AUDIO Postproduction for FILM and VIDEO

After-the-shoot solutions, professional techniques, and cookbook recipes to make your project sound better.

JAY ROSE

Second Edition

ROUTLEDGE

Audio Postproduction for Film and Video, 2nd Edition

Audio Postproduction for Film and Video, 2nd Edition

Jay Rose, C.A.S.

Focal Press
Taylor & Francis Group
NEW YORK AND LONDON

First published 2009 by Focal Press

Published 2013
by Focal Press
52 Vanderbilt Avenue, New York, NY 10017

Published in the UK
By Focal Press
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

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

Library of Congress Cataloging-in-Publication Data
Application submitted

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN: 978-0-240-80971-8 (pbk)
ISBN: 978-0-240-81051-5 (CD-ROM)

Typeset by Charon Tec Ltd., A Macmillan Company

Contents

Contents

Acknowledgments

While mine is the name on the cover, it took a lot of people to make this book happen. CMP Books' Dorothy Cox and Paul Temme helped develop the concept for the first edition, letting me cover every concept I considered important. Dave Moulton (a Grammy-nominated recording engineer), and Dan Rose (Assistant Chief Engineer at WBUR, and also my son), made sure it was technically accurate. Lydia Linker kept its language understandable to non-technicians.

That edition was successful. CMP subsequently sold its interest to a much larger publisher, Focal Press. After a few years, Focal's Publisher, Elinor Actipis and I decided industry changes justified a new version. Their Associate Acquisitions Editor Michele Cronin has been in charge of its creation, shepherding resources within the publishing house and in keeping me focused. She also assembled a panel of expert educators and practitioners—including Chris Anderson, Patrick de Caumette, Harry Cheney, Jaime Estrada-Torres, Fred Ginsburg C.A.S., Jane Jackson, and Paul Werner—to review my ideas and explanations. Focal's Senior Project Manager Paul Gottehrer turned the words into a book. Despite all this help, blame me for any inaccuracy. And please, send me a correction… after all, it's my name on the cover.

I'm particularly grateful to friends who've been sharing their projects with me for more than three decades. Among them, Rob Stegman of Bluestar Media, Marc Neger of Creative Media Group, Francine Achbar of High Impact Marketing and Media, Mike Kuell of Jetpak Productions, and Trev Gowdy of Gowdy Productions let me use their clips as examples on the CD. Mitchell Greenspan and the DeWolfe Music Library provided the music, John Moran of Hollywood Edge the sound effects, and busy PBS announcer Don Wescott some narration. The female narrator on the CD is my late wife Carla, a former actress (and one-time spokesperson for Howard Johnson Motels) who then became a writer. She wrote dozens of popular books about Photoshop and digital imaging; in fact, she wrote two of them while helping me through the creation of this book's first edition. I owe her for that and pretty much everything else good in my life. The drawings and photographs, however, are my own work: It's a good thing I specialize in sound and not graphics.

Supplementary Resources Disclaimer

Additional resources were previously made available for this title on CD. However, as CD has become a less accessible format, all resources have been moved to a more convenient online download option.

You can find these resources available here: http://resourcecentre.routledge.com/books/9780240809717

Please note: Where this title mentions the associated disc, please use the downloadable resources instead.

Introduction

There's a funny thing about the work I do. When there's good cinematography, lighting, editing, or special effects, you can see it on the screen. Good audio postproduction, on the other hand, is mostly invisible. You can't tell how we fine-tuned an actor's dialog. You can't know which noises came from on-screen action, and which were added later. Good music editing sounds like the song was played that way in the first place. Equalization, dynamics control, noise reduction, and all the other processes discussed in this book are usually used as subtle corrections, not special effects. *When we do our jobs right, you don't know we've been there at all.*

This makes it nearly impossible for filmmakers to teach themselves audio post. Even if you've got the best ears in the world, you won't spot most of the techniques. In fact, relying on your ears alone can be a mistake. If you don't know what to listen for, or how a process actually changes a sound, you might do things that temporarily seem to help. But they actually hurt the finished track.

I wrote this book for two different groups of people. Students and beginners can use the book to understand how to improve a soundtrack, even after the actors have gone home. Experienced film producers and music engineers can learn how we solve some unique technical and creative problems of sound for picture.

The book is full of tutorials, practical examples, and cookbook recipes. But you won't need a technical background. Everything is in plain English, with plenty of demonstrations and visual analogies. I believe you're an intelligent person—after all, you bought my book—but you might not have much training in science. Ordinary common sense, a little bit of grade school math, and a CD player are all you'll need to follow the explanations.

INTRODUCTION TO THE SECOND EDITION

Film and video sound have been separate disciplines for most of my career. Film used platoons of specialists, sometimes with rooms full of interlocked transports and editing machines, to turn tiny bits of acting into tracks that would work in a theater. Video had to be smaller and faster. Even big shows often had only a few audio people, working with much longer performances and mostly the same equipment as music production. Film and video both had "sound for picture" and many of the problems were similar. But the equipment, techniques, workflows, and business models were different.

Digital changed that. Over the course of about twenty years, both worlds adopted the new audio technologies: they were faster, easier to use, and higher quality. But since the infrastructures and business models were already in place, we all pretty much kept working the way we had… just with newer equipment. Part of the reason those things didn't change was that film's *picture* was still an optical and chemical medium, with many expensive stages before it could be released. Video's picture was instantaneous.

Then pictures caught up. Some small digital cameras and desktop editing systems are now good enough to tell stories that can be shown in theaters. Blockbusters costing many millions of dollars can use essentially the same tools as well-produced corporate videos. Even if the production is shot on film (and a lot of "films" are now shot on video), most of the visual editing and post-production happens electronically. So Hollywood started looking to video sound for ways to work with it efficiently. Meanwhile, independents who had been shooting smaller videos started making features, and their sound people had to deal with theatrical-style tracks.

I'm trying to bring these worlds together. Shooting a Web video? Some feature-film track techniques might make it work better. And you theatrical guys? The video side learned tricks that could knock your socks off.

Of course this edition also updates sections on equipment and software, and fixes one or two things that I wasn't happy about in the first edition. There's a lot more meat here.[1]

Software Agnostic?

New technologies adopt old words. Before personal computers, I knew *agnosticism* only from its dictionary definition: a philosophy that says you can never understand First Causes or truly know the Ultimate Being.

Well, I understand and believe in software. But this book isn't about platforms, programs, or versions. I've used some specific applications to demonstrate certain functions, but you'll find both Mac and Windows here. Actually, the techniques in this book are scalable. They'll work on any system, from a laptop to a giant 35mm mixing theater. They were developed over 75 years of radio,

[1]There's also more words, thanks to more pages and tighter layout in this edition. To make it all fit, I had to sacrifice a few small sections that just don't apply as much any more. I've noted those deletions in this text, and you'll find the original versions on my website: www.dplay.com/book/missing.

movies, and television. And they'll stay useful as new software and hardware is developed. Expect to keep these pages for a long time.

HOW THIS BOOK IS ORGANIZED

Chapter 19, set off in gray and starting on page 401, provides quick help for the most common audio problems. If the solution can be explained in a couple of sentences, you'll find it there; otherwise, you'll be directed to the appropriate chapter.

Chapters 1–4: Technical basics

Chapter 1 explains what sound is, and how the digital version works. It's the basis you'll need for the rest of the book. But it's not heavy reading, and I've tried to use visual analogies as much as possible. Even you think you already know this stuff, glance through these pages anyway. You could have been misled by some of the myth and marketing hype that surrounds audio.

Chapters 2 through 4 are for those who want to build an efficient and reliable audio post setup, whether it's a single desktop computer or a fully professional suite. It includes guidelines, technical tips, and time- or money-saving shortcuts. Those chapters deal separately with practical acoustics, equipment, and software.

Chapters 5–10: Elements

Here's where we start turning sounds into a soundtrack.

Chapter 5 helps you plan and budget audio post, and understand the steps necessary for it.

Chapters 6 and 7 are about getting sound into your NLE: what settings to use when transferring from a camera , video deck, or separate audio recorder; how to maintain sync; how to record studio narrations and replacement dialog; and what to do when things go wrong.

The next three chapters specialize on the three principal sound streams: voice, music, and effects. Chapter 8 will teach you a more efficient and accurate technique for editing voices than you'll find in an NLE manual. It's how the pros do it. Chapter 9 covers how to find music for a production. Then it has you practice a *different* editing technique, the one top music editors use. Chapter 10 covers the full range of sound effects: where to find them, how to record your own, and how to fit them to picture.

Chapters 11–18: Putting it together

Chapters 11 through 16 are about shaping sounds with processes like equalization, reverb, and noise reduction. This may be the most critical aspect of creating a track, as well as the one most often done wrong. Even if you've worked in a music studio, read these chapters. Sound for picture really is different.

Chapter 17 is about mixing. It's a big chapter. There's more to a mix than you might think.

Chapter 18 covers what to do after the mix, including necessary preparations for theatrical release, TV, Web, or DVD audio. Along the way, you'll learn how compression schemes like Dolby Digital and mp3 really work—and how to make them work best.

There's glossary and CD track listing at the end of this book.

About the Cookbooks

The processing chapters, 11 through 16, include recipes: step-by-step instructions for common operations. Most are also demonstrated on the CD.

While the recipes suggest specific settings for each process, it's almost certain they won't be exactly right for your track. Every recording is different, and a tiny difference in your microphone or shooting situation can make my settings inappropriate. Read the entire chapter—not just its cookbook—and you'll know how to adjust them.

If you don't find a recipe you're looking for, read through the chapter again; it'll probably give you enough insight to solve the problem. If you're still lost—or think I've missed something important—write to me. Explain what you're trying to do. I'll try to come up with specific steps, and post them on my site for others to use. Also write me if you want to share your own recipes (with appropriate credit, of course).

Just don't expect me to solve a problem on your deadline; my clients get first priority. And never send me unsolicited files. If necessary, I'll give you uploading instructions.

From time to time, in the recipes and descriptions of processors, I'll put terms in small caps. These are the names of specific knobs on a screen or control panel. When same term appears elsewhere in normal type, it's talking about what the process does rather than a particular setting.

About the CD

This book includes a CD with about an hour's worth of diagnostics, demonstrations, examples, and tutorial tracks. It's an audio CD, rather than a data disc, so you can play it through the best speakers you own. But none of it's copy protected, so you can load it into your NLE as well.

Despite the lack of copy protection, the tracks are covered by copyright. Buyers of this book may transfer them to their hard drive to practice techniques or run the diagnostics. Any other use requires written permission. Information about licensing specific music cues is in the text.

ABOUT THIS BOOK AND PRODUCING GREAT SOUND

My other Focal Press book, *Producing Great Sound for Film and Video*, 3rd edition (2008), covers production sound—the technical and creative process of getting good dialog, interviews, or event audio while shooting picture. It starts with issues you should be thinking about before you ever pick up a camera. Then there are chapters on how to choose the most appropriate mic, how to hold a boom or get the best results from a radio lav, and how to get the best sound from cameras and field recorders. It's all important, but none of it belongs in this book you're presently holding.

Some parts of the two books necessarily overlap. The basics of sound, digital recording, monitoring, and editing belong in both books. But I've written about them differently in each, to give you a better chance of understanding these important concepts.

In other words, it's entirely reasonable to purchase both *Producing Great Sound* and *Audio Postproduction*. My publisher would certainly be happy if you did. But if you want only one:

- Choose *Producing Great Sound* for an overview of the entire audio process, with a strong emphasis on sound at the shoot.
- Choose *Audio Postproduction* for an in-depth discussion of turning that sound into a polished, finished soundtrack.

ONE OTHER THING...

It should be obvious, by now, that I haven't written a formal textbook. I'm a working sound professional who just happens to love the medium, spent a long time learning about it, and wants to spread this knowledge around. (Eventually, that'll mean more good films and videos for me to enjoy.)

In that respect I'm like a lot of my industry friends. We're not territorial, we're approachable when not overwhelmed by work, and we enjoy solving problems. Who knows? Maybe you and I will even work on a soundtrack together one day.

Meanwhile, let's talk about how to make one.

Jay Rose, C.A.S.
www.dplay.com
jay@dplay.com
Boston, July 2008

CHAPTER 1

Vibrations to Volts to Bits

REMEMBER THIS:

- We hear the world as ratios between sounds, not as absolute values. Volume and pitch differences—expressed as decibels and musical steps—are always comparisons between two values, like “twice as big” or “one-eighth.” Don’t think of these measurements as fixed units like ounces or inches.
- Sound exists only in time, and there is no such thing as an “audio stillframe.” This affects everything you do with a track.
- Although sound itself is analog, there are good reasons to keep a soundtrack in the digital domain. Some of the rules for digital audio are different from those for analog.

A TREE FALLS IN A FOREST…

Jokes and koans aside, we’ll assume it makes a sound. We’ll also assume you’re doing a film about logging, and the sound is part of a title sequence. You need to record the crash. You need to edit, process, and mix it for maximum impact. And you need to do this in a way that assures the viewer hears what you intended.

You probably bought this book to help accomplish goals like that. If you’re impatient, you can open the Table of Contents, find the tasks you want to accomplish, and turn to those chapters. Or, if you have issues with an existing track, flip to the gray-tinted pages at the end of this book: They’re a list of common problems and what to do about them.

But I honestly believe you’ll get a far better track—and ultimately be a better filmmaker—if you read this chapter first. It tells how a sound gets started, how it travels through the air, and how it turns into analog and then digital signals that are linked to pictures. When you understand this process, good sound becomes intuitive and creative. Then the rest of this book can serve as guidance, inspiration, and professional tips—things to build on, rather than steps to blindly follow.

This isn't rocket science, just grade-school math and intuitive physics. But because it isn't visual, many filmmakers surround the process with myth and hype. Take a few minutes now to think about how sound works; it'll save you time and money later.

So: To the tree.

> ***Gotcha***
>
> In this book, you'll find a bunch of *Gotchas*. They straighten out audio myths, correct misapplied principles, and fix other audio mistakes that can affect your track. They're based on real-world confusions I hear from filmmakers, read on Internet forums, or even find in software tutorials.

HOW SOUND WORKS

If our tree fell on the moon, no one would hear it. Sound requires air,[1] whose tiny molecules surround us. Anything that moves in the air reacts with those molecules.

- As our earthbound tree's leaves pass by, they scatter air molecules aside (a soft *whoosh*).
- As branches and limbs break, they vibrate. The vibration is transferred to nearby molecules (*crackle*).
- When the thick trunk gets close to the ground, it squeezes a lot of molecules in a hurry (*bang*).
- When the trunk lands, the ground vibrates and moves molecules next to it (*thud*).

These movements eventually transfer to our ears, and we hear the tree come down. Let's concentrate on just one aspect of that sound: the tree approaching the ground.

Before the tree starts to fall, air molecules surround it evenly. Their individual positions may be random, but the overall density is the same on both sides of the trunk. If we could enlarge and see these molecules, they'd look like the black specks all around Figure 1.1.

[1] Unless you're making a science-fiction film. It's an accepted movie convention that explosions, rocket fly-bys, and other events make noise in the vacuum of space. Sci-fi filmmakers also generally ignore another fundamental rule of physics, as you'll learn in a couple of pages.

As the tree falls, it pushes molecules directly in front of it and squeezes them together (lower right in Figure 1.2). At the same time, it forms a partial vacuum behind it—where there was tree, now there's nothing. This vacuum pulls nearby molecules into it, spreading them out (upper left).

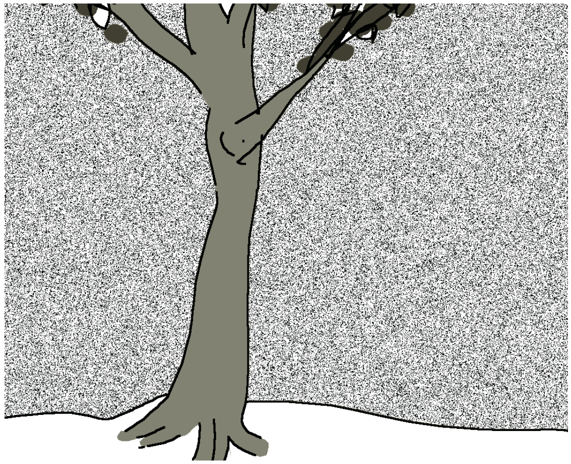

FIGURE 1.1 Air molecules distributed evenly around a standing tree.

The squeezed air molecules in front of the tree have to go somewhere, so they push against those farther out. Those newly-pushed molecules push others even farther, and so on. This creates a wave of *compression*, or higher air pressure, moving out from the tree. Meanwhile, the partial vacuum behind the tree forms a wave of low pressure, or *rarefaction*. It also moves out from the tree and draws nearby molecules toward it.

Trees aren't two-dimensional blobs like my drawing. In the real world, molecules flow around a three-dimensional trunk and branches. So as our waves spread out, other molecules rush in to equalize the pressure behind them. The result is a growing bubble of compression and rarefaction that constantly spreads out from the tree. If you could freeze it, it would look like Figure 1.3.

FIGURE 1.2 The falling tree squeezes molecules in front, and spreads out those behind.

The Speed of Sound

Push one end of a piece of wood, and the other end immediately moves. But molecules in air aren't connected that way. Push an air molecule, and it takes a moment for pressure to build up enough to move its neighbor. *This is a vitally important concept: The farther sound has to travel, the longer it takes.* Our pressure bubble spreads from the tree at about 1,100 feet per second.[2] It might seem pretty fast, but sound is a slowpoke compared to light.

When a sound is being picked up by two mics, differences in the length of time it takes to reach each can affect sound quality. We'll deal with that in Chapter 7.

[2] Actually, sound travels 1,087 feet per second at 32° Fahrenheit, gaining about 1.1 foot per second per degree, with minor variations based on pressure and humidity. We round it to 1,100 feet per second. That's close enough for filmmaking.

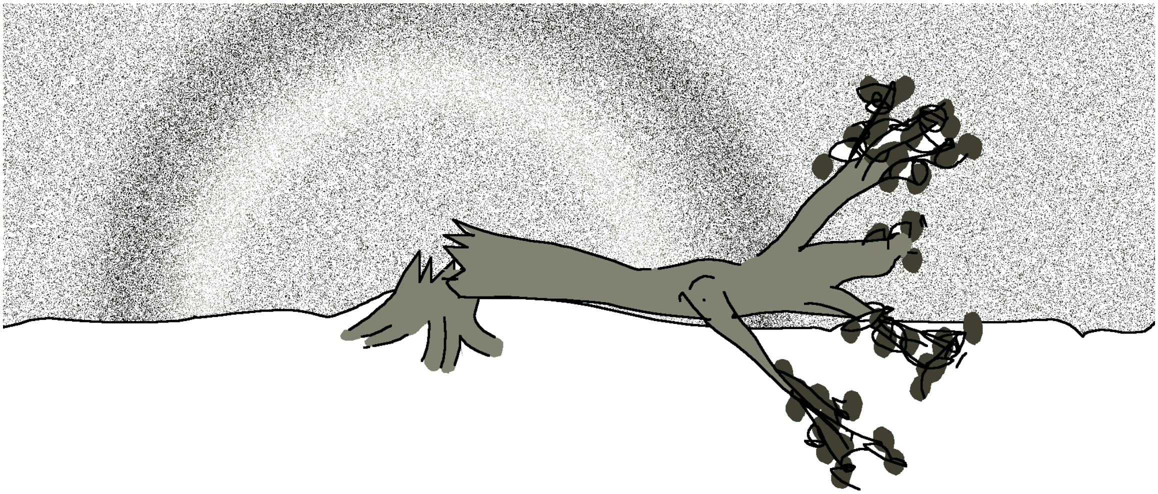

FIGURE 1.3
A growing bubble of compression and rarefaction spreads out from the tree.

Gotcha

Twelve yards is a frame! A video frame is roughly 1/30th second, and sound travels only about 36 feet in that time.[3] If you have a shouted conversation with someone across the street, their voice arrives about one frame after you see their lips move. If you see someone shoot a gun on the other side of a football field, the bang reaches you five frames after you see the barrel flash!

In other words, the real world is frequently out of sync. It affects the film world as well: In a very large theater, people in the back rows will hear dialog a few frames after it leaves the speaker mounted behind the screen.

Frequency

A tree falls just once, so our example creates a single pressure wave—something you're as likely to feel as to hear. But most things that make sound tend to vibrate for a while. Consider the imaginary guitar string in Figure 1.4. When it's pulled back (1.4a), it stores energy. When it's let go, it snaps forward quickly and creates a pressure wave. But since it snaps past its resting position (1.4b),

[3]An analog video frame in the United States and other National Television System Committee (NTSC) countries is 1/29.97 second; in Phase Alternating Line (PAL) countries it's 1/25 second. A film frame is 1/24 second. There are some other variations for special purposes, including digital television. These fractions are all pretty close, and have similar effects in terms of sync.

tension draws it back, creating rarefaction. The process repeats (1.4c) until all the energy from the pull is absorbed by the air.

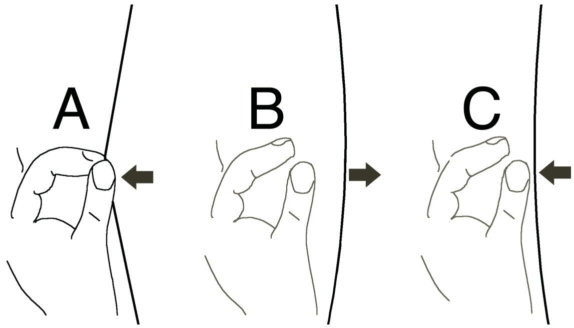

FIGURE 1.4
The guitar string vibrates back and forth after you let go.

The result is a regular, constant series of compression and rarefaction bubbles—a *sound wave*—spreading out from the string. If we could freeze it, it would look like Figure 1.5, with those bubbles spreading out from a source on the left. When the wave hits our ears, it vibrates our eardrums, moves the tiny bones inside, and is carried to the fluid in our inner ears. There it activates a few of the specialized nerves and we hear a *ping*.

Which nerves get activated is determined by how frequently the pressure peaks hit our eardrum,[4] which is how fast the string was vibrating. We call this the string's *pitch*. Hi-fi books often say we can hear pitches between 20 and 20,000 peaks per second, though the term is *Hertz* (Hz) and it's called a *frequency range* of 20 Hz to 20 kHz. Some people say sounds as high as 30 kHz are critical for truly hearing music.

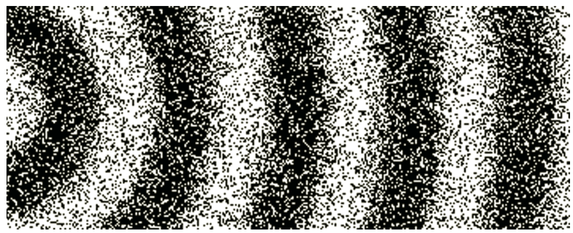

FIGURE 1.5
If you could freeze a sound wave, you'd see a regular pattern of compression and rarefaction.

A few exceptional humans can hear 20 kHz. Many young people can hear up to 17 kHz or so. But even the best high-frequency hearing deteriorates with age and can be destroyed by prolonged exposure to high-level sound (at the workplace, in a club, or because of overdriven headphones and car stereos). Very few adults hear quiet sounds above 15 kHz, though trained ears can often sense a sort of openness when higher frequencies are present.

Fortunately, there's not much going on up there. With very few exceptions, we perceive frequency as a ratio between two pitches, not as an absolute number of Hertz. So a difference of a few Hertz on the lower end of the audible band can be as important as a few hundred at the top. You can see how this works in Figure 1.6, a piano keyboard with three major thirds (a common musical interval) highlighted. I've written the frequencies above each note:

- Middle C is at 261.6 Hz (rounded to the nearest tenth). The E above it is 329.6 Hz, a difference of 68 Hz.
- The C that's two octaves below is 65.4 Hz, and its closest E is 82.4 Hz—only 17 Hz higher.
- The C two octaves above is 1046.5 Hz, and its E is 1318.5 Hz. That's 272 Hz higher.

---

[4] And by how much pressure the peaks have. This pressure component is the secret behind perceptual encoders like mp3, as you'll learn in Chapter 16.

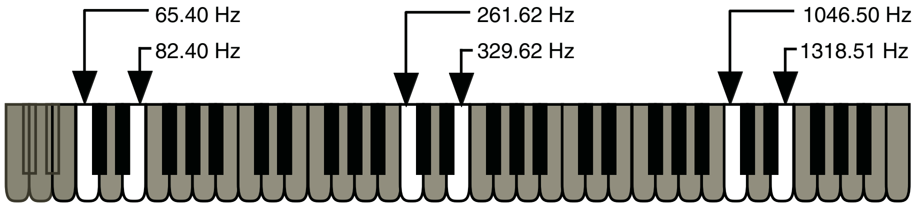

FIGURE 1.6
Musical thirds always sound like the same size jump, no matter how many Hertz are involved.

Yet the size of the jump from C to E sounds the same in each case, as you can hear on part one of Track 1 of this book's CD. It's also always the same ratio, about 1:1.25.

Gotcha

The misplaced middle. This nonlinear way we hear means that lower frequencies contain more information per hertz than upper ones. Many filmmakers—and quite a few programmers writing software to deal with sound—get this wrong.

Since the audio range is usually quoted as reaching 20 kHz, you might think that 10 kHz should sound like it's right in the middle. But that's not the case. We hear pitch differences as ratios, so the middle of the audio band is the frequency that has about the same ratio to the bottom as it does to the top. That works out to be about 1 kHz: it's 20 times higher than 50 Hz (a very low tone indeed), and of course it's 1/20th of 20 kHz. Keep listening to Track 1—a series of pure tones with voice identification—and you'll hear for yourself. Learning to identify approximate frequency ranges is an important step in building a soundtrack. Track 1 will help you practice this.

Hear for yourself

Track 1 of this book's CD is in two parts: First the jump from C to E in different octaves, then a series of tones with voice identification.

HARMONICS

Those pure tones on track 1 sound electronic, because real-world sounds are seldom pure. Purity has a precise meaning here: A pure tone has energy at only one frequency. Engineers call it a *sine wave*.[5]

[5] That's because a graph of its pressure looks like a graph of the geometric function, something important to mathematicians.

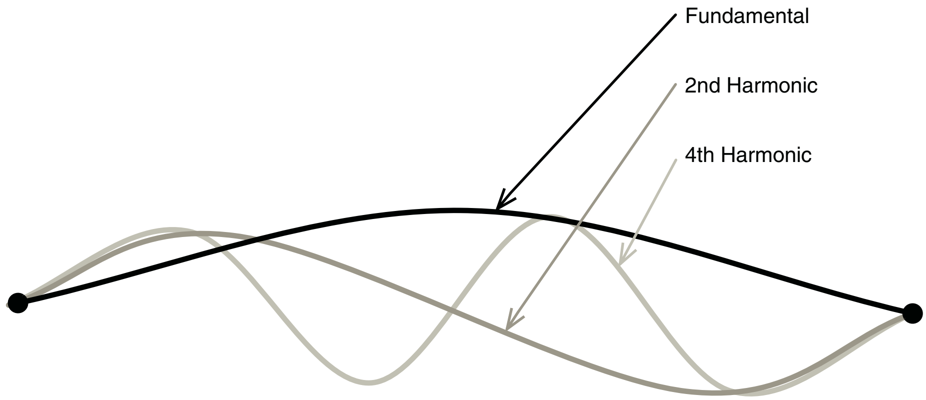

FIGURE 1.7
A fundamental and two harmonics on a vibrating string.

Remember our imaginary guitar string vibrating at one pitch? For any particular string and tension, the pitch is determined by what size vibration fits exactly from one end of the string to the other. This pitch is called the *fundamental frequency*. But guitar strings are flexible: they bend in the middle, so multiple instances of shorter waves at higher pitches can also fit neatly from end to end. These are called *harmonics*, and they're in precise ratios to the fundamental. Figure 1.7 shows a fundamental, along with harmonics at twice and four times the fundamental pitch.

Because fundamental and harmonics happen simultaneously, they combine and force the vibrating string into complicated, constantly-changing wiggles. You can see this with an oscilloscope, a gadget which displays changing voltages over time. (Normally, you'd use a microphone to convert changing air pressure to voltage for the scope. But since we're dealing with an imaginary string, I created its wave in software.)

Hear for yourself

Track 2 has two sounds: a pure tone at 200 Hz, and one adding just the two harmonics shown in Figure 1.7 (400 Hz and 800 Hz). Even though this is a very simple sound, you should hear a difference. The first tone will sound electronic. The other is more like a pipe organ.

Figure 1.8 is a scope photo of the combined fundamental and harmonics from Figure 1.7.

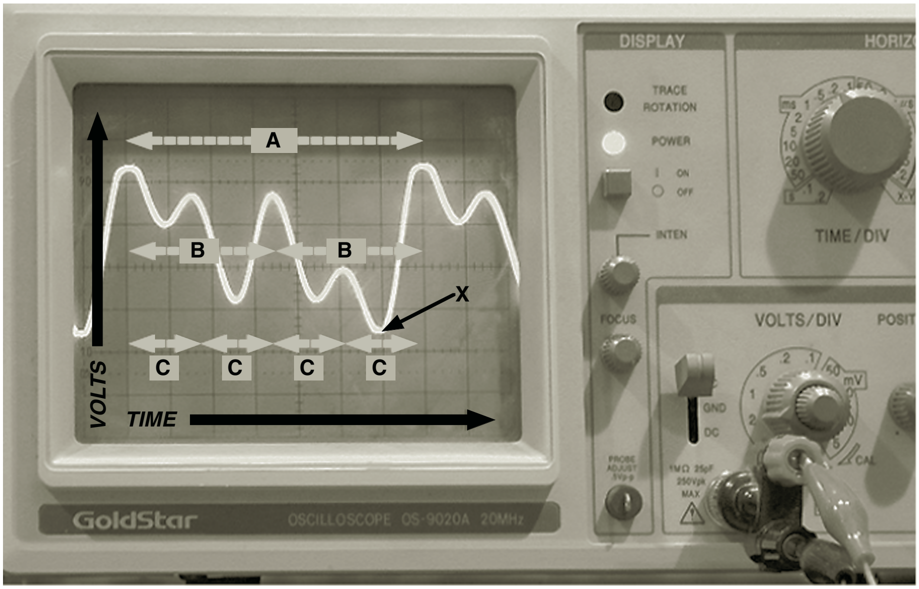

FIGURE 1.8
The complex wave from Figure 1.7, as shown on an oscilloscope.

- The area A is the fundamental, measured from one compression peak to the next.
- B is the second harmonic. We see exactly two of them for each fundamental wave, meaning its frequency is exactly twice as high.
- Note how the compression phases of both A and B add together at the start of the wave.
- C is the fourth harmonic, exactly four times the frequency of the fundamental.
- Rarefaction phases of all three waves coincide at point X. They also add together, giving this point the lowest voltage of the entire wave.

There's no theoretical limit to how many harmonics can exist for any fundamental (though you might not hear the highest ones). In these examples, I've shown only even-numbered harmonics; real-world sounds can also have odd-numbered ones. Also, I've shown the fundamental and harmonics as having the same strength. But real harmonics' strengths are determined by mechanical properties of the vibrating element, resonances in the instrument body, and even where you plucked or hit it. The strength of each harmonic changes over time. It's these differences in harmonics that make a Fender Classic sound different from a classical guitar, or from a student's classroom guitar.

FREQUENCY RANGE

Without harmonics, instruments playing the same note would sound virtually the same. You wouldn't be able to tell the difference between a violin and a flute. That's why it's important for a sound system to carry frequencies above the 5 kHz fundamental of the highest orchestral instruments.

But the system doesn't have to extend much higher. Most instruments pack a lot of harmonics below 10 kHz. This is a good thing, because most real-world playback systems are useless half an octave above that limit. Despite how speaker manufacturers may brag about "20 Hz to 20 kHz ranges," most computer, multimedia, and consumer-level stereo speakers fall far short of that goal. The advertising relies on fudged numbers or unspecified qualifiers. Test your system (and your ears) with Track 3 of the CD.

Gotcha

Few sounds have just one frequency! Real-world sounds are rich, moving tapestries of harmonics that extend across the audible spectrum—far more than the two in our simple example. So don't expect to find a magic frequency in an equalizer that will emphasize only violins in the orchestra, or separate male voices from female.

Hear for yourself

Track 3 consists of a few repeating musical selections[6] that alternate between being played at full fidelity and through a filter that cuts off at specified high frequencies. If your speakers (and ears) are really hearing above those frequencies, you'll be able to hear when the filter switches in and out. But if the highs aren't making it through your system, the filter won't make any difference.

Try playing Track 3 through your best hi-fi speakers and note how high you can hear. Then try loading it into your NLE and listening to it there… and then, listen on a laptop. You may be shocked at how much you're missing. But also notice that even when I apply a fairly low cutoff filter (7.5 kHz), enough harmonics are present that the music is still recognizable—if somewhat dull.

By the way, this track was produced with very high-quality digital filters, far sharper than those provided with most NLEs or audio programs. If you try to create a similar test, be sure your filters aren't lowering sounds below the specified cutoff; otherwise, the results will be meaningless. There's more about filters in Chapter 11.

UNPITCHED SOUNDS

Some sounds don't have regular vibrations. Our falling tree trunk, a gunshot, or the clap of a film slate create a single pressure wave of no particular frequency. Many continuous sounds don't have a frequency either; the crunch of breaking branches, the hiss of the phoneme /s/, or the fizz of a glass of soda are

---

[6] "Swing Out Brother" (J. Trombey, DWCD219/10), "Queen of the Night" from *Magic Flute* (Mozart, DWCD142/2), "Rock Hits" (R. Hardy/B. White, DWCD293/11). All from the DeWolfe Music Library, protected by copyright and used by permission. You'll learn more about this library, as well as a lot more about working with music, in Chapter 9.

collections of essentially random movements. The best we can do is describe their predominant frequency ranges. When we talk about the low pitch of rumbling thunder, we're not talking about a specific pitch but about how this random noise centers around low frequencies.

Gotcha

Hiss doesn't have frequency. Unpitched sounds are virtually impossible to remove with an equalizer, unless you don't mind doing serious damage to dialog. This includes a lot of the noises that plague film and video shoots: ventilation systems and computer fans, traffic, and the electronic junk from wireless mics or poor camera preamps. The best you can do is make these sounds less annoying; see Chapter 15 for some tips.

HOW HIGH IS ENOUGH?

While 20 kHz is the standard upper frequency limit for DV tape, don't assume every soundtrack must extend that high:

- The upper limit for U.S. analog television and FM stereo is 15 kHz. Frequencies higher than that cause transmission problems. Knowing this, TV manufacturers don't worry about building higher-frequency capability into their products.
- Even digital broadcasting may be limited to near 15 kHz for a while. Station hard-disk servers and satellite links often cut off at that frequency to conserve resources.
- Most Hollywood films made before the 1970s—including most popular musicals—stop around 12.5 kHz, the limit for optical soundtracks. Outside of major cities, you'll find a lot of theaters still don't go much higher.
- Acoustic and orchestral music has surprisingly little energy above 15 kHz, though a good ear can certainly hear when it's missing. (Heavily processed pop music sometimes emphasizes very high frequencies electronically, often to the point of harsh distortion.)

Loudness

Our falling tree will move a lot more molecules with more force than the guitar string; that's why we hear it as "louder." The human ear can handle an amazingly wide range of loudness; the sound of a nearby jet plane hits your eardrum with about 10,000,000,000 times stronger molecular movement than the quietest tones used for testing your ears. We cope with this range by hearing loudness as ratios instead of absolutes, just as we do with frequency. The *difference* in loudness between one tuning fork and a pair of them sounds the same as the difference between one tuba and two.

So in the technical world, we use ratios to gauge the strength of a sound: The sound is compared to a standardized loudness, and its volume is expressed as the resulting fraction. The fastest way to deal with complex fractions is a math shortcut, *logarithms*. You don't have to know how logarithms work. Just remember that they stand for fractions, but with one important twist: adding two logs actually multiplies the numbers they represent. Subtracting them divides the numbers. In other words, $log\,1 + log\,2 = log\,3$ really means *ten times a hundred equals a thousand* … something every schoolchild knows.[7]

Decibels

The common measurement to express the level of sounds and electronic signals, the *decibel* (dB), is a logarithm. Unlike a watt or a degree Fahrenheit, which are measurable quantities, decibels are fractions of something else. They're mostly meaningless until you know what they're fractions *of*.

When we talk about the loudness of a sound in air, we're almost always referring to the fraction formed by it and a theoretical "softest sound a human can hear," the *threshold of hearing*.[8] Log 0 means a fraction of 1/1, so something at exactly that threshold would be written as 0 dB SPL (sound pressure level). Few things are that soft. The quietest recording studios are about +30 dB SPL. A good shooting stage would be around +40 dB SPL. Average dialog level at a boom mic may be around +65 dB SPL. Average playback level in a movie theater is supposed to be 85 dB SPL, but is usually much higher, particularly during coming attractions. A jackhammer, close up, would be around +125 dB SPL.

Gotcha

A sound can't have "so many decibels." Decibels make sense only when you know their reference. We've talked about dB SPL for sounds in air because SPL is a standard reference. Sounds in a digital signal chain are almost always measured as dBFS (defined later in this chapter); those on an analog wire may be dBu, dBm, or dBV (discussed in Chapter 3). But decibels without extra initials—just plain dB—are simply fractions, useful only when talking about how much a sound should be boosted or lowered … or when you want to impress a novice with meaningless audio jargon.

The Inverse Square Law

Remember how the falling tree created a spreading bubble of compression and rarefaction? As the bubble gets farther from the tree, its surface area grows, just like the skin of an expanding balloon. But the total energy can't change; that would require another tree. When the bubble was small, its energy was directed at only a few molecules. When the bubble is bigger and there are more molecules around it, it can't push each one as hard.

[7] That's if the logs have a *base* of 10, which is common in audio. Mathematicians use other log bases for other purposes.
[8] A pressure fluctuation of 0.0002 microbars, which is an energy of 0.0002 dynes per square centimeter. It may be theoretical, but it's set by international standard.

This is a long way of explaining what you already know: a falling tree sounds louder when it's nearby and softer when it's far away. But few people appreciate *how much* the volume changes based on distance. The effect is geometric, because the surface of a bubble grows much faster than its diameter.

Each time you *double* the distance from a sound source, the sound's power is one quarter as much. Each time you *halve* the distance, it's quadrupled. And these effects multiply. If you quadruple the distance, the sound's power is one-sixteenth. This is called the *inverse square law:* Sound power will change according to the square of the distance change.

Gotcha

Closer is better. A lot of filmmakers never consider how many ways the inverse square law applies to their work. A camera mic at four feet will pick up only half as much dialog as a boom at two feet. But noises and echoes can come from all around a room, so getting farther from the actor doesn't mean you're getting farther from these problems. In practical cases, doubling the distance from desired sound usually makes noises and echoes twice the volume by comparison. Unless you're dealing with ideal acoustic circumstances, there's no way a camera mic at six to eight feet can sound as good as a properly used boom or lav. This is just physics; it doesn't depend on how much you spend for that mic.

The same principle affects postproduction setups. When you listen to monitor speakers, you also hear reflections from nearby walls. Unless you're in a perfectly tuned acoustic space, the reflections subtly change what you hear, which can affect your mixing and processing decisions. Move the speakers closer to your ears, and you'll make better decisions.

Volume Changes Over Time

While the fundamentals of a violin and piano playing the same note may be identical, their *envelopes* are different.

Very few sounds, outside the test lab, are absolutely constant. Their volume can vary over times ranging from a few hundred milliseconds to many seconds. For example, a violin string starts vibrating weakly as the bow begins to move across it, builds strength, and then continues to sound while the bow keeps moving. A piano string, on the other hand, is violently struck by a hammer and starts making noise immediately. But because there's no other hammer hit, it starts to fade down almost immediately. How a sound's volume changes over time is its envelope.

Figures 1.9 and 1.10 compare envelopes of a violin and piano, shown in an audio program's waveform display.[9] Each picture represents about two seconds.

---

[9] I couldn't use an oscilloscope for this, because envelopes are too slow to display on that device.

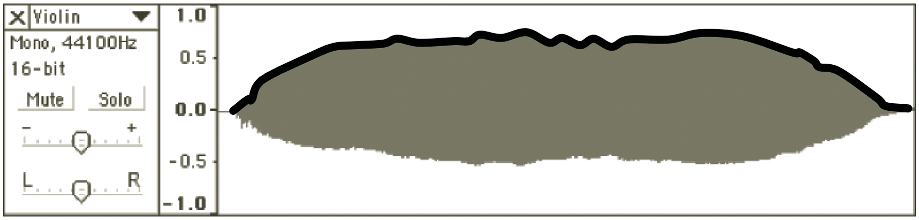

FIGURE 1.9
A violin sound grows slowly, but continues as long as the instrument is bowed.

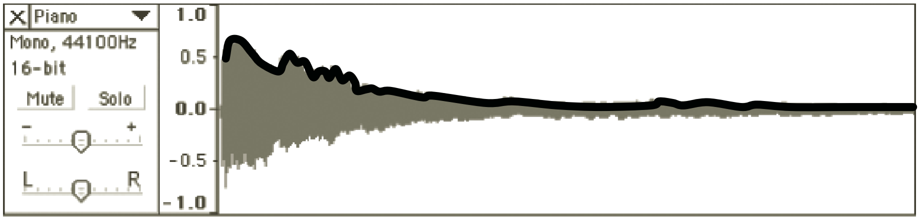

FIGURE 1.10
A piano note starts suddenly, then begins to fade almost immediately.

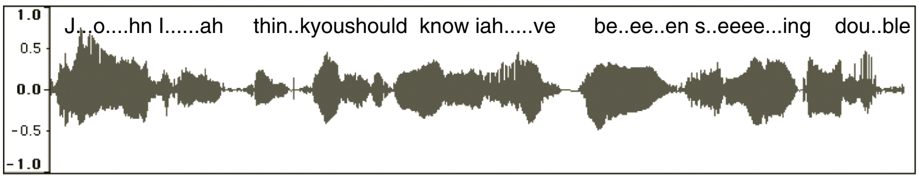

FIGURE 1.11
Spoken-word envelopes don't necessarily correspond to individual words.

I've drawn heavy black lines along the tops to trace the envelopes. The jaggy bits you see on the bottom are individual soundwaves.

The envelope of human speech depends on what's being said but doesn't necessarily follow individual words. Figure 1.11 shows an actress saying, "John, I think you should know I've been seeing...." She's speaking smoothly, but some words are jammed together and others are slowed down. I added text above each individual sound so you can see how uneven her envelope is.

Hear for yourself

Track 4 is the sounds whose envelopes we've just looked at: violin and piano (both with and without harmonics), and the woman's dialog clip.

Gotcha

When envelope and frequency overlap. The *j* sound that starts "John" is about 1/20th of a second long,[10] but as we've discussed, frequencies around 20 Hz can be audible. So is the *j* sound part of the envelope or a single soundwave by itself?

In this case, it's part of the envelope, both because the woman's natural voice doesn't go down as low as 20 Hz and because it's nonrepeating. This issue becomes important when you're dealing with dynamics processing to automatically control levels. If the system isn't set properly, you'll hear a kind of muffling. Learn to avoid it in Chapter 12.

Analog Audio

So far, we've examined sound mostly as variations of air pressure. Electronic equipment deals with electrical pressures (voltage) and quantity (current), not air pressure changes. The basic way to turn sound into electricity—dating from the very first telephones—is to make a changing voltage that corresponds exactly to the sound wave. When there's compression, the voltage is positive; when there's rarefaction, it's negative. The amount of voltage at any moment reflects the strength of the wave. In other words, it's an *exact analogy* of the sound.

You may be working with digital cameras and editors, but this analog audio is still at the heart of the process. Microphones and speakers are analog devices—even if they have digital connections—and most postproduction studios have a lot of analog wiring. But analog signals are fragile. Things that change the voltage add distortion or noise. If you don't understand what's going on with analog audio, bad things can happen to your track.

PRESSURE TO VOLTAGE AND BACK

Microphones and speakers work like tiny generators and motors. The dynamo at a power plant uses a rotating turbine to spin a coil inside a magnetic field; this creates a changing voltage. The dynamic mics used for handheld interviews and voice-overs work exactly the same way, except the coil doesn't rotate. Instead, it's attached to a thin diaphragm that flexes in response to air pressure.

A speaker is a microphone in reverse (Figure 1.12). Instead of a diaphragm, it has a large paper cone attached to the coil. Changing voltage—the audio signal—is applied to the coil, making it move back and forth inside a magnetic field. This vibrates the cone, which then pushes and pulls against air molecules.

Dynamic mics and speakers are so similar that their functions can be interchangeable. Many intercoms use a speaker to both radiate and pick up sound.

[10] Actually, the sound *j* consists of a very fast "d"—about 1/60 of a second long—followed by a drawn out ZH "zh" (say "John" very slowly and you'll hear what I mean). This kind of analysis is important for serious dialog editing, and is covered in Chapter 6.

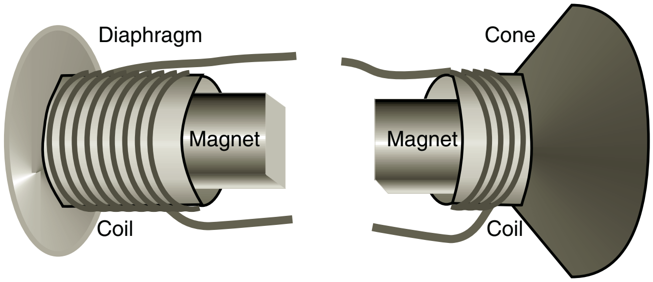

FIGURE 1.12
Dynamic mics (left) and speakers (right) have very similar construction.

Conversely, radio station engineers have been known to play tricks on air talent by temporarily sending the talkback signal into a mic. Done right, the mic itself will start to whisper at the announcer just before a broadcast.[11]

The condenser elements in most dialog mics don't have magnets and won't work as speakers. But their function is the same: turning changing air pressure into changing voltage. There's a deeper discussion of mics in Chapter 7.

Balanced Wiring

When you connect a mic to a camera, or an NLE to a mixer or audio input, the analog wiring can act like an antenna and pick up nearby electric fields. Video monitors and data cables broadcast high frequencies; power cords and the wiring in the walls radiate low-frequency noise. Since you can't avoid these fields, the trick is keeping them from mixing with your track.

As you probably remember from grade school or model trains, any electric circuit requires two conductors. Most consumer line- and mic-level wiring accomplishes this with a single conductor for each audio channel, surrounded by a foil or braid shield that serves both as the return conductor and a barrier to high-frequency noise. This scheme, *unbalanced wiring*, is cheap and moderately effective. Unfortunately, the shield isn't very good at stopping power-line hum, timecode, or crosstalk from other audio channels.

The solution is to have two separate but identical wires, twisted very close to each other, inside the shield. While one wire is carrying current in one direction, the other's current is going the opposite way. Analog audio constantly changes polarity to reflect the compression and rarefaction of the sound waves. But at any given instant, the voltage on one wire will be positive while the other is negative. That's why it's called *balanced wiring*. It's connected to an input circuit that looks just at the voltage difference between the wires. Figure 1.13 shows how this works.

---

[11] Done wrong, it can damage the equipment. I was a little devil of a radio engineer when I was much younger, and made sure to do this hack properly.

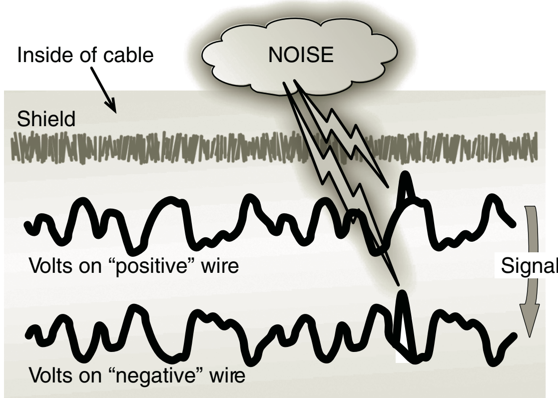

FIGURE 1.13
Balanced wiring uses symmetrical signals to reject asymmetrical noise.

Since the wires are physically identical and very close together, any noise that bursts through the shield is picked up by both. The voltage that results from the noise is the same on both wires, so subsequent equipment doesn’t detect it. The math is easy:

Conductor	Original signal	Noise	Total on wire
A	+1.0 v	+0.5 v	+1.5 v
B	−1.0 v	+0.5 v	−0.5 v
Difference between A and B	2.0 v	0 v	2.0 v

Balanced wiring does not require XLR connectors, nor do these connectors guarantee a circuit is balanced. But they are often used together, particularly in microphone cables. Mic signals are so small that they’re easily polluted by external noise.

Gotcha

Balanced wiring requires balanced equipment. Balanced wires don’t actually reject the noise; that magic happens in whatever the wires get plugged into. Both the signal source and the destination need special circuits for balancing. These circuits can be expensive and are usually left out of prosumer equipment.

While unbalanced wiring doesn’t offer this protection, it doesn’t have to be noisy if you plan it properly. Chapter 3 discusses practicalities of both wiring schemes and how to connect between the two.

DIGITAL AUDIO

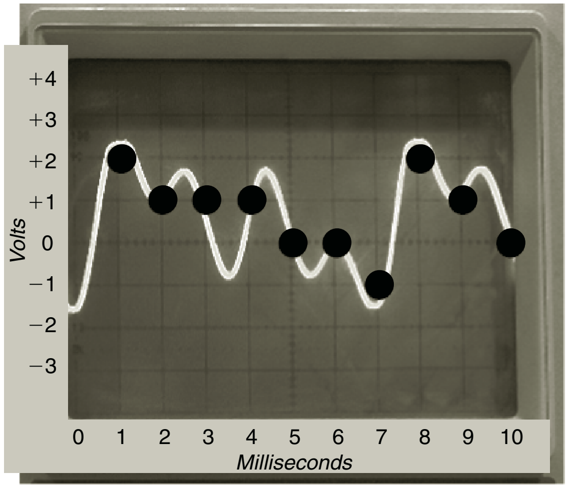

FIGURE 1.14 Sampling a soundwave's voltage, once per millisecond.

Those bugaboos of noise and distortion in analog audio are cumulative. Each time a signal goes through a different processor or is stored on another generation of analog tape, it gets slightly noisier or more distorted. In the days before digital recording, a movie soundtrack could go through dozens of processing circuits and six magnetic or optical generations between stage and moviegoer. Finding the best compromise between noise and distortion was a constant battle.

But computer data doesn't deteriorate that way. The words I'm typing now are captured flawlessly to RAM, written to hard disk, and e-mailed to a publisher. Editors, tech reviewers, and designers will work on successive copies before it's sent to the printer. Yet, no data errors are introduced by these multiple computer generations.[12] Digital ones and zeros are unambiguous and easy to copy, and computers include safeguards to correct many errors.

If we turn an analog audio signal into digital data, it should be just as robust. Of course the process of turning constantly changing analog voltages into data and back has its own pitfalls. Twenty-five years ago, when the technology was new, these problems gave digital a bad reputation among audio purists. Today, we've learned to deal with all of them. Digital, when done right, is the best recording medium we've got. Musicians may argue that some of analog's imperfections are part of their art, but for professional film and video, digital's advantages are so great that the older method has been virtually abandoned.

Digitizing

The process of turning analog to digital is easy to visualize with an oscilloscope. For any moment, you can measure exactly what the voltage is. Scopes have grids printed on their displays to track time and voltage; the values depend on how you set the controls. In Figure 1.14, I've assigned the grid very simple numbers: each vertical unit represents one volt, and each horizontal is one millisecond. The wave is the dual-harmonic we examined earlier.

When children first learn to count, they do it in whole numbers: *1 apple, 2 apples, 3 apples*. Their math skills don't yet include the concept of "half an apple." Neither does a computer's. One bit is the smallest possible unit, and there can be no such thing as half a bit. But we can assign that smallest bit any

[12] I'm not guaranteeing this text is error-free. Just that the errors are human (and probably mine).

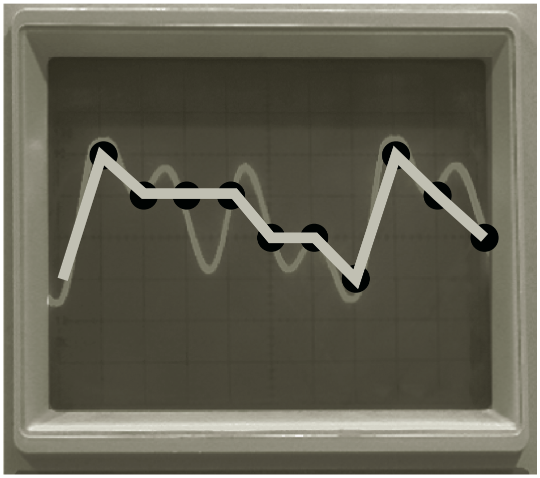

FIGURE 1.15
Our low-resolution recording isn't very accurate.

unit we want. It can be something whole, like a penny. Or it can be a fraction of a unit, like a hundredth of a volt.

These values may be tiny, but if you string bits together to make words, you can express much bigger numbers. A one-bit digital word has two possible values, 1 or 0. A two-bit word has four possible values: 00, 01, 10, 11. Word processors typically express characters and letters with a *byte*, or 8-bit word. Modern computers frequently string multiple bytes together into a single word. Using two bytes for a 16-bit word, you get 65,536 possible values.

Let's keep things simple. For our first digital audio, we'll use three-bit words, which have eight possible values. We'll count in volts, scaled from −3 to +4 on our scope screen. Digital 000 would be −3 volts, digital 001 is −2 volts, and so on. The wave fits nicely in that range.

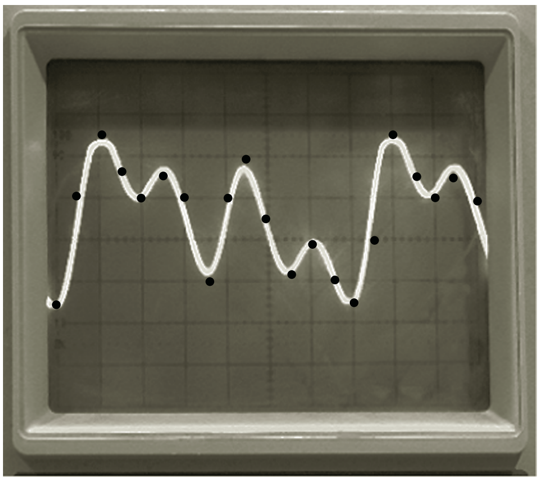

FIGURE 1.16
Doubling the number of bits lets us mark half-volts. Doubling the speed gives us twice as many samples.

Using the grid, let's turn our wave digital. Once per millisecond (that is, once per horizontal unit) we'll read the voltage. Since our counting unit is a volt, we round to the nearest whole one:

At 1 ms it measures	+2 volts
2 ms it measures	+1 volt
3 ms it measures	+1 volt
… etc.	

This string of analog numbers, *+2, +1, +1…* could be recorded on digital tape as the three-bit words: *101, 100, 100…* I've marked the value for each millisecond with a black dot on the screen. You can see in Figure 1.14 how the dots follow the sound wave.

To play this string of numbers back as sound, we create an analog wave that has the right voltage at the right time. Visually, this is as easy as connecting the dots (it isn't much harder in audio circuits). But the result—the light gray line in Figure 1.15—doesn't resemble the original much. What went wrong?

Actually, two things are at fault. Those rounded whole volts—the best we can do with our three-bit words—are too far apart to accurately record the wave. And a sample rate of once per millisecond isn't fast enough to accurately capture an 800 Hz harmonic. Both problems are easy to fix.

Each time you add one more bit to a digital word length, you double the number of possible values. So let's bump up to four-bit recording: instead of just eight possible values between −3 and +4, we now have 16. This lets us count in half volts. Let's also double the speed and take a fresh sample every half millisecond (you'll learn why that's sufficient in the next few pages). Figure 1.16 shows these new samples as smaller black dots: twice as many as in Figure 1.14, now rounded to the nearest half volt.

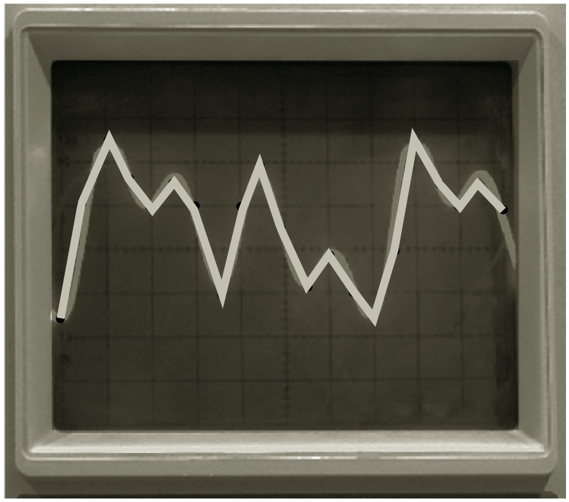

FIGURE 1.17
This "doubled" recording gives much more accurate results.

We haven't added much data—four bits would be considered a very low-resolution system. Two thousand samples per second (usually written as a *2 kHz sample rate*, or *2 kHz s/r*) is only fast enough for bass-effect channels. But the reconstructed wave is a lot more accurate. Figure 1.17 shows how well it follows the original. Even more important, CD Track 5 lets you judge the result: an actual four-bit, 2 kHz s/r recording of the multi-harmonic wave from Track 2. It was converted back to CD standard so you could play it, but that didn't restore the lost data. You're hearing what four bits at 2 kHz s/r really sounds like!

Hear for yourself

Track 5 lets you hear the actual wave produced by Figure 1.17, along with the original wave for comparison.

Bit Depths and Full Scale

Track 5 sounds as good as it does because even low-bit systems can handle simple waves recorded at the right level. Even today's basic systems use 16-bit audio words for considerably more accuracy and volume range. When the signal has to be converted to or from analog, circuit designers choose a maximum possible voltage. This is defined as when all 16 bits are ones, or *full scale*. The smallest bit is then 1/65,536 of full scale, or about 0.00013 dB of it.

Some rounding off is still necessary because even those tiny bits aren't the same as a constantly changing wave. The worst case is when a signal is halfway between two minimum bits; it causes a maximum error of about 0.00007 dB of full scale—far less than humans can hear. But each time you manipulate the signal—say, by changing volume or adding a processor—you're doing some math on the samples. Since the result isn't likely to fall perfectly on a whole-bit value, it's rounded off again. These errors are cumulative. Do enough processing, and the result is audible distortion or noise. For that reason, audio professionals now use 24-bit words, for a possible error of 0.00000026 dB. They do as much as possible in that domain, and then reduce the final output to 16 bits when necessary for compatibility with other media.

There's another way to consider bit depth. Each bit doubles the possible number of values, increasing the ratio between media noise and full scale by 6 dB[13] (that ratio is often called *signal/noise*, or *s/n*). An eight-bit signal, common in early computers, has 48 dB s/n. Sixteen-bit DV and CDs have a theoretical 96 dB s/n, though the analog circuits are seldom that good. Twenty-four-bit signals have 144 dB s/n—beyond the range of the most golden-eared humans, as well as most electronics—but it provides significant margin for error.

Some DV cameras also support 12-bit recording at a reduced sample rate, to double the number of audio tracks on a videotape. Twelve bits means 72 dB s/n, which isn't bad compared to older analog media. But 12-bit recordings are frequently noisy. That's because there's no safety margin, and it's almost impossible to set levels accurately enough in the field to take full advantage of the range.

Gotcha

Digital signals have no relation to voltage. The digital audio values on a CD or stored on your hard drive represent fractions of full scale. But the actual value of full scale depends on the analog circuits: it might be 0.01 volts at a camcorder's mic input, 3.5 volts on a balanced line connector, 8 volts in a hi-fi speaker, or whatever the equipment designer prefers.

This is a good thing. It means while the signal is in the digital domain, you don't have to worry about absolute values at all. The software doesn't care whether a track is destined for a thousand watt theater system, the Internet, or your own headphones.

For this reason, digital audio levels are always expressed as dBFS, a decibel ratio to full scale. Full scale is the absolute maximum, so a ratio of 1:1 (0 dBFS) is the loudest you can get. Because no voltage changes can be recorded above full scale, if the original analog voltage exceeds the design limit, the digitized result has an absolutely flat-topped waveform: total (and very ugly) distortion.

Circuit designers frequently flash the "overload" light when a signal approaches −1 dBFS to avoid this distortion. But there really isn't a digital overload. That would require a digital signal greater than 0 dBFS, which is impossible.[14]

[13] Actually, 6.0205999 dB.But you can round it off. What's a few hundredths dB among friends?
[14] Except in the case of one particular and ugly-sounding test signal, which never occurs in a soundtrack.

DITHER

Humans can't hear a single sample's worth of rounding error, but a bunch of them in a row produce a distortion that's annoying because of the way it follows the signal. The solution is to add analog noise—dithering—at about one-third the level of a single bit. This randomizes the error so it doesn't sound like distortion any more. The ear is good at tracking desired sounds even when they're obscured by noise, so dithering can effectively extend the dynamic range. Dithering can also be noise-shaped, making the random signal stronger at frequencies where the ear is less sensitive. Its randomness continues to hide distortion, but we don't hear the noise as much.

Gotcha

What signal to noise? When measuring s/n in the analog world, it's customary to refer to a nominal signal level, typically 0 VU. That's because while analog circuits start distorting once you reach that level, you can still record a louder signal. Momentary peaks 6 dB higher than 0 VU can sound fine, even though longer sounds that loud would reveal the distortion.

Since digital recording has an absolute limit of 0 dBFS, that level is usually used for digital s/n measurements. But prosumer audio gear often has its nomal operating level set to −12 dBFS (−20 dBFS for pro gear) to leave room for peaks. When comparing digital s/n to analog, you must take that safety margin into effect. So while 16-bit digital has 96 dB s/n, a 16-bit miniDV isn't capable of sounding better than an 80 dB analog s/n.[15]

Audio programs may give you some choices for dithering when you need to reduce bit depth. If your program does this, use dithering and play with the noise-shaping options. The effects may be too subtle for most media production, unless you're translating a 16-bit signal to 8 bits. Most NLEs' audio capabilities aren't sophisticated enough to use dithering at all.

Sample Rates

Sound exists only when pressure *changes* over time. Eliminate the time component, and all you've got is air pressure—useful for weather forecasts, but not soundtracks. So it makes intuitive sense that how often you sample a signal's level influences how accurately it gets recorded.

This principle was known long before practical digital recording was invented. In the 1920s, a Bell Labs engineer, Harry Nyquist, proved mathematically that the highest frequency you can faithfully reproduce has to be less than half the sample rate. Sounds above the *Nyquist Limit* start to mix with the sample rate itself. The resulting data implies additional frequencies, which aren't

[15] That's a theoretical maximum, with ideal electronic circuits. Many mass-marketed cameras don't even come close.

harmonics of the original wave. When you play back the data, those frequencies are re-created as well as the desired one. What you get is a horrible squeaking or whistling overlaid on the original.

To avoid this, analog-to-digital conversion circuits use *antialiasing* filters to reject high frequencies before they get digitized. Digital-to-analog circuits have similar *reconstruction* or *smoothing* filters. They're often set to about 45 percent of the sample rate.

- CDs are set at 44.1 kHz s/r so they can carry a 20 kHz audio signal, generally accepted as the minimum requirement for high fidelity.
- Computer media often use a 22 kHz s/r to save data. This implies an upper audio limit of 10 kHz, about the same as AM radio.
- Analog TV and FM stereo sound are cut off at 15 kHz; the transmitter uses higher frequencies for other purposes. So broadcasters often set a 32 kHz s/r in their storage systems and satellite links. These systems might be part of an HD radio or TV chain as well, even though the transmitters use 44.1 kHz (HD radio) and 48 kHz (HDTV).
- U.S. telephones traditionally cut off at 3.5 kHz. So when switching networks and long-distance lines were converted to digital, a sample rate of 8 kHz was chosen.
- Our example in Figures 1.15 and 1.16 has a second harmonic at 800 Hz. This puts it safely below the Nyquist Limit of our 2 kHz sampling. That's why Track 5 can accurately play the harmonic.

Higher is not necessarily better. Professional video is standardized at 48 kHz sampling. But most equipment that handles 48 kHz s/r also supports 44.1 kHz s/r and uses the same filters for both. That means there's no advantage to the higher rate. Some professional equipment uses 96 kHz sampling, but very few engineers believe we can hear as high as 40 kHz, and it's certainly not necessary for dialog. The higher rate was chosen to keep filter artifacts—which increase as you reach their cutoff frequency—far from critical audio.

DIGITAL VS. ANALOG

Many music producers like the distortion created by overloaded analog tape, and will dub their digital masters through vintage studio recorders just to get it. Then they'll redigitize for CD release. But that's actually adding distortion, not compensating for any limitations of digital recording.

Arguments about the superiority of one medium over the other can be intense as Presidential elections. Some people insist that audio hit its peak with vinyl records and has gone downhill since; they point to awful mp3s on the Internet as proof. Others assert that digital is inherently "steppy" or can't reproduce certain shapes of waves. Those last two assertions are absolutely true. But they ignore how analog circuits suffer from problems that are exactly equivalent.[16]

---

[16] Want proof? See my book *Producing Great Sound,* or my Web site www.dplay.com/gotcha.

It's not worth fighting about. If you're doing any serious soundtrack production these days, digital audio will be part of your life. What's important—particularly if your background is analog—is understanding some fundamental differences between the two systems.

Analog Handles Overloads More Gracefully

When an analog mixer, processor, or recorder receives a signal that's too loud, it starts to distort. The distortion is gentle at first and increases as the input gets louder. If an unexpected signal peak causes a sudden overload, the distortion may be tolerable.

That's not true for digital. Anything louder than full scale—even for a single sample—gets badly distorted. This often shows as a crackling that obscures the signal. And it doesn't have to happen during the original recording: if a signal is boosted too much inside a digital processor or while you're mixing in an NLE, the distortion is the same. In some systems, you can't even spot the distortion until after things are rendered. The only cure is to watch all your levels carefully. If the software doesn't allow that, listen to the final product and be prepared to redo if there are problems.

Analog Deteriorates with Each Error; Digital Conceals Errors, then Fails Totally

In a digital system, an error is missing or inaccurate data. In an analog system, it's usually a burst of noise. Neither medium is particularly prone to errors, but either can have them because of malfunctioning equipment, dirty tapes, noisy environments, and other problems.

It sometimes seems like digital has more things to go wrong. Problems like unsynchronized clocks (Chapter 3) can cause clicking or hiccups that are hard to trace, and there's no equivalent error in the analog realm. But the fact is, analog has its own unique array of problems. We've simply been dealing with them longer.

Digital systems have two kinds of safeguards against problems. Small errors can be corrected perfectly, with mathematical techniques. Larger ones are detected but can't be corrected; instead, missing parts of the waveform are interpolated from the surrounding good data. Usually, this error concealment is undetectable. Very large sections of bad data cause the system to either produce noise or mute automatically. In practice, if a digital signal is getting progressively worse, you can keep hearing clean audio through much of the damage and not even know there's a problem. Then it disappears entirely.

Analog systems have no such safeguards. Once noise mixes with the signal, it's there to stay. But the signal's there, too, and you'll probably still hear it. In practice, as analog signals get progressively worse, you hear the changes and know something's wrong.

Analog Degrades with Each Process; Digital Doesn't Have To

Even the best analog equipment adds a tiny amount of noise and distortion. Knowledgeable engineers strive for as simple an analog signal chain as possible, and avoid extra tape generations except for special effects. But good digital recorders don't degrade digital input signals at all, and good digital processors do so little damage that you'd have to string together hundreds before you'd notice anything.

However, you still have to be knowledgeable if you want digital good sound. Processors can be misused. Each conversion from digital to analog and back adds some generation loss. Avoiding these problems isn't hard, once you're armed with some basic understanding and the proper techniques. This chapter's been about the understanding. The rest of this book is about the practicalities.