book on digital audio. After all, he explained, it would be simple to put everything down on paper. I naively accepted his bargain, and *Principles of Digital Audio* was published in 1985. The book's publication fortuitously coincided with a tremendous surge in the development of digital audio technology.

Digital audio's commercial popularity was matched only by the demand for explanations of its workings. Audio engineers, and students wishing to join their ranks, assimilated every scrap of information they could find, further propelling development of the science. The first edition gave way to the second, which stepped aside for the third, which submitted to the fourth, which yielded to the fifth. Despite occasional half-hearted promises that I would never write another edition, I could not ignore the fact that digital audio technology continued to expand, and there were increasing numbers of people eager to learn about it. Clearly, new editions became inevitable. As I slavishly authored one edition after another, I had only the royalty checks to console me. With the arrival of this sixth edition, I suppose I should start planning for the seventh.

Readers familiar with earlier editions will see the same patterns in the renovations in this latest installment. The essential nature of topics such as discrete time sampling has not changed, but our appreciation of them has. Similarly, the relative importance of topics continually changes; advances in technology diminish the magnitude of some issues, while magnifying others. Moreover, and even more significantly, in the years since the last edition digital audio theory has evolved and entirely new applications have been developed.

Examination of the contents of this sixth edition will show both losses and gains. A number of topics have been eliminated to conserve the page count; most notably and not surprisingly, the magnetic tape, DAT, and MiniDisc chapters have been dropped. In their place, numerous fresh topics have been introduced. For example, readers will find two new chapters on Blu-ray disc and speech coding, and the perceptual audio chapter (renamed "Low Bit-Rate Coding") has been split into two chapters to allow more thorough discussions of this important topic. Also, the bibliography has been updated as needed.

As in past editions, a user-friendly approach has been retained, but the greater depth given to some material has somewhat increased the reading sophistication required. Of course, readers may pick and choose according to their level and need. Also, the wider scope of topics should satisfy a broader range of professional practitioners and students. One thing has not changed. This book is neither a compendium of every possible fact nor an advanced treatise. It is an introductory text that attempts to provide lucid explanations, and to strike that all-important balance between mere information and understanding. In other words, this is a learning tool, sharpened by years of refinement.

A final note: The material in this book stems from the work of the many pioneers and leaders in the field of digital audio technology. We owe a tremendous debt to those wonderfully creative and hard-working engineers who breathe life into digital audio. Clearly, their vision of the potential of this science is transforming both our industry and our society.

Ken C. Pohlmann
Durango, Colorado

CHAPTER 1

Sound and Numbers

Digital audio is a highly sophisticated technology. It pushes the envelope of many diverse engineering and manufacturing disciplines. Although the underlying concepts were well understood in the 1920s, commercialization of digital audio did not begin until the 1970s because theory had to wait 50 years for technology to catch up. The complexity of digital audio is all the more reason to start with the basics. This chapter begins our exploration of ways to numerically encode the information contained in an audio event.

Physics of Sound

It would be a mistake for a study of digital audio to ignore the acoustic phenomena for which the technology has been designed. Music is an acoustic event. Whether it radiates from musical instruments or is directly created by electrical signals, all music ultimately finds its way into the air, where it becomes a matter of sound and hearing. It is therefore appropriate to briefly review the nature of sound.

Acoustics is the study of sound and is concerned with the generation, transmission, and reception of sound waves. The circumstances for those three phenomena are created when energy causes a disturbance in a medium. For example, when a kettledrum is struck, its drumhead disturbs the surrounding air (the medium). The outcome of that disturbance is the sound of a kettledrum. The mechanism seems fairly simple: the drumhead is activated and it vibrates back and forth. When the drumhead pushes forward, air molecules in front of it are compressed. When it pulls back, that area is rarefied. The disturbance consists of regions of pressure above and below the equilibrium atmospheric pressure. Nodes define areas of minimum displacement, and antinodes are areas of maximum (positive or negative) displacement.

Sound is propagated by air molecules through successive displacements that correspond to the original disturbance. In other words, air molecules colliding one against the next propagate the energy disturbance away from the source. Sound transmission thus consists of local disturbances propagating from one region to another. The local displacement of air molecules occurs in the direction in which the disturbance is traveling; thus, sound undergoes a longitudinal form of transmission. A receptor (like a microphone diaphragm) placed in the sound field similarly moves according to the pressure acting on it, completing the chain of events.

We can access an acoustical system with transducers, devices able to change energy from one form to another. These serve as sound generators and receivers. For example, a kettledrum changes the mechanical energy contributed by a mallet to acoustical energy. A microphone responds to the acoustical energy by producing electrical energy. A loudspeaker reverses that process to again create acoustical energy from electrical energy.

The pressure changes of sound vibrations can be produced either periodically or aperiodically. A violin moves the air back and forth periodically at a fixed rate. (In practice, things like vibrato make it a quasi-periodic vibration.) However, a cymbal crash has no fixed period; it is aperiodic. One sequence of a periodic vibration, from pressure rarefaction to compression and back again, determines one cycle. The number of vibration cycles that pass a given point each second is the frequency of the sound wave, measured in Hertz (Hz). A violin playing a concert A pitch, for example, generates a waveform that repeats 440 times per second; its frequency is 440 Hz. Alternatively, the reciprocal of frequency, the time it takes for one cycle to occur, is called the period. Frequencies in nature can range from very low, such as changes in barometric pressure around 10^{-5} Hz, to very high, such as cosmic rays at 10^{22} Hz. Sound is loosely described to be that narrow, low-frequency band from 20 Hz to 20 kHz—roughly the range of human hearing. Audio devices are generally designed to respond to frequencies in that general range. However, digital audio devices can be designed to accommodate audio frequencies much higher than that.

Wavelength is the distance sound travels through one complete cycle of pressure change and is the physical measurement of the length of one cycle. Because the velocity of sound is relatively constant—about 1130 ft/s (feet per second)—we can calculate the wavelength of a sound wave by dividing the velocity of sound by its frequency. Quick calculations demonstrate the enormity of the differences in the wavelength of sounds. For example, a 20-kHz wavelength is about 0.7 inch long, and a 20-Hz wavelength is about 56 feet long. Most transducers (including our ears) cannot linearly receive or produce that range of wavelengths. Their frequency response is not flat, and the frequency range is limited. The range between the lowest and the highest frequencies a system can accommodate defines a system's bandwidth. If two waveforms are coincident in time with their positive and negative variations together, they are in phase. When the variations exactly oppose one another, the waveforms are out of phase. Any relative time difference between waveforms is called a phase shift. If two waveforms are relatively phase shifted and combined, a new waveform results from constructive and destructive interference.

Sound will undergo diffraction, in which it bends through openings or around obstacles. Diffraction is relative to wavelength; longer wavelengths diffract more apparently than shorter ones. Thus, high frequencies are considered to be more directional in nature. Try this experiment: hold a magazine in front of a loudspeaker—higher frequencies (short wavelengths) will be blocked by the barrier, while lower frequencies (longer wavelengths) will go around it.

Sound also can refract, in which it bends because its velocity changes. For example, sound can refract because of temperature changes, bending away from warmer temperatures and toward cooler ones. Specifically, velocity of sound in air increases by about 1.1 ft/s with each increase of 1°F. Another effect of temperature on the velocity of sound is well known to every woodwind player. Because of the change in the speed of sound, the instrument must be warmed up before it plays in tune (the difference is about half a semitone).

The speed of sound in air is relatively slow—about 740 mph (miles per hour). The time it takes for a sound to travel from a source to a receptor can be calculated by dividing the distance by the speed of sound. For example, it would take a sound about one-sixth of a second to travel 200 feet in air. The speed of sound is proportional to elasticity of the medium and inversely proportional to its density. For example, steel is 1,230,000

times more elastic than air thus the speed of sound in steel is 14 times greater than the speed in air, even though the density of steel is 6000 times greater than air. Sound is absorbed as it travels. The mere passage of sound through air acts to attenuate the sound energy. High frequencies are more prominently attenuated in air; a nearby lightning strike is heard as a sharp clap of sound, and one faraway is heard as a low rumble, because of high-frequency attenuation. Humidity affects air attenuation—specifically, wet air absorbs sound better than dry air. Interestingly, moist air is less dense than dry air (water molecules weigh less than the nitrogen and oxygen they replace) causing the speed of sound to increase.

Sound Pressure Level

Amplitude describes the sound pressure displacement above and below the equilibrium atmospheric level. In absolute terms, sound pressure is very small; if atmospheric pressure is 14.7 psi (pounds per square inch), a loud sound might cause a deviation from 14.699 to 14.701 psi. However, the range from the softest to the loudest sound, which determines the dynamic range, is quite large. In fact, human ears (and hence audio systems) have a dynamic range spanning a factor of millions. Because of the large range, a logarithmic ratio is used to measure a sound pressure level (SPL). The decibel (dB) uses base 10 logarithmic units to achieve this. A base 10 logarithm is the power to which 10 must be raised to equal the value. For example, an unwieldy number such as 100,000,000 yields a tidy logarithm of 8 because $10^8 = 100,000,000$. Specifically, the decibel is defined to be 10 times the logarithm of a power ratio:

$$\text{Intensity level} = 10 \log \left(\frac{P_1}{P_2} \right) \text{ dB}$$

where P_1 and P_2 are values of acoustical or electrical power.

If the denominator of the ratio is set to a reference value, standard measurements can be made. In acoustic measurements, an intensity level (IL) can be measured in decibels by setting the reference intensity to the threshold of hearing, which is 10^{-12} W/m^2 (watts per square meter). Thus the intensity level of a loud rock band producing sound power of 10 W/m^2 can be calculated as:

$$\begin{aligned} \text{Intensity level} &= 10 \log \left(\frac{P_1}{P_2} \right) \text{ dB} \\ &= 10 \log \left(\frac{10^1}{10^{-12}} \right) \\ &= 130 \text{ dB SPL} \end{aligned}$$

When ratios of currents, voltages, or sound pressures are used (quantities whose square is proportional to power), the above decibel formula must be multiplied by 2.

The zero reference level for an acoustic sound pressure level measurement is a pressure of 0.0002 dyne/cm^2 . This level corresponds to the threshold of hearing, the lowest SPL humans can perceive, which is nominally equal to 0 dB SPL. The threshold of feeling, the loudest level before discomfort begins, is 120 dB SPL. Sound pressure levels can be rated on a scale in terms of SPL. A quiet home might have an SPL of 35 dB, a busy

street might be 70 dB SPL, and the sound of a jet engine in close proximity might exceed 150 dB SPL. An orchestra's pianissimo might be 30 dB SPL, but a fortissimo might be 110 dB SPL. Thus its dynamic range is 80 dB.

The logarithmic nature of these decibels should be considered. They are not commonly recognizable, because they are not linear measurements. Two motorcycle engines, each producing an intensity level of 80 dB, would not yield a combined IL of 160 dB. Rather, the logarithmic result would be a 3-dB increase, yielding a combined IL of 83 dB. In linear units, those two motorcycles each producing sound intensities of 0.0001 W/m^2 would combine to produce 0.0002 W/m^2 .

Harmonics

The simplest form of periodic motion is the sine wave; it is manifested by the simplest oscillators, such as pendulums and tuning forks. The sine wave is unique because it exists only as a fundamental frequency. All other periodic waveforms are complex and comprise a fundamental frequency and a series of other frequencies at multiples of the fundamental frequency. Aperiodic complex waveforms, such as the sound of motorcycle engines, do not exhibit this relationship. Many musical instruments are examples of the special case in which the harmonics are related to the fundamental frequency through simple multiples. For example, a complex pitched waveform with a 150-Hz fundamental frequency will have overtones at 300, 450, 600, 750 Hz, and so on.

Overtones extend through the upper reaches of human hearing. The relative amplitudes and phase relationships of those overtones account for the timbre of the waveform. For example, a cello and trumpet can both play a note with the same fundamental pitch; however, their timbres are quite different because of their differing harmonic series. When a cellist plays the note D4 as a natural harmonic, the open D string is bowed, which normally produces a note of pitch D3, and the string is touched at its midpoint. The pitch is raised by an octave because the player has damped out all the odd-numbered harmonics, including the fundamental frequency. The pitch changes; because the harmonic structure changes, the timbre changes as well. Harmonic structure explains why the ear has limited ability to distinguish timbre of high-frequency sounds. The first overtone of a 10-kHz periodic tone is at 20 kHz; most people have trouble perceiving that overtone, let alone others even higher in frequency. Still, to record a complex waveform properly, both its fundamental and harmonic structure must be preserved, at least up to the limit of hearing.

The harmonic nature of periodic waveforms is summarized by the Fourier theorem. It states that all complex periodic waveforms are composed of a harmonic series of sine waves; complex waveforms can be synthesized by summing sine waves. Furthermore, a complex waveform can be decomposed into its sine-wave content to analyze the nature of the complex waveform. A mathematical transform can be applied to a waveform represented in time to convert it to a representation in frequency. For example, a square wave would be transformed into its fundamental sine wave and higher order odd harmonics. An inverse transform reverses the process. Likewise the information in any audio signal can thus be represented in either the time domain or the frequency domain. Some digital audio systems (such as the Compact Disc) code the audio signal as time-based samples. Other systems (such as MP3 players) code the audio signal as frequency coefficients.

Given the evident complexity of acoustical signals, it would be naive to believe that analog or digital audio technologies are sufficiently advanced to fully capture the

complete listening experience. To complicate matters, the precise limits of human perception are not known. One thing is certain: at best, even with the most sophisticated technology, what we hear reproduced through an audio system is an approximation of the actual sound.

Digital Basics

Acoustics and analog audio technology are mainly concerned with continuous mathematical functions, but digital audio is a study of discrete values. Specifically, a waveform's amplitude can be represented as a series of numbers. That is an important first principle, because numbers allow us to manage audio information very efficiently. Using digital techniques, the capability to process information is greatly enhanced. The design nature of audio recording, signal processing, and reproducing hardware has followed the advancement of digital technology; the introduction of software programming into the practical audio environment has been revolutionary. Thus, digital audio is primarily a numerical technology. To understand it properly, let's begin with a review of number systems.

The basic problem confronting any digital audio system is the representation of audio information in numerical form. Although many possibilities present themselves, the logical choice is the binary number system. This base 2 representation is ideally suited for storing and processing numerical information. Fundamental arithmetic operations are facilitated, as are logic operations.

Number Systems

It all begins with numbers. With digital audio, we deal with information and numbers, as opposed to an analog representation. Numbers offer a fabulous way to code, process, and decode information. In digital audio, numbers entirely represent audio information. We usually think of numbers as symbols. The symbology is advantageous because the numerical symbols are highly versatile; their meaning can vary according to the way we use them.

For example, consider my classic 1962 BMW R50/2 motorcycle, 500 cubic centimeters, registered as 129907, and shown in Fig. 1.1. Several numbers describe this machine; not so obvious is the important context of each. R50/2 represents the motorcycle's model number, 1962 is the year of manufacture, and 500 represents the quantity of cubic centimeters of engine displacement. The license number 129907 represents coded information that allows my speeding tickets to be properly credited to my account. These

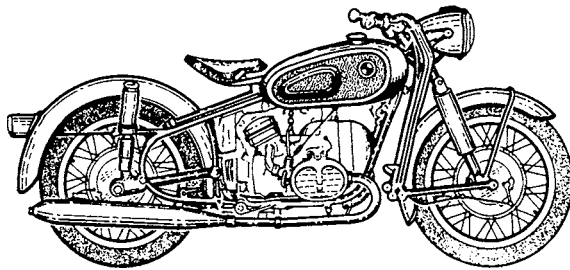


FIGURE 1.1 The author's classic 1962 BMW R50/2 motorcycle.

various numbers are useful only by virtue of their arbitrarily assigned contexts. If that context is confused, then information encoded by the numbers goes awry. I could end up with a motorcycle with license number 1962, manufactured in the year 500, with an engine displacement of 129907 cubic centimeters.

Similarly, the numerical operations performed on numbers are matters of interpretation. The tally of my moving violations determines when my license will be suspended, but the sum of my license plate numerals is less problematic. Numbers, if properly defined, provide a good method for storing and processing data. The negative implication is that numbers and their meanings have to be used carefully.

For most people, the most familiar numbers are those of the base 10 system, apparently devised in the ninth century by Hindu astronomers who conceived of the 0 numeral to represent nothing and appended it to the nine other numerals already in use. Earlier societies were stuck with the unitary system, which used one symbol in a series of marks to answer the essential question: how many? That is an unwieldy system for large numbers; thus, higher-base systems were devised. Babylonian mathematicians invented a number system that used 60 symbols. It was a little cumbersome, but even today, 3700 years later, the essence of their system is still used to divide an hour into 60 minutes, a minute into 60 seconds, and a circle into 360 degrees.

Selection of a number system is a question of preference, because any integer can be expressed using any base. Choosing a number system simply questions how many different symbols we think are most convenient. The base 10 system uses 10 numerals; the radix of the system is 10. In addition, the system uses positional notation; the position of the numerals shows the quantities of ones, tens, hundreds, thousands, and so on. In other words, the number in each successive position is multiplied by the next higher power of the base. A base 10 system is convenient for 10-fingered organisms such as humans, but other number bases might be more appropriate for other applications. In any system, we must know the radix; the numeral 10 in base 10 represents the total number of fingers you have, but 10 in base 8 is the number of fingers minus the thumbs. Similarly, would you rather have 10,000 dollars in base 6, or 100 dollars in base 60? Table 1.1 shows four of the most popular number systems.

Binary Number System

Gottfried Wilhelm von Leibnitz, philosopher and mathematician, devised the binary number system on March 15, 1679. That day marks the origin of today's digital systems. Although base 10 is handy for humans, a base 2, or binary, system is more efficient for digital computers and digital audio equipment. Only two numerals are required to satisfy the machine's principal electrical concern of voltage being on or off. Furthermore, these two conditions can be easily represented as 0 and 1; these binary digits are called bits (*binary digits*). From a machine standpoint, a binary system is ruthlessly efficient, and it is fast. Imagine how quickly we can turn a switch on and off; that speed represents the rate at which we can process information. Imagine a square wave; the wave could represent a machine operating the switch for us. Consider the advantages in storage. Instead of saving infinitely different analog values, we must only remember two values. Only through the efficiency of binary data can digital circuits process the tremendous amount of information contained in an audio signal.

Whatever information is being processed—in this case, an audio signal that has been converted to binary form—no matter how unrelated it might appear to be to numbers, a digital processor codes the information in the form of numbers, using the base 2

Hexadecimal (Base 16)	Decimal (Base 10)	Octal (Base 8)	Binary (Base 2)
0	0	0	0000
1	1	1	0001
2	2	2	0010
3	3	3	0011
4	4	4	0100
5	5	5	0101
6	6	6	0110
7	7	7	0111
8	8	10	1000
9	9	11	1001
A	10	12	1010
B	11	13	1011
C	12	14	1100
D	13	15	1101
E	14	16	1110
F	15	17	1111

TABLE 1.1 Four common number systems.

system. To better understand how audio data is handled inside a digital audio system, a brief look at the arithmetic of base 2 will be useful. In fact, we will consistently see that the challenge of coding audio information in binary form is a central issue in the design and operation of digital audio systems.

In essence, all number systems perform the same function; thus, we can familiarize ourselves with the binary system by comparing it to the decimal system. A given number can be expressed in either system and converted from one base to another. Several methods can be used. One decimal-to-binary conversion algorithm for whole numbers divides the decimal number by 2 and collects the remainders to form the binary number. Similarly, binary-to-decimal conversion can be accomplished by expressing the binary number in a power of 2 notation, then expanding and collecting terms to form the decimal number.

The conversion points out the fact that the base 2 system also uses positional notation. In base 2, each successive position represents a doubling of value. The right-most column represents 1s, the next column is 2s, then 4s, 8s, 16s, and so on. It is important to designate the base being used; for example, in base 2 the symbol 10 could represent a person's total number of hands.

Just as a decimal point is used to delineate a whole number from a fractional number, a binary point does the same for binary numbers. The fractional part of a decimal number can be converted to a binary number by multiplying the decimal number by 2.

- [read Night Shift book](#)
- [click The Temple of Solomon: From Ancient Israel to Secret Societies here](#)
- [**Tomorrow's People: How 21st-Century Technology is Changing the Way We Think and Feel online**](#)
- [click The Freud Encyclopedia: Theory, Therapy, and Culture book](#)

- <http://diy-chirol.com/lib/Night-Shift.pdf>
- <http://aneventshop.com/ebooks/Emigrating-Beyond-Earth--Human-Adaptation-and-Colonization.pdf>
- <http://unpluggedtv.com/lib/Essentials-of-MATLAB-Programming--2nd-Edition-.pdf>
- <http://redbuffalodesign.com/ebooks/Perfectly-Miserable--Guilt--God-and-Real-Estate-in-a-Small-Town.pdf>