Audio Anecdotes III

Tools, Tips, and Techniques for Digital Audio

Edited by
Ken Greenebaum
Ronen Barzel

Audio Anecdotes III

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Audio Anecdotes III

Tools, Tips, and Techniques for Digital Audio

Edited by

Ken Greenebaum

Ronen Barzel

A K Peters
Wellesley, Massachusetts

Editorial, Sales, and Customer Service Office

A K Peters, Ltd.
888 Worcester St., Suite 230
Wellesley, MA 02482
www.akpeters.com

Copyright © 2007 by A K Peters, Ltd.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

Library of Congress Cataloging-in-Publication Data

Audio anecdotes : tools, tips, and techniques for digital audio / edited by Ken Greenebaum, Ronen Barzel.
p. cm.
Includes bibliographical references.
ISBN 13: 978-1-56881-215-1 (vol. 3)
ISBN 10: 1-56881-215-9 (vol. 3)
1. Sound--Recording and reproducing--Digital techniques. I. Greenebaum, Ken, 1966- II. Barzel, Ronen.

TK7881.4.A93 2003
621.389'3--dc22

2003057398

Cover art: Pablo Picasso (1881–1973), *Guitar on a Pedestal*.
Copyright © 2006 Estate of Pablo Picasso / Artists Rights Society (ARS), New York.

Cover design by Darren Wotherspoon.

Printed in India
11 10 09 08 07 10 9 8 7 6 5 4 3 2 1

Contents

Preface

This third volume completes the collection of *Audio Anecdotes* that was begun in 2004 and that has involved many contributors and advisors. The extensive and positive feedback from the digital audio community has convinced me that our combined efforts have been well worthwhile and that we have achieved our goal of helping researchers and practitioners to be more effective in developing and integrating digital audio solutions. I hope that, in a small way, we have been able to raise the bar for digital media quality and capabilities.

Amazing strides have been taken since the time this series was conceived (with much progress even in the relatively short period of time since the second volume was published). It was during this period that digital media technologies have come into their own and become mainstream instead of being limited to deep-pocketed professionals, dedicated early-adopting enthusiasts, or the curious wealthy.

The following are only some examples of the technologies that have spawned exciting new products and sometimes entire industries.

Voice over IP (VoIP) has been a technology buzzword for a long while, but with the growing popularity of Skype, it has become not only mainstream but also a valuable enterprise, as Skype's recent multibillion-dollar acquisition impressively underscores. While not yet mainstream, IP-based PBXs, such as the open-source Asterisk, have already begun outselling conventional PBXs and may just become mainstream in the near future.

Personal digital media players have become a common accessory, lead by the phenomenal success of the iPod, and they continue to evolve, offering video capabilities and integrating with other equipment, such as car stereos.

Legal, convenient, digital audio content download and streaming services are rapidly gaining in popularity and are already outselling all but the largest brick-and-mortar outlets for music. Purchase and subscription models are being experimented with in the laboratory of the marketplace.

Digital video downloads of television and movie content are now available from services such as CinemaNow, MovieFlix, and iTunes. Services such as these will forever change how media is distributed and may just eliminate physical distribution entirely.

The verb *to TiVo* has entered our lexicon with PVR *time-shifting* capabilities rapidly becoming a ubiquitous offering for hard-disk video recorders, cable television boxes, and media PCs.

Perhaps the most exciting phenomena has been the democratization of media creation and distribution.

The word *podcast* has been added to the venerable *OED*, and the *New Oxford American Dictionary* declared it to be "the 2005 word of the year." Perhaps, this record-time for a neologism to be added to the self-proclaimed "definitive record of the English language" underscores individuals' desire to share their thoughts whether written (blog), spoken (podcast), or performed (video blog, video podcast).

Video production and distribution are no longer limited to those governments, individuals, or corporations who can afford studio time and television broadcast rates. Do-it-yourself video is not only possible but has never been easier or more affordable with now inexpensive digital video cameras, the ubiquitous high-speed digital video interfaces such as USB2.0 and 1394 on computers, and even non-linear digital video editing applications available inexpensively or even bundled free with many new computers. Even the bandwidth costs associated with distributing video over the internet have been defrayed by technologies such as *bit torrent*, which uses networks of interested parties to redistribute already downloaded content, in effect scaling content servers with the popularity of the media being (re)distributed.

I continue to be amazed by the wonderful group of people who have helped take *Audio Anecdotes* the concept and make it a successful reality. Most of the people who made these books possible are explicitly recognized as contributors, without whose dedication and patience *Audio Anecdotes* would not have been possible. I want to take a moment to thank others who have made major contributions.

Thanks to Alice, Klaus, and the rest of the wonderful A K Peters publishing family.

Special thanks to those who volunteered for the project: Howard Good (our build and CD meister), Eric Lee (for his help and enthusi-

asm), Robert Quattlebaum (for his desire of excellence and Mac OS X expertise), John Nordlinger (for his strategic vision), and Michelle Steinberger (for her constant support even through the late nights and lost weekends).

Finally, thanks to all my dear friends, colleagues, and family, who have contributed ideas, read early drafts, leveled with me when I failed to communicate, and otherwise helped and encouraged me through this long project.

Ken Greenebaum, Cupertino, June 2007

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Introduction

Welcome to *Audio Anecdotes III*! Those of you already familiar with the series will find a variety of exciting new content in a familiar format. Those of you new to *Audio Anecdotes* should be able to dive right in and quickly find yourself at home; however, for the best experience please have the earlier volumes available since articles often build on and reference articles from earlier volumes.

Continuing the series, the articles in this volume also explore creating, recording, processing, and analyzing many forms of sound and music. While the book discusses all manner of audio phenomena, it emphasizes techniques for digitally representing, manipulating, and processing media. These digital-media techniques present profound opportunities over former analog methods and are increasingly enabled by the seemingly endless progression toward ever more powerful, less expensive, and universally available digital computation.

To ensure the broadest impact of the material presented, the accompanying CD-ROM provides audio and video files, interactive demos, and cross-platform open-sourced computer source code. We go to this effort to help the reader not only understand the material but also hopefully incorporate it into existing projects, and ultimately to help stimulate the creation of new products and services.

As the near ubiquity of audio-centric consumer electronic devices like the cell phone and, more recently, the portable digital audio player remind us, it is still far too easy to forget in our visually focused society just how fundamental our sense of hearing is. Perhaps we in this society forget or undervalue sound both because hearing is such a subtle sense that we

often don't recognize its effect on us and because most of us no longer rely on it to hunt or save us from becoming prey.

Yet, sound's influence is profound, and it influences us in a powerfully primordial way. Consider that the sound of our mother's beating heart may be the first sound of which we are aware, and no matter what our age, resting our head on a loved one's chest and listening to their heartbeat may be the safest and most comforting place we know. Or, consider how music can seemingly affect our emotions directly, whether we try to resist or not. Music can quickly stir our hearts and senses. Finally, it is both fascinating and enlightening to consider that most of the world's religions teach that creation began with a sound; whether that sound be an utterance or the universal Om.

In *Audio Anecdotes* we explore both sound and our sense of hearing; the one sense which never sleeps and works omni-directionally across vast distances. *Audio Anecdotes* attempts to present opportunities to improve the audio experience where sound already exists or to encourage the integration of sound into presently mute applications, leading to richer, more expressive, more engaging, and ultimately more valuable applications.

Structure

Each *Audio Anecdotes* volume is composed of articles that cover a wide range of audio-related topics. Written by experts, the articles' topics span the breadth of sound- and music-related fields and take a number of forms. The types of articles include topic introductions, essays, in-depth technical exploration of algorithms, and practical presentation of tools and techniques.

Audio Anecdotes is different than other books. We encourage our authors to write using their individual voices and many articles contain the authors' personal anecdotes and hard-earned experience from having worked in the trenches. We therefore encourage readers to consult the biography section at the end of the book to learn a little about an author's background before diving into their article.

Our articles explore deep topics that individually could fill entire books. Consequently, the articles are designed to act as *jumping-off points* for readers to discover new topics, receive motivation as to why a technique or algorithm might be appropriate, facilitate experimentation via interactive demos, provide explanation with a bit of background, and finally, point to other references to further explore the topic. Each article contains an annotated list of references that serve not so much to docu-

ment the sources of the article, but to direct readers to significant texts and further sources of information on the topic area. Where possible, articles reference other articles in this or other *Audio Anecdotes* volumes. We introduce a topic in an early *Audio Anecdotes* volume and then return in a future volume to cover the same topic in a deeper or more abstract way.

Articles in each *Audio Anecdotes* volume are grouped into chapters by topics organized along an arc spanning the following topic areas:

- Fundamentals: the physics, measurement, and human perception of sound
- Recording and playback of sound: whether of music, voice, or nature
- Synthesis: rendering sounds including the synthesis of musical instruments, voice, or sound-effect (Foley Sound)
- Signal processing: the mathematical analysis and manipulation of sound
- Signal processing applications: from compression techniques to signal detection and recognition
- Computer techniques: efficiently implementing robust, low-latency, precisely synchronized audio systems
- Music theory: the mathematics of both western and non-western music
- Creative topics: music composition and sound design
- Nature, mind, and body: how sound exists in nature and affects the mind and body

The motivation for this topic arc is rooted in the belief that to understand any topic, or to be able to make informed engineering trade-offs in design or optimization, a solid understanding of the physics and human perception of the phenomena is required. Great engineering accomplishments, such as the design of the telephone system, color television, and digitally compressed media such as the DVD, all demonstrate a mastery of the interplay between physics and human perception. From the fundamentals, the arc extends to the abstract through the applied and creative, to again revisit human perception from a different perspective.

While each *Audio Anecdotes* volume can't include articles covering every topic area, the articles are organized according to this arc. *Audio Anecdotes III* contains the chapters described below.

Chapter 1. Recording Music

This chapter greatly expands the basic audio recording articles found in the first two *Audio Anecdotes* volumes.

We begin with a pair of articles that provide music recording industry insiders' perspectives on recording techniques. The first article explores the analog multi-track era, providing many examples from classic contemporary recordings, many of which the author participated in. The second article describes the current situation where most recording is performed digitally and is edited using easy-to-use, nondestructive, nonlinear editors. Not surprisingly, it takes time to refine new tools and to create appropriate methodologies. This article discusses some of the strengths and perils of using these new digital tools.

The next article provides an incredibly intuitive introduction to the artistry of sound mixing. The author explains mixing from an almost painterly perspective instead of the classic clinical perspective of level setting and equalization. The article describes a large number of techniques and provides audio versions of these on the accompanying CD-ROM.

Audio Anecdotes last explored audio scene analysis in the first volume. The final article in this chapter applies the principles of audio scene analysis to audio mixing, providing a theoretical framework to understand how audio recording and mixing techniques are perceived by the listener. This article may also provide insight into anecdotal techniques and rules of thumb.

Chapter 2. Sound Synthesis

In this chapter we return once again to explore sound synthesis, a topic featured in every *Audio Anecdotes* volume.

This chapter begins with a detailed introduction to granular synthesis, a technique that uses many simultaneous instances of sound, called seeds, to create results that are very difficult to emulate using other synthesis techniques. This technique is particularly effective at creating the sounds of nature such as the roar of waterfalls or the crash of waves. However, other sounds directly lend themselves to this technique such as synthesizing the sound of the maraca, a traditional instrument usually consisting of a hollow guard filled with seed and shaken. Code and examples of these and more exotic sounds are included on the CD-ROM.

The final two articles build on the second volume's introduction to physical modeling. The first article adds a model of string excitation due

to *stiction* to the physically modeled traveling wave equation presented in the previous volume. This is needed to emulate the effect of rosin applied to the bow sticking, melting, and solidifying in rapid succession in the bow-string interface that allows a violinist to so beautifully excite their instrument's strings. Code and examples of a bowed string violin synthesizer are provided on the CD-ROM.

The second article of the pair presents modal synthesis, a highly efficient alternative to physical modeling complex resonant models. The technique employs a bank of resonators tuned to emulate the measured resonances of actual objects. The result is a realistic, high fidelity model of the actual object that can be excited to simulate the effect of striking or scraping the modeled object. Code and examples are provided on the CD-ROM.

Chapter 3. Voice Synthesis

This chapter introduces the topic of voice synthesis to *Audio Anecdotes*. Voice synthesis has been possible for a long time; however, it has become more mainstream as the trend toward designing computers and other devices to interact directly with people on human terms accelerates. Increasingly specialized, computer-centric interfaces, such as computer keyboards and video displays, are being displaced by other, more natural devices and modes of interaction whenever possible. An example, with which most of us have interacted, is a telephone-based banking, reservation, or other system that employs speech synthesis and, more recently, voice recognition.

While devices have employed the playback of voice since before the invention of digital computers, these devices have traditionally been limited to repeating the same pre-recorded phrases or sentences and consequently could only provide content variation by recording every anticipated permutation of the message. We begin this chapter with a practical article describing how to record and process segments of speech so that they may later be dynamically stitched back together in real-time to form full sentences with variable detail. Without a methodology such as the one introduced in this article, it can be very difficult to form natural sounding sentences.

The next article provides an unusual perspective on voice synthesis. While developers have traditionally focused on improving the clarity and intelligibility of their voice synthesis systems, this author's research involves creating realistic voice especially as an artificial replacement voice

for people who are loosing their own ability to speak. Since our unique voice is a large part of our individual identities, the ultimate goal of such synthesis is to recreate the unique qualities of an individual's natural voice. This article provides an introduction to the subject by surveying the history of speech synthesis techniques.

Phonemes are the distinct units of sound that make up the spoken word. Surprisingly, every language uses an overlapping but distinct subset of the sounds a human vocal tract can produce. While modern speech synthesis algorithms don't string together recordings of phoneme sounds to form words (this produces crude and mechanical sounding speech), it is very helpful to be familiar with phonemes. This chapter's final contribution is a table of the phonemes found in the English language.

Chapter 4. Speech Processing

This chapter continues Chapter 3's theme by considering the processing of speech. The human vocal tract is a very specialized instrument that creates a very unique signal. Consequently, conventional sound processing can fail horribly when applied to speech, or at least not perform as well as those designed specifically for speech. Speech processing requires a solid understanding of both the physics and perception of voice.

The first article begins by providing an introduction to the physics, physiology, and acoustics of speech. This article describes the vocal tract and identifies the distinct features unique to voice based on the physics of the human vocal tract. It is specifically these features that the next two articles exploit.

The remaining pair of articles explores the manipulation of the speech signal. The first articles describes a method to speed up or slow down the rate of speech without significantly affecting it's intelligibility or its perceived pitch. Such a system is useful whenever we might want to slow speech down (when transcribing), or speed voice up (when reviewing recorded notes, or catching up on a missed television episode in a fraction of the originally aired time).

The second article of the pair describes, conversely, how to use related techniques to change the apparent pitch of the speech without affecting the playing time. This technique could be useful for correcting the pitch of voice or song to make it *on key* or to simply alter the quality of the speaker's voice.

Chapter 5. Applied Signal Processing

In this chapter we return to the subject of signal processing, this time exploring applications.

The first article introduces audio dynamic range compression. Not to be confused with data compression (reduces the encoded size of data), dynamic range compression (and its opposite, expansion) changes the relative amplitude of the loudest and quietest portions of a signal. Dynamic range compression is heavily used in the music recording and broadcast industries for both creative and practical purposes and has many applications in other fields. This article deeply explores compression theory, application, and challenges. MATLAB models are provided.

An important component of speech communication systems is the speech, or energy, detector. For instance, a speech detector enables a system to be able to discriminate between valuable speech and unwanted noise. Such an equipped system could disable its transmitter when speech is not present thus increasing battery life. Bi-directional speech detectors are used to eliminate feedback in speakerphone systems when only one party is speaking. The next article presents a simple and computationally inexpensive speech activity detector translated from its original 8-bit assembly language implementation.

The last article expands on the signal detection theory articles from the first volume by providing an introduction to sound classification. The implementation for an actual sound classifier that can distinguish between the sounds of different propeller aircraft is presented to help illustrate these concepts. The working classifier with example sounds is provided on the accompanying CD-ROM.

Chapter 6. HRTF Spatialization

Audio Anecdotes II included a chapter on multiple-speaker spatialization techniques as well as articles on binaural sound. In this chapter we return to the topic of sound spatialization by introducing the head-related transfer function (HRTF). The HRTF models the spectral filtering caused by the human torso and pinna (outer ear). The characteristics of this filter are highly dependent on the sound's incident angle to the head, which allows us to sense the sound source's position in space. While most sounds heard in the environment are processed in this way, it has also become possible to synthetically process sound to control spatial characteristics.

We begin the chapter with an article describing the nature of true three-dimensional sound, why it is best listened to on headphones, and how the spatial cues break down when listened to on stereo loudspeakers. Be sure to listen (with headphones!) to the collection of binaural recordings the author recorded himself, using a special in-ear microphone apparatus, on the accompanying CD-ROM.

The next article describes a novel application for spatialized sound; a dance club. This article demonstrates that spatialized sound is not limited to computer graphic applications or headphone wearers.

The final article provides a mathematical model for deriving and understanding the HRTF from first principles. The author provides a mathematical derivation of the HRTF angle-dependent spectral filtering based on a simplistic model of the head (resembling a bowling ball). This analysis is unusual since the HRTF is usually not synthesized but rather is constructed based on measuring in-ear microphones' spectral response to a movable sound source.

Chapter 7. Synchronization

A major goal of *Audio Anecdotes* is to share algorithms, techniques, and actual code to help individuals build robust applications that combine sound and other media. This chapter extends the audio-sample *plumbing* articles presented in previous *Audio Anecdotes* volumes to address the challenging subject of synchronizing digital media.

The first article attempts to provide a solid introduction to the often-confusing subject of synchronization. To do this it defines the major terms, provides motivation for solving the problem, and describes some of the challenges and strategies for solution, including a variety of approaches.

The next article adds a historical context for synchronization by describing the twenty year-long technical struggle to marry a synchronized soundtrack to the then silent motion picture. This effort finally yielded the *talkie* that we today recognize as the modern movie. This is a colorful story from which we can still learn today as we attempt to perfect multimedia synchronization on modern computers.

The final two articles describe in detail the two major components of synchronization first presented in the introductory article: *start synchronization* (the process of ensuring that media streams are begun at individually appropriate times to ensure that they are synchronously presented at a designated time in the future) and *dynamic synchronization*

(the process of constantly comparing the relative positions of multiple streams in a presentation, measuring their drift from ideal, then dynamically adjusting their rates to keep the streams approximately synchronized). Pre-roll, resampling, and control theory are among the subjects introduced. Code examples are provided on the CDROM to use and experiment with.

Chapter 8. Music Composition

Audio Anecdotes II included a chapter introducing music theory. This chapter builds on that base to provide two rather different articles on composition.

Music has generally fit into two categories: the composed repertoire (consider classical music) and improvisational (consider jazz). While many live performances may blur these distinctions, it is certainly true that all recorded music is presented virtually identically every time it is played. The recent application of computer technology to the performance of music has begun to change the static, linear, repeatable nature of recorded music. Unlike in a movie where scenes are a set length and order, in an interactive video game the user may take a different length of time to complete a goal, and increasingly video-games are being designed to allow more free-form exploration that allows the user to direct the order of encounters or scenes. The first article describes special considerations for composing music to accompany such non-linear experiences as video games strive to become more and more cinematic in both quality and presentation.

The second article expands on the description of musical meter from the second volume by introducing and exploring polyrhythm, a topic that can be alien to the western mind and ear. The article is designed to be interactive, suggesting exercises that the reader can perform to experience polyrhythm themselves. Further, this article provides a perspective from ethnomusicology on rhythm and polyrhythm, suggesting how different cultures came to develop very distinct rhythmic structures.

Chapter 9. Human Experience

Every *Audio Anecdotes* volume closes with a chapter that returns to the human experience of sound.

As scientists, we attempt to understand the physics of phenomena (like the creation, propagation, and perception of sound described in *Audio*

Anecdotes I). As engineers, we attempt to make calculated tradeoffs to create reliable, efficient devices that produce output perceived to be as high quality as possible, using technologies that are currently available (topics *Audio Anecdotes* returns to again and again). However, as artists, we have a different set of goals that are more difficult to describe, and we constantly change the rules. The first article suggests how video-game sound engines may be modified to produce a less technically accurate but much more engaging sonic experience for the user. This manipulation mimics the inaccurate and often completely unrealistic sound commonly employed in film and theatre.

Audio Anecdotes I extensively explored the human perception of sound and the mathematics used to measure sound based on human perception. Our second article continues this theme by introducing the fundamentals of psychophysics: detection, discrimination, scaling, and identification. The article then uses these principles to explore the psychoacoustic implications for quantization in the digital audio playback. For instance, does the 16-bit quantization of sound (as used in the CD Red-book standard) exceed human perception? If 16 bits are insufficient, then what would be the optimal number of bits of quantization before the human ear could no longer recognize an improvement?

Finally, we close the chapter and this book with an essay examining the history of technology and the human condition that ponders the future of man-machine interaction. The author suggests that traditional computing environments have reached "a crisis of complexity" that requires novel approaches to solve. Future machines will need to fully engage all human senses including hearing (under appreciated in our culture for historical reasons). Hence, the inevitability of the "audiocomputer."

Glossary, Contributor Biographies, and Index

Following the main chapters are an extensive glossary (defining many of the audio terms used throughout the book), contributor biographies, and an index.

CD-ROM

Audio Anecdotes III is accompanied by a CD-ROM containing materials intended to supplement the articles: audio files, video files, and executable demo programs including C-language source code. Demos support the Mi-

crosoft Windows, Apple OS X, and Linux platforms. Wherever possible, articles reference these materials so that readers can immediately listen to examples and experiment with the concepts introduced in the articles. Please be sure to explore the CD-ROM's contents, via the HTML-based tour, since materials are constantly being added to the CD-ROM and may not be explicitly mentioned in the text. In addition to the executable formats mentioned above, programs are also distributed as C-language source code with a makefile-based build system to facilitate experimentation and to allow code to be easily incorporated into the reader's own projects.

This material is distributed on the CD-ROM as tar balls (*tar*'ed compressed archives). A wizard-based installer is provided for automatic installation on our supported platforms. README files provide installation information if the installation wizard doesn't automatically start upon CD-ROM insertion. Once installed on your computer, the demo material is organized by chapter and author.

Unless otherwise specified, the contents of the CD-ROM are protected by the BSD license, and the reader may use the source code provided on the CD-ROM for any purpose as long as the following statement is prominently displayed: This product includes code from *Audio Anecdotes III*, edited by Ken Greenebaum and Ronen Barzel, published by A K Peters, 2006. The code is to be used at your own risk: Ken Greenebaum, Ronen Barzel, and A K Peters make no claim regarding the suitability of this material for any use.

A Note on Patents and Trade Secrets

Our authors have certified that their articles do not contain trade secrets. In some articles, authors have explicitly stated that the algorithms that they describe are patented. However, even algorithms that lack such statements may be under some form of patent protection. For many reasons, including the long gestation of patent applications (so-called submarine patents), we cannot vouch for the suitability of using these algorithms for any use other than educational purposes.

Please Participate

Visit us at the *Audio Anecdotes* website (http://www.audioanecdotes.com) to find errata, download code updates, or find out what's new.

Audio Anecdotes was created as a forum to share tools and techniques with the greater audio community. The subjects covered in this volume only scratch the surface of topics that we would like to cover. If you have been inspired by *Audio Anecdotes*, we encourage you to share your own tools, techniques, or experiences. If you find an error in the text or code, or have a code improvement, please send it to errata@audioanecdotes.com.

A Final Thought

We wanted to create books that would be fun to leaf through or read cover-to-cover, books that would be useful both as a reference and a source of creative inspiration. We hope that we have succeeded!

Recording Music

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

How Recordings Are Made I: Analog and Digital Tape-Based Recording

Daniel J. Levitin

Most modern movie-goers and television watchers are aware of the various forms of "trickery" involved in bringing scenes to cinematic life. We take for granted that there were probably multiple takes; that the dialog might have been dubbed in later to fix poor location recording; or that the sound of a blender mixing up a smoothie or a pistol being fired were added later in a sound effects suite. What most people don't realize is that this same level of sophisticated production is found in most modern audio recordings. The techniques used in music recording are fascinating in their own right, and they can enhance one's appreciation of the final product. See also Rogers' article "The Art and Craft of Song Mixing" (page 29) later in this chapter for a discussion of how such techniques are used to artistic effect.

I'll start this article by providing some background about the "traditional" hardware that is available in the recording studio. (Until recently, I would have called this the "modern" hardware, but the development of digital hard-disk-based recording is changing studio hardware, as discussed in the next article. Still, the traditional principles and techniques described in this chapter carry forward into that world.)

1 Multitrack Recording

Most popular music (rock, country, alternative) CDs use *multitrack* recording, in which different instruments (or different parts of an instrument) are recorded on distinct, separate regions of recording tape or a computer's hard disk. The most common systems use 24-tracks. In tape-based recording, several of these machines can be linked together to create 48-track and 72-tracks. In virtual or disk-based recording, additional tracks are subject to the number of *buses* available, the disk access speed, and the memory limitations of the computer.

If this concept of multiple tracks is new to you, consider your stereo cassette player or CD player. These have two tracks known as left and right, that is, two independent channels of audio information. The information on one track is processed using completely separate electronics from the other track, and this is why you are able to hear separate information coming from your two stereo speakers. (If you have more than two speakers, in a surround arrangement, the information coming from the third through *n*th speakers used to be extracted artificially from the two stereo tracks, and was not created in the original recording session. True multichannel audio recordings are just beginning to be commercially released on DVD-audio and SACD). Now, by convention, what we hear coming from the two speakers are parts of the same song and they are time-locked (synchronized) so we can listen to both tracks together and they make sense. But this does not have to be so; I have a CD of Leonard Bernstein discussing Beethoven's "5th Symphony," in English on the left channel (one of the stereo tracks) and in German on the right channel (the other stereo track). Using the balance knob on my amplifier, I can choose to listen to only one of the tracks or both. Theoretically, record companies could manufacture CDs with two mono tracks in parallel, of different performances, and you would get twice as much music on one CD. So for example, on older recordings of Duke Ellington's Orchestra (made before there was stereo), you could have two Ellington albums on one CD–you'd just have to set the balance knob so that you wouldn't hear the cacophony that would be created by playing back both at the same time.

Now, extend the concept of two tracks to a multitrack tape recording system which has 24, 32, or 48 independent tracks. The output of each of these tracks feeds a separate preamplifier built into a mixing console in the studio, or a virtual console on your computer monitor. Instead of having a balance control with 48 positions (awkward, to say the least), a recording engineer can decide which of the tracks to play by adjusting a

separate volume control for each, or turning each track on and off with a switch (called the *mute* button). It is important to understand that these 24 (or however many) tracks are both time-locked and distinct. They can be recorded or played back one at a time or in any combination, without interfering with each other. This simple fact enables a number of interesting recording techniques.

First, the musicians don't all have to perform their parts at the same time. If a band decides to add a saxophone solo after they've finished recording a song, the sax player just adds her part to an empty track. It doesn't disturb parts that were already recorded. Conversely, if the group decides that they don't want to use a guitar solo they had recorded earlier, they just don't turn that track on (they can even erase it) and the rest of the parts remain undisturbed. Many groups exploit this feature of multitrack recording and add all kinds of parts just to see what they sound like–background vocals, horns, strings, and so on–and let the producer or mixing engineer decide later what to keep and what to throw out. The mixing engineer is the engineer who combines all the tracks into a two-channel "mix," and decides how to allocate the various instruments to the left-right stereo soundfield.

Second, a given musician can play more than one instrument, and listen back to the previously recorded instruments while he is doing so to provide a reference. The guitarist and inventor Les Paul was the first to employ this technique, and Stevie Wonder, Prince, and The Beatles have all used it to great effect.

A third advantage of separate, multiple tracks is that each track can be modified or processed *individually* without affecting other tracks. Signal processing devices, such as compressors, expanders, tonal equalizers, noise gates, digital reverberation simulators, and digital delays can be applied to any one or multiple tracks, and they can be applied after the sound was recorded. Most high-end recording consoles and digital audio workstations have built-in parametric equalizers (EQ) on every track, allowing the engineer a wide range of tonal control over every track. For example, suppose that an electric guitar, electric bass, and acoustic guitar are recorded on three separate tracks. Maybe the electric guitar sounds too shrill, the bass sounds too muddy, and the acoustic guitar sounds too dark. Any time during the recording process, the engineer can modify these sounds by applying EQ to them individually. Multiple signal processing devices can be chained, so in this case, the engineer might EQ the bass to make it less muddy, run it through a noise gate to get rid of hum that was present in the background of the studio that day, then run it through a compressor (to even out the overall volume of the

performances), and finally, another stage of EQ. This specific scenario is actually not all that uncommon.

2 The Basic Tracks

The typical way that rock and country music are produced is to record the rhythm section first–usually the drums, bass guitar, and maybe a rhythm guitar. At this time, the vocalist records a *scratch* vocal—a temporary vocal track just to help the rhythm players keep track of where they are in the song. The vocalist typically doesn't give it his all at this stage and the engineer doesn't always bother to set up a particularly good microphone, because the plan is to replace this vocal (overdub it) later with a better performance. You can often find a lot of joking around on these scratch vocal tracks.

John Lennon was working on a new album in 1980 which eventually became *Milk and Honey.* He had recorded scratch vocals to accompany the musicians' basic tracks, but he was killed before any final vocals were recorded. The vocals you hear on the version of the album that has been released were what Lennon had intended only as temporary vocals, and so they contain a certain degree of casualness—and an absence of full voice singing—that would not normally be found on a final vocal.

The various instruments used in the rest of the piece are usually added one at a time. Musicians adding a new part can listen to any combination of the instruments already recorded, in any volume mix that they choose. A rhythm guitarist might want to hear lots of bass and drums so he can keep time; a lead guitarist might want to hear lots of keyboards so he can hear the chord changes better.

This is the norm in popular and country recording. Traditional jazz, classical, bluegrass, and folk have followed a different tradition. In these genres, the musical communication between players is considered an essential part of the performance, and they would never consider playing separately from one another. Neil Young is an example of a rock artist who tends to favor *live* recordings with minimal overdubs, but he is an exception in the rock world. One of the issues here is purely technical: To create a clean rock recording with loud electric guitars is difficult to do when the guitar amps, the drums, and the vocalist are all playing in the same room at the same time, because the sound of the instruments leaks into the microphones of the other instruments, creating a muddy sound. If you care to, listen to *Led Zeppelin III* and *Houses of the Holy* to hear the radical difference in recording quality as the group moved from live

recording to an overdub approach, the latter of which allowed for sonic isolation between the instruments and the attendant improvements in sound quality.

There is also a movement, at the vanguard of audio engineering, to use as little audio processing as possible. These engineers often boast on album covers that they have used no EQ, no digital reverberation, etc. The results can sound stunningly lifelike, but pulling this off requires a great sounding musician to begin with, and a great deal of skill on the part of the engineer. One famous example of an album with no equalization is Steely Dan's *Countdown to Ecstasy*, recorded by Grammy-award winning Roger Nichols. To record an entire album without any outboard effects is a challenge, but it does not guarantee a superior product. Some of the best engineers in the world—Roger Nichols, Bruce Swedien, and George Massenburg, for example—use outboard signal processing devices judiciously to create beautiful recordings, and in many cases, to create interesting *hyperrealities*.

3 Soundscape

3.1 Illusions of Perspective: Realism versus Hyperrealism

One of the most interesting aspects of cinematography is that we are able to see on the movie screen things that we could never see in real life. A classic example of this is the movie chase scene. In the theater, we can see the pavement speeding by from a camera mounted on the door of the car, or we can see the road ahead from a camera mounted on the front bumper. In a sense, these are very *unrealistic* vantage points—we rarely are able to put our eyeballs in these positions. An even more startling example of an impossibility is when the director cuts from one of these cameras to another, allowing you to see two very different perspectives in rapid succession. What the director and cinematographer are conveying is an intentionally unrealistic view of the world; they are providing a set of impossible perspectives in order to provide excitement and a sort of *hyper*realism. Please see Bailey's article "Spatial Emphasis of Game Audio" (page 399), where such techniques are applied to video games to create hyperrealistic cinematic experiences.

Of course, chase scenes aren't the only use of techniques that create unreal perspectives. Even simple *head shots* of someone talking give the illusion that your eye is only three inches from the person's face, revealing pores and details most of us never see. Modern recording also uses technology to create hyperrealities.

3.2 Microphone Placement

One common technique is based on a simple concept—microphone placement. For example, when recording an acoustic guitar, the engineer might use two microphones, one at each end of the guitar, and record these onto two separate tracks. During mixing, one of these tracks is assigned to the left stereo field, and the other to the right stereo field. If you listen back at home and your speakers are eight feet apart, it sounds like the guitar is eight feet wide! (It also sounds like your head is right in the middle of the guitar, which of course it couldn't be in real life, or the guitarist would be strumming your face.) In headphones, the illusion of your head being right inside the guitar is even more compelling because there is virtually no air between the transducers and your ear. The guitarist Alex deGrassi records his acoustic guitars using this technique, which is particularly evident on his albums *The World's Getting Loud* and *Slow Circle*.

Any instrument can be recorded in this way, known as *stereo mics split panned*. *Split panning* refers to the two mics being split in the stereo image, so that one is assigned completely to the left channel and the other is assigned completely to the right channel (the *pan pot* used for *panning* is an abbreviation for the control knob which is officially called a *panoramic potentiometer*). With only one mic, the instrument can be assigned to one speaker or the other, or to any arbitrary point between them. Only by rendering the signal with two mics, however, can the sound break free of *point source localization* and begin to take up more space in the stereo image, the ultimate being the illusion that the instrument is surrounding the listener. Grand pianos are often recorded this way, too, in popular, jazz, and classical music, because it gives the listener a sense of being enveloped in sound.

Other instruments lend themselves to different spatial effects. Drums are typically recorded with one microphone on each individual drum, and these are panned in a semicircular arc, emulating the sound that a drummer would hear sitting at the drums: the high-hat just to the left, the ride cymbal on the right, the snare and kick drums in the middle, and the tom-toms sweeping around the arc of a semicircle, from left to right. The sound we hear through our speakers and headphones, however, is typically much better than the drummer actually hears; because the mics are placed adjacent to each sound source, each percussive component conveys the sound it would if your ear were right up next to it. Stevie Wonder was one of the first to do this, working with engineers Malcolm Cecil and Bob Margouleff, on his album *Music of My Mind*.

The same is true with vocals—the engineer typically places a very sensitive microphone an inch or two in front of the singer. This makes it sound as though your ear is just in front of the singer's mouth. In ballads, this adds intimacy to the performance, especially when listening back in headphones; in heavy metal, it adds a great deal of power, and gives the vocals a presence that keeps them from being swallowed up by the other instruments in the mix. Again, in real life, our ears are never just two inches from the singer's mouth, but through recording we experience this illusion. For years, my favorite example of this was Paul McCartney's vocal on "Honey Pie" from the Beatle's *White Album*. The mic—probably a Telefunken M49—is so close to his mouth, you can actually hear his lips part just before he pronounces the "p" in the word "pie"; when he sings the word "crazy," you can hear the air moving as he sets his mouth to pronounce the "c." Recently, I found a recording that conveys this effect even better—Aimee Mann's vocals on "Jacob Marley's Chain," from her album *Whatever* (recorded with Neumann's version of the M49, a U49). She uses vocal dynamics artfully to create the illusion she is practically whispering the song in your ear. Mixing engineer Bob Clearmountain added a great deal of compression to the vocal to even out the dynamics, so that loud and soft passages appear to be at the same volume, even as Aimee goes from very soft to very loud. Now imagine listening to a group and all of the instruments have been recorded with the microphones right on top of them—this is called *close miking* and it is how most rock records are made. The listener experiences the ultimate in hyperrealistic perspective—hearing each instrument as though her ear was right up against it, all at the same time! This is equivalent to the rapid edits in a movie, except with albums, you, the listener, get to decide when to switch your attention from one instrument to another, or whether to take in the whole scene.

It is interesting to consider the cognitive differences between seeing and hearing. Because visual information is spread out across *space* and auditory information is spread out across *time*, the two sensory experiences are fundamentally different. When we shift attention from one visual stimulus to another, we have to move our eyes. To shift attention from one auditory stimulus to another, we don't move our ears; we simply focus our attention on a different aspect of the sound that is impinging on our eardrums. In a musical performance, we can concentrate on an individual instrument or on the whole (the *Gestalt*). In a visual performance, such as a movie, we can only have the equivalent degree of control if we are provided with multiple views—for example, if the director splits the image up into several parts. Note also that in a movie, the director and cine-

matographer often use an assortment of lighting and image-composition tricks to get you to look at exactly the part of the screen they want you to, whether that's focusing on the face of a character who's making some significant expression or looking off to the side in anticipation of a monster about to jump in. Audio engineers can accomplish some of these same framing effects by the proper use of signal processing: equalization that carves out a notch in frequency space for a particular instrument, for instance, or reverberation that places certain instruments at a specific depth in the overall auditory space of the recording.

3.3 Reverberation

In the old days, engineers would take the signal of Elvis Presley's vocal, play it through a speaker in the corner of a small, tiled room, and pick up the sound of the room reverberating with a microphone suspended from the ceiling. In recent years, the acoustic echo chamber has all but been replaced by digital reverberation simulators. Whereas the live echo chamber provided only one sound (adjusting parameters like reverb time required moving the mic around in the tiled room), the modern devices simulate dozens of spaces, such as a small tiled room or a large woodpaneled church. Because each instrument can be run through special effects separately, you can hear something else on albums you never hear in the real world, a band in which the snare drum sounds like it is inside a 50-gallon oil drum, a guitar that sounds like it's underwater, and a lead vocal that seems to be coming from the far end of the concert hall.

The various microphone and mixing techniques described earlier define the location of a sound in the left-right plane; reverberation defines a sound's location in depth. The three-dimensionality of recordings comes from the listener's impression that the various instruments occupy different places in depth as well as in the left-right stereo field. By applying different reverb programs to different instruments, the depth of a recording is greatly increased, giving the sense that each instrument occupies its own place in the sonic landscape.

Additional tricks can be applied to alter location in the sideways or $x - y$ plane. With clever manipulation of phasing, engineers can make it seem as though sound is coming from *beyond* (outside of) the stereo speakers (not just between them). Engineer Bruce Swedien experimented with these placements on Michael Jackson's *Bad* and *Dangerous* albums. Of course, if used indiscriminately, all these techniques can create a cheap, gimmicky sound, but if used properly, they can create excitement.

Pink Floyd, The Beatles, and Laurie Anderson pioneered the use of the studio as another musical instrument to enrich their artistic product, and this has now become commonplace.

4 Editing Parts

Multitrack recording brings with it another possibility—the ability to edit individual parts. Remember that in the typical case, a rock band might record the *basic tracks* of their song first—the drums, bass guitar, and rhythm guitar. Because each instrument can be recorded on its own track, it is a simple matter to repair any mistakes on a given track without altering the other tracks. If the producer and the band decide that a particular take has the right *feel*, they might decide to use it even if it contains some mistakes. If the bass player played some wrong notes, or her timing was off by a bit, it is simple to go back and fix *just those* notes. The engineer plays the tape back to the bass player and she plays along with the tape. When the tape gets to the part where the mistake occurred, the engineer hits the *record* button for the bass player's track only. Now the bass player's new performance is put on tape, erasing the old one, and the engineer can hit the *stop* button any time to stop recording and return to the part that was formerly recorded.

This technique is called *punching in.* It is simple to punch in and out of very tight spots—it is not unusual, for example, for a musician to try to repair a passage with only a 16th note space on either side of it. As long as an instrument was recorded on a separate track, and was isolated from the sound of other instruments during recording, it is difficult to tell a repair from the original. After spending three and a half minutes recording one take for a basic track, a group might spend hours making repairs to those basics.

Soloists and vocalists also routinely punch into a track to repair or improve performances. If a vocalist misses a high note, there is no need to redo the whole performance, obviously—he can just punch in and fix the troublesome phrase. If you listen carefully, you can actually hear where the punches are on Michael Jackson's vocals on some of the songs off *Thriller*, and on Crosby, Stills and Nash's song "Helplessly Hoping." You can hear the punches because they occurred while the singers were taking a breath and the punch interrupts the sound of them breathing in. Interestingly, a musician with only marginal technique can use punching in to make himself sound better than he really is, creating flawless performances that he would never be able to otherwise execute in real time.

(I am a marginal guitarist in real life, but on tape I sound pretty good, only because I, like many of my friends, used to spend six hours recording one eight-bar solo.)

Conceptually, punching in is equivalent to the old-time method of editing analog tape with a razor blade and splicing tape. The difference is that punching in only affects one or a few tracks at a time and editing usually involves cutting the entire piece of tape and splicing it to a new one. Symphony orchestras typically record an entire performance, and then go back and replay any sections that had mistakes. Later, an editing engineer splices in the fix. In traditional jazz, the combo might play several versions of the same song, but it would be an artistic scandal if two different takes were edited together; because jazz is primarily an improvisational form, each take is considered a completed and inviolable work.

Since jazz and classical sessions are generally recorded without overdubs, you might think they don't need all these tracks, but they are still commonly recorded multi track so that mixing engineers can make balance decisions about the relative levels of instruments after the performance. In the case of classical, many people believe this is the conductor's job, and that engineers should not presume to change the balance from that which the conductor and orchestra have so carefully achieved. Engineers may work closely with conductors to achieve the conductor's ideal of how the instruments ought to sound. This idea of punching in fixes is extended in the technique of *compositing* performances.

5 Composite Performances—Creating a Master Take

The ability of an artist to punch in and out of a track to make fixes eventually spawned the idea of creating composite tracks. Originally, an artist might have recorded two or three takes of their vocal, and then, along with the producer, picked the best take and systematically fixed any problems by punching in. Some time ago, a clever engineer figured out that he could mix and match the various parts of these three vocal takes, taking the best parts from each one and dubbing them into an empty track on the tape.

The way this is often implemented now is that the vocalist will sing the song across several different days, compiling maybe 20 different vocal takes of the song. Then, the vocalist, engineer, and producer will sit down with a lyric sheet and listen carefully to every take, indicating which

take contains the best version of a particular musical line. Then the engineer creates a composite vocal track that combines all these distinct performances.

Vocalists who are really compulsive (they shall remain nameless) sometimes even edit down to the syllable level. I've observed several of these *compositing* sessions in which the poor engineer had to extract a "th" from one track and an "e" from another to create the perfect "the." In this ultimate application of punching in, what you end up with is a performance that is better than the artist had actually done—a truly *master* performance. Once a composite master has been compiled, the artist, whether she is a singer or a guitarist or whatever, studies and practices this master so that they can duplicate it in concert. An example of a composite guitar solo is Jimmy Page's solo on Led Zeppelin's "Stairway to Heaven." This solo was pieced together from several different solos, to create the unified piece we now hear.

6 Impressionism and Realism

For several hundred years, beginning in the Renaissance, painters strived to bring increasing realism into their works. The discovery of the use of perspective, which had eluded earlier artists, laid the groundwork for fantastic advances in rendering scenes in oil with lifelike qualities. Around the middle of the nineteenth century, a popular movement overshadowed the realists; the impressionists strove to create scenes that didn't rely on realistic depiction to convey their emotional message. What caused this sudden change in style?

One explanation of art historians is that the invention of the camera around 1840 meant that everyone, without any special training, could suddenly capture scenes realistically. Impressionism and, subsequently, cubism were the styles adopted by artists to create engaging artistic works as a reaction to the ease with which realism could now be created.

For many years in audio, recording engineers strove to create ever more realistic recordings; to recreate the sound of a musical group on stage inside everyone's living room. In classical, folk, and traditional jazz, this is still the norm. Naturalistic microphone techniques and a minimum of processing are used to accomplish this. One notable exception is the cycle of Beethoven symphonies recorded by Herbert von Karajan and the Berlin Philharmonic in the early 1980s. Karajan insisted that the instruments be close-miked. The result was a complete loss of the normal depth in the soundscape of the orchestra. Instead of the French horns sounding as

though they were off in the distance, they and all the other instruments sounded as though they were right in your face. Many critics and the public found the recordings so disorienting as to be unlistenable. The Maestro was unabashedly pleased with the result, commenting that for the first time in his life, he could now hear the orchestra as he had always heard it in his head.

In the 1970s, recording technology reached the point where it succeeded in recreating the sound of a live band with great fidelity. The cutting edge of audio production since then has been to create something more than reality—to sculpt sound pictures that evoke feelings and thoughts unconstrained by reality: soundscapes that push the envelope of the technology available to create a sort of auditory impressionism. Artists, producers, and engineers are now able to create all the sounds that they hear in their heads, not just the ones that would occur in real life.

These days in rap, hip-hop, house, techno, and electronica, it is common for the engineer to entirely compose and perform music by grabbing samples from previous works, looping them, combining them with drum machines, editing, etc. The distinction between engineering and performing has become increasingly blurred, and the sounds created for a track can be more important than chords and notes. In acousmatic music, a branch of electroacoustic music, compositions are created out of *found sounds*, environmental sounds—such as jack hammers, breathing, turbine engines, and waves crashing—that have been recorded and reprocessed, then sculpted together to create a composition. Samplers and editing stations are considered to be musical instruments by many musicians these days, and have allowed a greater number of people to participate in the making of music. The increased sophistication and affordability of advanced technology has been a great equalizer, making music creation accessible to a larger number of people, and not just a select few with conservatory educations.

We continue this discussion in "How Recordings Are Made II" (page 15) by exploring how recent advances in computers and digital recording have affected the recording industry.

How Recordings Are Made II: Digital Hard-Disk-Based Recording

Jay Kadis and Daniel J. Levitin

1 The Brave New (Digital) World

The last five years or so have seen what may be the biggest change in recording technology since the introduction of multi-track recording in the 1960s—the move to hard disk-based digital recording. This may represent an even more profound change than the introduction of digital recording in the 1980s because hard-disk-based recording allows for editing and manipulation of the signal in ways that are fundamentally different from that which came before, even with digital tape. Because developments in this domain have been so rapid, there is a danger that anything we write today (summer of 2005) may become quickly outdated; nevertheless, we will attempt to discuss principles, techniques, and technologies as they exist today. While some details may change, we believe that the fundamental principles will apply for some time (as the principles and techniques of tape-based recording, discussed in the previous article, still apply). We will examine the many advantages that computer-based recording systems provide along with the related implications and difficulties.

Recording engineers and composers have long eyed the computer's potential for making and manipulating sound recordings, but early computers were too slow and expensive to be practical. As personal computers became faster and cheaper, desktop computers acquired the ability to perform functions that previously necessitated a building full of expensive electronic devices. This fueled a race among software developers to find

new ways of making and processing music. The now-widespread availability of tools previously available only to professionals has facilitated an explosion of music created by musicians previously unable to realize their ideas. The democratization of the recording process also brings the predictable result whenever powerful tools are placed in the hands of less-skilled operators: overuse of gimmicky production tricks and poorly composed and performed music in abundance. Simply using professional tools does not guarantee the production of professional-quality recordings, and providing access to a professional recording environment does nothing to improve musical composition. Whether boon or bane, we will see how the personal computer has altered the relationship between music creators, their tools, and the listener.

The fundamental difference between the older and newer systems for recording and manipulating sound is the manner in which the signals are encoded and stored. Prior to the development of digital audio recording, sound recordings were made by processing continuously varying voltages generated by microphones or electronic instruments. These voltages were converted into proportional magnetic fields in analog tape recorders and stored on magnetic tape. The term analog indicates that the signal voltage is directly proportional to the original sound pressure level and is continuous in nature (the term "analog" comes from the same root as the word "analogy"). Any manipulation of the analog sound representation had to be made in what is called *real-time*, meaning simultaneously as the musicians or tape recorder played.

In contrast, digital audio devices first convert the continuously varying voltage signals into a series of numbers that represent the signal amplitude. This process, called *sampling*, requires that the measurements be made very frequently so that the digital representation closely reflects the analog signal. To understand sampling, imagine that you wanted to obtain an estimate of how much traffic passes by your living room window, the room in which you plan to set up your new recording system. If the cars are going by slowly enough, you can see a car, look down to your notebook, and write it down. If they're going by quickly, in the time it takes you to look down in your notebook and write it down, another car or two may have passed and you will have missed counting them. You can see intuitively that the amount of time you spend looking out the window (sampling) the traffic has to be related to how long it takes a car to pass by and how long it takes to write down each entry. In the case of an audio signal, the rate of voltage change is somewhat analogous to the speed of the cars: the higher the signal frequency, the faster the voltage is changing.

In audio sampling, we store only the values that we measure at the sample times and do not save information about what happens between samples; thus, sampling a signal results in a discrete representation, one that is not continuous. When we choose the sampling rate properly (according to the Nyquist theorem of 1928), we don't have to worry about missing important information. Another analogy comes from the world of film-making. A film camera does not record continuous images; it samples them, putting each image into a portion of a continuously moving piece of film, called a *frame*. The standard frame rate for professional film cameras is 24 frames per second (fps), standard NTSC video in North America is 30 fps, and faster frame rates exist also. This sampling rate is sufficient to give the flicker-free illusion of smooth motion when the projected image is viewed (the projector double shutters to display 48 images per second, which is beyond the human eye's fusion rate). This system works well as long as no element of the picture moves too quickly with respect to the frame rate. For example a quickly spinning wagon wheel, whose rotation takes less than two frame times to complete, will not appear to spin at the correct rate when projected and might even appear to rotate slowly backwards. This phenomenon is known as aliasing and can be prevented by limiting the frequency of the signal being sampled to less than half of the sampling rate (that's the Nyquist theorem again).

The very high speed of modern computers allows the digital samples representing the recording to undergo significant signal processing in the short time between when new samples are acquired or played. Effects may also be applied to existing digital recordings. This is especially useful for effects that require too many resources to run in real-time on today's processors. Many effects that we could not easily accomplish using analog techniques are possible, from editing performances in very small pieces to shifting individual notes in time by tiny fractions of a second or correcting the pitch of a singer's performance. We can create sounds directly by computer synthesis and use sampled sounds and loops of music to create new compositions. We can make mathematical models of instruments that allow computer programs to simulate the physical behavior of real and imaginary instruments, and the sounds of those instruments can be played by the computer. We can cut and paste musical sounds as easily as we do text in a word processing program, moving them freely in time or pitch space to create special effects or to improve a flawed performance.

The advances mentioned thus far apply to digital recordings: music stored as a series of digital samples of acoustic pressure over time as ac-

quired by microphones fed into analog-to-digital converters. In addition, computers enable the capture and manipulation of the gestural information pertaining to how a musician actually manipulated an instrument to create the performance. Specially outfitted instruments must be used, which encode the musician's performance as a stream of instructions pertaining to which note was struck, when it was struck, how hard it was struck, and when it was released. This stream of information may then be edited and played back through a synthesizer or even an actual instrument such as a modern player piano. This entire approach is reminiscent of the player piano rolls of old, but it retains more gestural information than piano rolls did.

This is the result of a technology known as MIDI (Musical Instrument Digital Interface), providing a digital instruction language and hardware connection standard that links electronic instruments to each other and to computers. MIDI sends data that contain instructions on when to play notes (note-on), how dynamically the notes sound (velocity), how loud it should play (volume), and when the note should stop (note-off or velocity 0) as well as many special controller values and device-specific commands (sys-ex). MIDI allows synthesizers to play back digital scores stored in computer memory as sequences of events. MIDI permits selecting and altering the sounds to use for playback and any effects like reverberation that may be produced by the devices. MIDI networks connect racks of different synthesizers so that they operate as one instrument, allowing composers and arrangers to hear a symphony played by sampled instruments, for example. Popular music may be produced entirely in this manner without the creator being able to play any instruments at all! MIDI may also be used to connect the devices in a studio so that their operation is controlled from a central location, often the mixing console, using MIDI machine control instructions. MIDI time code can be used to synchronize playback from different sources such as tape recorders and synthesizer sequencers. While MIDI allows the fledgling arranger to try out ideas and textures before bringing them to live musicians, it will not prevent the arranger from writing notes that cannot be played within the range of a real instrument or requiring playing in other impossible ways.

Not everything in the production of music has changed fundamentally as a result of moving to digital audio. Many techniques of mixing and creating effects are still accomplished in a fashion similar to that of the analog studio. For instance, a mixing console is as central to a digital studio as it was to an analog studio, although it may now exist completely in software with virtual knobs and sliders on the computer

display which may be moved with a mouse instead of a hand. The engineer still needs to have the ability to change several parameters quickly and perhaps more than one at a time. The mouse is a poor substitute for a mixer full of controls; consequently, digital studios often interface devices with physical knobs and sliders reminiscent of the analog mixing panel to their software mixers. The digital audio workstation, or DAW, is modeled on the recording studio of old because the methods developed in the traditional recording studio are still sensible and adequate ways to record and assemble multitrack recordings, albeit the DAW does have some advantages over studios of yore.

One of the major advantages of digital audio is the nondestructive nature of editing and recording. Now we can have as many tracks as the speed and memory of the computer will allow rather than the fixed number of tracks provided by a dedicated tape recorder. If we wish to add material, we can *punch in* without overwriting the previously recorded tracks. We may select parts of tracks to be combined in a *playlist*, a listing of pieces of sound that are to be played back in the order we specify regardless of when the pieces were originally recorded. This process is automated and recallable, so we can make small changes and listen to multiple versions easily. The ability to independently manipulate pieces of data is known as *nonlinear editing*. Using analog systems, tracks recorded synchronously to tape cannot easily be moved in time relative to each other whereas digitally recorded tracks can be accessed freely and independently.

Another advantage of the computer-based studio is the cost savings of using software for simulating the hardware devices found in the traditional studio. Signal processors like compressors, limiters, delays, reverberators, and other special-effects devices were plentiful but expensive in the traditional studio. These hardware devices are now being modeled in software (often called *plug-ins* because they are add-ons to the basic software package), creating programs that process the digital information much like the analog gear of old. The advantage is that rather than buying a new compressor when we exhaust our supply, we need only click a button to install another copy of the software emulation and we're ready to go on with the mix. The number of copies that we may use is limited only by the power and speed of our computer. Add-on signal-processing cards are available to increase the amount of processing power a computer may provide, including both complete recording and mixing systems like Digidesign's ProTools TDM and add-on digital signal processing (DSP) systems like the Universal Audio UAD-1, designed to augment host-based recording systems like Apple's Logic.

2 Trouble in (Digital) Paradise

Digital audio continues to evolve rapidly, taking advantage of the constant stream of faster computers and new interface formats that allow more data to be transferred and stored. The speed of this evolution is both an advantage and one of the potential problems plaguing users of the new technologies. Issues of incompatibility are exacerbated when the elements of the system are all changing quickly, and often independently. While analog technologies were also constantly improving, an analog tape recorder (like Studer or Ampex) or analog consoles (like Neve or SSL) were usable for decades. Digital hardware and software change quite rapidly and completely, requiring frequent upgrades to remain compatible. This state of constant change makes it harder to train new engineers and the need to continually learn new software and hardware takes time away from the job of recording music for experienced engineers. The software is complicated and may contain *bugs* that cause system failures in special circumstances not discovered by the programmers, leading to computer crashes, lost information, and angry customers. Selecting add-on audio and MIDI interfaces for the computer is complicated by differing requirements of the various computer systems and recording software packages. Recording engineers now also need to be computer technicians to be assured of the ability to keep their studios operational.

For all its advantages, digital audio introduces some problems that did not exist with the analog approach. A friend of ours, a famous producer/engineer who has a studio in his home, spent several months getting the various components of his system running: the computer hardware, operating system, recording software, sequencing software, various input/output devices, plug-ins, etc. During this time, a minor upgrade to the computer's operating system was released, but since it wasn't compatible with the plug-ins, our friend decided (of course) not to do the upgrade.

He began to record an album, and after four months, his computer's motherboard failed. The manufacturer offered to send our friend a brand new computer, with a faster processor and more hard disk space, and the whole package was sent by next day air. Unfortunately, the new computer came with a newer version of the operating system (one that was incompatible with the plug-ins) and the new computer's logic board was incapable of operating under the previous system. This meant that the album production had to stop completely. With the plug-ins and various other I/O devices not working, the producer could not maintain continuity between what he had done before and after the hardware change. This

demonstrates that the complex interdependency of hardware, peripherals, operating systems, drivers, and applications demands a thorough understanding on the part of those depending on such systems for production.

Software and hardware incompatibilities aren't the only source of potential trouble. Time synchrony is required when sampling, since we must guarantee that the timing of the sample measurements is correct for every track we record. While most digital systems are able to automatically synchronize with themselves and with each other, the accuracy of the clock that determines the sample time must be consistent throughout the entire system.

For instance, if devices clock each sampled word at different times, clicks may be generated in the audio data stream. Or, if the reference clock is not perfectly regular, digital data may be output with slightly different intervals between words, altering the audio data produced and potentially affecting the perceived stereo image. There is also an issue of delays created in the digital system as data are moved around inside the devices. In order for many tracks of recorded music to be played simultaneously, each must have undergone exactly the same internal delays, or *latency*.

If simultaneously-recorded, live tracks are played back with different delays, they combine to produce peaks and dips in the frequency response as certain frequencies reinforce and others cancel to produce an unwanted comb filter effect. This is a particular problem with stereo tracks where the sound is simultaneously recorded but played back slightly apart, causing the stereo image to shift. (When tracks are overdubbed, this is not so much of an issue since the sounds were not time-locked to start except by the musician's timing accuracy, which is nothing like the microscopic resolution of the sample clock.) Each process, analog-to-digital conversion, data storage, data manipulation, and digital-to-analog conversion must delay all tracks the exact same amount in order for them to be played back synchronously.

Loss of synchronization is not just an abstract problem but one that occurs frequently in practice. In the analog world, tape machines typically have two *heads* over which the tape passes and which read or write material onto the tape. One head is used for high quality playback, and the other is used to record. This is because the physical processes of recording and reproducing magnetic signals are different and optimization requires different head designs for each process. If one wants to *overdub* (add an instrument to one that has already been recorded), a potential problem exists.

Suppose that you've recorded drums and now you want to add bass. If the record head and the playback head are not in the same physical

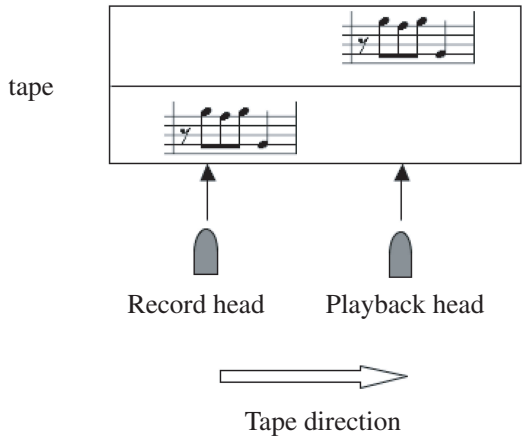

Figure 1.

location along the tape's path, there will exist an asynchrony. You'll hear the drums playing the moment the tape passes the playback head, of course. As you play bass along with those drums, the record head will put the signal on tape, but if the record head is an inch or so away from the playback head, your bass part will be out of synchronization with the drum part by the distance between the two heads. At 15 inches per second (IPS) with an inch between the heads, this one inch would cause the bass part to be recorded 66 ms later, or the equivalent of 1/16th note at 64 beats per minute (bpm).

This problem was solved in analog recording by combining lower fidelity playback electronics on the same physical component as the recording head, in what is often called a *sync head*. This permitted the musician to hear back the previously recorded track (at somewhat lower fidelity) and add something to it virtually instantaneously, and in (virtually) perfect synchrony.

In the analog world, no matter what happens to the master tape, the performances on it will remain synchronized with one another so long as they started out synchronized; not necessarily so in the computer-based digital world. The two of us had to mix a song once for which the session file had become corrupted. We still had access to the individual sound files of the performance—the basic tracks and all the overdubs—but the ProTools session file that contained instructions for how the tracks were to line up in time had been destroyed. We had to import each individual sound file—for the kick drum, snare drum, bass guitar, vocals, etc.—into

a newly created session file. Because some of the drum files were edited after recording to get rid of noise at the beginning, and because some of the files were from overdubs not recorded with the basic tracks, lining up all the sound files in the mix/edit window did not succeed in synchronizing them. In fact, there was no information available at all to tell us where to place the files with respect to one another. Lining up the snare drum and kick drum was time-consuming but not complicated because it was obvious how those parts were intended to fit together (and bleed-through in the mics gave clues). Lining up the vocals and the guitar solo, however, was extremely difficult. In some cases we knew roughly (within half a second or so) when they were supposed to occur, but the vocals and solo had been performed with a particular feel, a very specific relationship to the beat that we were unable to recapture. We spent hours nudging the parts around by 10 and 20 milliseconds to get what sounded good to us, with no objective information about how the singer had intended to place his vocals with respect to the beat. To make matters worse, there were pieces of vocal and guitar performances—a few notes here and there—that had been recorded on separate tracks as repairs or fixes, and we had no idea where *they* were supposed to go or what they were intending to repair or fix.

The ease with which files are created by computer recording often leads to sessions with huge numbers of individual sound files. This demands careful attention to file naming and recordkeeping. Every time you create a new track, give it a name that indicates what it is. Good names: Lead Vocal I, Snare Drum, or Replacement Rhythm Guitar. Bad names: Audio 6 (the default name the computer software might assign), RE20 (the name of the microphone you used—but what instrument did you record?), or July 7. Take as many notes as possible about the track and write them directly to the computer or in a project journal: the microphone used, mic pre-amp, time of day, compression settings, and what the part was intended to do (that is, how it was intended to fit into the final mix). The microphone and signal processing notes will help you if you want to go back and recreate that sound. Be sure to distinguish tracks that were intended as retakes or replacements as opposed to primary parts, so that you or some poor mixing engineer doesn't waste days trying to figure out if two parts are redundant or not.

In the digital domain, it takes a small but often noticeable amount of time for a digitally-recorded signal to be converted to analog. Suppose that it takes 20 milliseconds for a signal to wind its way through an A-to-D or D-to-A converter, and you want to play bass along with a previously recorded drum part. The drums take 20 ms to get through the converters,

so your speakers get them 20 ms after the computer "plays" them. Now, if you're sitting 8 feet from your speakers, assume that it takes another 7 ms for the sound to reach your ears. (While we usually think of overdubs employing headphone monitoring, some performers, reportedly including Frank Sinatra, dislike performing with headphones and favor loudspeaker monitoring for overdubs, allowing a more natural performance.) You play along as best you can, but it takes the computer another 20 ms for the sound from your bass to get through the A-to-D converter on the way back in. Your sound is now recorded 47 ms *after* the drum track (not unlike the case with analog overdubs that we just spoke about). For more information see Derek DiFilippo's "Perceivable Audio Latencies" in *Audio Anecdotes I.*

The software designers, in theory, know how long the hardware converters take to process sound, and part of their job is to build in a synchronization function that should synchronize your overdub with the previously recorded track automatically. The only thing that they can't take into account is the small delay from the speakers to your ear because that will vary from room to room. The hypothetical 7 ms in our example is truly insignificant when it is the only delay, but it can make a noticeable difference when it adds up with other sources of delay. Most semiprofessional and professional software recorders have a way to minimize, though not eliminate, converter delay. Known by names such as *low latency* or *overdub mode*, this options shrinks the buffer size during playback to give you the fastest playback possible. This is usually at the expense of being able to use large amounts of built-in processing, such as compressors, equalizers, reverbs, etc., which take time to employ but which can be added back in during mixdown. Alternatively, you can monitor inputs directly and not listen through the digital device, eliminating the digital monitoring delays entirely.

3 Summation for the Defense?

As we are beginning to realize, every technological revision comes with a cost. When fuel injection replaced carburetors in automobiles (new cars haven't had carburetors since the early 1990s), it provided a more reliable system for delivering fuel. Gone are the cold mornings when the car wouldn't turn over at all, the sudden sputtering during high altitude climbs, and much of the pollution caused by imprecise air/fuel mixtures. Fuel injection works well without requiring maintenance for many times the miles between carborator tune-ups, but when the fuel injection com-

puter fails, it fails all at once and no amount of tinkering will get you back on the road. A carburetor failed gradually, gracefully—you had warning coughs and sputters as gaskets decayed or springs lost their tension, and if you knew what you were doing you could enrich or lean out the mixture to accommodate changing climate, elevation, or wear.

When analog tape, or for that matter tube electronics (amps, compressors), fail, they often do so gradually; the tubes begin to leak and performance is compromised, but the hardware is still useable. Old analog tapes eventually wear out as a function of age, poor storage conditions, or too many playbacks, but on their way to wearing out they are still useable. The loss of high-frequency information or occasional drop-outs are often the first clue that a tape is deteriorating. But, with digital recording as found on digital tape, hard-disk recording, or even CDs and DVDs, the deterioration is masked by error-correction schemes. Error correction ensures that even imperfect media (ever hold a CD up to a light source and notice the pinholes?) can play back bit-for-bit accurate sound—that is, until the deterioration exceeds a threshold where the error correction can no longer cope, potentially rendering the recording utterly and instantly unusable. While some professional gear will report digital media's bit error rate to allow an engineer to monitor potential deterioration of media, this feature is far from universal. In a pinch, we could always use a partly compromised analog tape (as was done for many CD reissues—that's why some of your favorite albums don't sound all that great on CD), but a digital recording that has been corrupted cannot be used at all.

What advantages does digital recording technology offer to the musician? With MIDI-based sampling systems like Tascam's GigaStudio, a composer can hear a close approximation of a symphony playing her composition using only the computer. A musician can record music in a bedroom studio and experiment with different arrangements, perfecting songs without requiring the participation of other musicians. Recordings may be edited to produce near-perfect performances, and slight timing and pitch problems may be eliminated. Bands can record their own albums, taking as much time as they desire without spending a fortune for studio time. Musicians in far-flung areas of the world can collaborate in recording by sending music over the Internet or through the mail for others to contribute to and return. Musicians can now record, master, duplicate, and offer for sale at performances their own CDs. Digital music is easily distributed over the Internet, so unknown bands can find an audience that they could never meet physically. This represents a true democratization of the recording and distribution process.

Recent developments in Internet technology may result in an even more active method of collaboration: real-time interactive performance over the network. It is now possible to reduce the time lag associated with digital audio systems connected to the Internet to a delay short enough to allow musicians around the world to play together. Our research groups at Stanford and McGill have worked on particular algorithms for reducing broadband latency as much as possible, and members of the Stanford and McGill jazz community have conducted several live, "low latency" internet jam sessions to demonstrate the technology. We are still at the beginning of such possibilities, and there are some inherent limitations to the process, but digital audio promises to provide new capabilities to musicians and music lovers in the future that are difficult to imagine.

The Internet and digital music also present a problem for the creators of music due to the ease with which sound files can be exchanged. Current digital delivery media like the compact disc contain no method of copy prevention, so anyone can extract, or *rip* the digital audio files from a commercial CD to their computer and send them over the Internet to anyone else with a computer. Preventing unauthorized copying is one reason for developing new digital media. It has also helped promote the development of new digital techniques of representing music that both provide improved sound quality and easier protection from copying. Super audio compact discs (SACD) use different encoding of the audio data that simplify playback circuitry while preventing computers from reading or playing the discs. Future media for music distribution will likely be secure from unauthorized copying. While this protects the commercial producers of music, it also makes it more difficult for independent musicians to use the technology.

While many of the signal-processing devices, synthesizers, and sequencers now available provide preset, instant sounds, it is important to fight the tendency to overuse them! In most cases, the designers of these presets intended them as a starting point for exploration, not an end point. The danger of using presets is that music will become increasingly homogenized as everyone starts to use the same effects, reverbs, or synth sounds. It also introduces the danger of music sounding dated: as soon as new sounds come out, everyone rushes to be the first to use them, and then after several years those sounds become associated with a particular era. As always, the best way to make fresh, creative, and high-quality recordings is to let your ears guide you. Whether it is a digital reverberation device, a new string synthesis program, or a guitar amplifier simulator, we recommend that you play around with and modify