

Sound for Digital Video

SECOND EDITION



SHOOTING

EDITING

MIXING

MASTERING

TOMLINSON HOLMAN
ARTHUR BAUM



Sound for Digital Video

This page intentionally left blank

Sound for Digital Video

Second Edition

**Tomlinson Holman and
Arthur Baum**



Focal Press
Taylor & Francis Group

NEW YORK AND LONDON

First published 2005 by Focal Press

This edition published 2013

by Focal Press

70 Blanchard Road, Suite 402, Burlington, MA 01803

Simultaneously published in the UK

by Focal Press

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

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

© 2013 TMH Corporation

The rights of Tomlinson Holman and Arthur Baum to be identified as authors of this work have been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Notices

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

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

Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging in Publication Data

Holman, Tomlinson.

Sound for digital video / Tomlinson Holman and Arthur Baum. – Second edition.

pages cm

Includes bibliographical references and index.

1. Sound—Recording and reproducing—Digital techniques. 2. Sound—Recording and reproducing—Equipment and supplies. 3. Digital video. I. Baum, Arthur.

II. Title.

TK7881.4.H627 2013

777'.53—dc23

2012042247

ISBN: 9780415812085 (pbk)

ISBN: 9780203384053 (ebk)

www.focalpress.com/cw/holman

Typeset in Palatino

By Cenveo Publisher Services

Contents

Preface	xi
Chapter 1 Basic Concepts	1
Digital Video Landscape	4
Minimum Standards for Audio	5
The Four Dimensions of a Sound Track	11
Digital Sound	17
Features of Digital Video Recording Formats	21
Chapter 2 Introduction to Digital Video	25
Basic Digital	26
Some Basic Video for Audio People	27
Frame Rates	27
Interlaced Video	30
“Film” Look	32
Resolution and Aspect Ratio	36
Under- and Over-Cranked Camera	40
Operational Matters	41
Quality Modes and Codecs	41
Interchangeability	42
Off-Line/Online	44
Time Code	45
User Bits	50
PAL Formats	50
Locked versus Unlocked Audio Sample Rate	51
Instant Playback	52
Interconnecting Video	53
Conclusion	53
Chapter 3 Preproduction Planning	59
The First Big Step: A Dedicated Sound Person	59

Location Scouting	61
Choosing a Sound and Camera Workflow	64
Single-System Recording	64
Double-System Recording	68
Testing Sync	72
Mic, Line, and Speaker Level	73
Reference Levels and Headroom	74
Setting Up the Recording Chain	77
Line Levels	78
Mixing Balanced and Unbalanced Connections	80
Chapter 4 Production Sound I: General Considerations	85
Coverage	85
Scene Coverage and Microphone Technique	88
What Can Be Done with an On-Camera Microphone?	92
How to Use the Two Channels	94
Other Items Recorded during Production Sound	98
Microphone Accessories	99
Booms/Fishpoles	99
Shock Mounts	102
Windscreens	103
Pop Suppression	105
Case Studies	106
Mounting Lavs	106
Radio Mic Usage	109
Disposable Mics	109
Production Sound Example	109
Boom Operator's Job	113
Common Problems	114
Logging	115
DV Production Sound Log	116
Sound Kit Accessories	117
Chapter 5 Production Sound II: Microphones	119
Power	120
Dynamic Microphones	123
Polar Patterns	123
Differences Between Microphones Having	
Various Types of Polar Patterns	128
The Radio Part of Radio Mics	133

Chapter 6	Production Sound III: Dealing with the	
	Output of Microphones	139
	Recording Level Parallels to Early Cinematography	142
	Cries	148
	Whispers	153
	Cries and Whispers	155
	Multiple Level Controls in the Chain	158
	Another Kind of Overload Distortion and	
	How to Avoid It	161
	Combining Features for Best Wind Performance	162
Chapter 7	Media Management	167
	What Counts as a Backup?	167
	Types of Media	170
	Streaming Transfers and File Transfers	171
	Streaming Transfers	173
	File Transfers	185
	Audio File Formats	188
	Managing Double-System Recordings	189
	Syncing Footage	192
	Time code Autosync	193
	PluralEyes and DualEyes to Autosync	194
	Syncing by Hand with a Clapperboard Slate	195
	If a Slate Wasn't Used (What to Do in Case of Emergency)	196
	Common Problems in Digital Audio File Transfers for Sound	
	Accompanying Picture	197
Chapter 8	Sound Design	201
	Film Sound Styles	208
	Realism	208
	Stretched Reality	213
	What is Seen versus What is Heard: On-Screen	
	versus Off-Screen	214
	Hyper-reality	215
	Surrealism	216
	Montage	218
	Shifting Levels of Reality	219
	Sound Design as an Art	222
	Spotting	223

Chapter 9 Editing	225
Non-Linear Editing	225
Random-Access Editing	227
Non-Destructive Editing	227
Visual Waveform Editing	228
Edits and Fade Files	229
File Management	230
How to Edit	231
Fine Editing of Production Sound	234
Stealing Presence	237
Where Presence Is Used	237
Documentary Considerations	238
Fixing Bumps	239
Sound Effects	242
Ambience/Backgrounds	245
Foley Effects	246
Cutting Music	247
Scene Changes	247
Plug-Ins/Processes	249
Tracks and Channels	261
Busses	261
Pan Pots	262
Solo/Mute	262
Grouping Tracks	262
Differences Between Picture and Sound Editing Systems	263
Picture–Sound Sync Resolution	264
Chapter 10 Mixing	267
The Mixing Hourglass	270
Audio Processes	271
Processes Related Mainly to Level	272
Level Controls	272
Gain Staging	274
Hand Compression	274
Compression	275
Limiting	276
De-esser	277
Noise Gate	277
Downward Expander	278

Processes Related Mainly to Frequency	278
Equalization	278
Filters	281
Combinations of Level and Frequency	284
Time-Based Devices	284
Reverberation	284
Other Time-Based Effects	286
Other Plug-Ins	287
Panning	288
Routing, and Limitations Caused by Routing	288
Busses, Channels	288
Delay Compensation	289
Voice Limitation	290
How to Mix	290
Chapter 11 Masters and Monitoring.	295
Choice of Sound Format on Edit and Delivery Masters.	297
Uses of Surround Sound	304
Mastering for Level	305
Background on –20 versus –12 dBFS Reference Level	306
Level Calibration	307
When You Can’t Calibrate with Test Signals	310
Best One-Size-Fits-All Approach	311
Mastering for DVD, Blu-ray, Digital Broadcast, and Digital Satellite Television.	313
Monitoring	322
Film versus Video Mixes.	326
Index	329

This page intentionally left blank

Preface

This book is for those who wish to improve sound quality and make it more interesting for digital video productions. Many theoretical topics such as descriptions of basic sound, psychoacoustics, and applications to theatrical entertainment may be found in Mr. Holman's book *Sound for Film and Television*. Instead, this book is for anyone interested in making a wide variety of digital video projects more impactful through the use of sound. We have striven to include practical advice every step of the way, from production through distribution. Here you will find specifics for the use of digital video technology, audio accompanying high definition picture, dialogue and sound effects editing, single and double-system audio recording, and tips for selection and use of equipment, all rooted in best practices from the world of motion picture audio. As each major topic is introduced in turn, the fundamentals come first, followed by greater elaboration. So, if you are coming to sound for the first time, you may find it easiest to skip through the book, starting on each topic and moving on to the next when you get enough information to do the job at that stage. In order to facilitate this approach, each chapter ends with a section called "The Director's Cut," which is the briefest possible statement of the ideas in the chapter. The notion behind the appropriation of that phrase for the first edition was to alert sound people to "what the director should know," and the information in each "Director's Cut" has been streamlined and brought fully up to date for this new edition of the text.

The scope of coverage in *Sound for Digital Video* runs the gamut from pre-production planning, equipment selection, and production sound recording to editorial processes, postproduction mixing, and distribution in multichannel audio formats on the latest media. After two chapters introducing basic concepts in digital audio and video technology, we begin in earnest with a new focus on setup and preproduction planning in Chapter 3. For those who must start a production just after picking up this book, Chapter 3 is probably the place to start thanks to its treatment of setting up

equipment and making important decisions en route to the best sound possible. The next three chapters provide expanded coverage of production sound, a core topic in this book and in the field of sound for video. Recording good production sound is important even for simple projects because if you have good production sound, then editing and mixing are relatively easy, whereas if you don't, at the very least you are going to spend more time and money getting a decent result.

Production sound is divided into three chapters of increasing complexity. The first chapter explains the various microphone techniques commonly employed, what can be done with an on-camera microphone, how to use the audio channels available on a medium, and what to record on location besides speech. Chapter 4 also covers microphone accessories, includes case studies of various microphone techniques, and features an analysis of raw production sound tracks available as extras on a commercial film release, with information derived from an interview of the production sound recordist. A walkthrough of the boom operator's job is followed by advice for common sound problems, information on logging recordings, and suggested sound kit accessories for practical use.

Chapter 5 concentrates on microphones: their selection on the basis of how they do their job and how they respond to sound coming from various angles, powering, radio microphones, and so forth. Chapter 6 deals with properly recording the electrical signal coming out of the microphone, which may cover a very wide range that has to be accommodated by recording equipment to result in noise- and distortion-free recording. This includes a particularly salient treatment of how and where to set the level in a recording chain.

Chapter 7 presents a host of new information on media management, including managing physical media on set for the wide range of formats encountered in video today, as well as the transfer and sync of audio and video media across devices from production to postproduction. Chapter 8 continues the transition to postproduction with a treatment of sound design, a useful approach to thinking about sound tracks. Sound design is the art of getting the right sound in the right place at the right time, and this chapter explains that process.

Chapters 9, 10, and 11 cover sound editing, mixing, and mastering and monitoring—in one word, postproduction. Today, the traditional boundaries of film sound editing and mixing are being broken down through the

introduction of plug-ins for editing systems that allow editors to access what were traditionally mixers' techniques. Those same plug-ins, when used correctly, give independent producers the ability to radically improve the quality of sound in even the simplest digital video productions. The question today is: where do I optimally perform these processes so that the project is run the most efficiently, with the best art applied in the least time? Some considerations for answering this question are given here.

A course in *Sound for Digital Video* might start with the material in Chapters 3 and 4, then move back to 1 and 2 before continuing with Chapter 5 and so forth. Chapter 7 could certainly serve as more of a reference alongside the earlier and later chapters in the text. Still, for a person who wants a fuller explanation of the basic concepts and the technology underlying digital video, starting at the beginning and going straight through the book is perhaps the best approach. The scope and utility of this book are greatly influenced by the large volume of activity in the field of audio production. In this edition, we have made every effort to focus on the principles behind the important decisions that must be made at each step of the process so that, even though one piece of equipment or audio editing system may be in vogue at the time of printing, the advice given here is universal and will remain practical and reliable for years to come.

This page intentionally left blank

CHAPTER 1

Basic Concepts

Attend a shooting of *Will and Grace*, watch one of the most accomplished directors in the world while the cameras are rolling, and you might ask the question: Where is he looking? James Burrows (*Mary Tyler Moore*, *Cheers*, and many more) is staring at the ceiling. He could be looking at the actors, but their routines have been blocked out days ahead, and they are highly accomplished. He could be looking at the monitors to see if the four cameras are covering the action as planned, but he has a fully professional crew and has been through camera rehearsals, so he knows they'll get it right. Why is he looking at the ceiling? Because he's *listening* to the delivery of the lines, the ultimate consideration of getting the show on film, constantly asking "is it funny?" The sound of the show is its bottom line. The question a sound person doesn't want to hear from such a director is "why is that distorted?" or "noisy?" or "reverberant?" This book explains how at each juncture—recording, editing, mixing, and mastering—mistakes can be avoided, sound quality improved, and sound design fit to the story.

It doesn't take a Hollywood studio environment to produce programs today. In fact, more and more people are turning to inexpensive digital video technology to create professional quality content. New advances have put high-definition video and audio within reach for amateurs and low-budget filmmakers alike, bringing cheaper and better methods for storytelling to a wider pool of people than ever before, and new distribution platforms such as Internet video have further democratized the media. But even with all these advances in digital video production and distribution, knowing how to use the technology effectively remains just as important as having access to it. In a landscape where anyone can make programs that stimulate, amuse, inform, frighten, and do virtually all of the things that cinema and television have been doing for more than one hundred years, what most often separates amateur efforts from professional looking (and sounding) programs is the filmmakers' attention to aesthetic detail—especially the quality of the sound.

Technologically, we have the capability to do more with video and audio today than at any point in the history of motion pictures. The digital video revolution has given rise to cameras and camcorders with greater image resolution and cleaner sound reproduction than that of older analog formats. Moreover, digital recordings can be transferred with no generation loss into picture and sound editing systems, edited, mixed, and exported back to tape or disc, all while preserving the fidelity and clarity of the original recordings. This is important because older analog methods suffered from generation loss and required extensive off-line/online facilities for editing professional programs. While consumer analog video formats were fine for home video use, copying and editing analog video inevitably led to imperfections in the final product.

Analog copies are like those from copy machines: if you copy an original, then copy that copy, and so forth, the picture will get fuzzier and have more contrast than the original after several generations as slight differences in each copy build up and become noticeable. On the other hand, digital recordings ingested into an editing system are copied bit-for-bit so that each copy is an exact clone of the camera original. The same is true of transfer between editing and mixing systems and back to tape or disc for archival purposes and distribution. The fact that each copy is a literal clone means the work throughout the process is more visually and audibly true to the original than ever before, with the potential for truly lossless transfers along the entire path from the captured expression of an actor or documentary subject to the performance seen and heard by an audience. Of course, digital editing systems also offer the possibility of processing and modifying recordings in creative ways to produce a whole greater than the sum of its parts. Once modified or created digitally, new files are also copied faithfully with no further generation loss throughout the finishing process. Your objective may not always be to make perfect clones of the original recordings, but digital recording and transfer still offer the most transparent route from production through postproduction, and it is important to know that the system is transparent when it needs to be.

Unfortunately, a common misperception among producers is “now that it’s digital, we don’t need any postproduction, right?” The truth is that limitations on the quality of production sound recordings are caused by practical considerations such as noise on the set more commonly than the actual technical limitations of the medium. Furthermore, editing a scene together from several shots means that voices recorded differently from shot to shot have to be blended together smoothly in postproduction, making them sound as though they come from one continuous recording

so that technical matters do not distract the listener. So even today there is plenty of work in postproduction, both to improve on the original production sound recordings and to add effects, music, and the like to create a compelling professional soundtrack.

Although digital video affords filmmakers a new level of quality and simplicity of use today, the main item hindering better sound design and sound quality is, frankly, most often the training and experience of those doing the job, including not just sound-specific people but everyone involved in the task. Viewing art house films, produced on video with small budgets but with great passion, reveals this to be true. Almost all of them have a carefully controlled look, formed by the synergy that occurs among location selection, production design, costume design, makeup, lighting, and cinematography, because everyone is aware, more or less, that these visual elements are important in telling stories on film or video. However, the sound quality of independent films varies tremendously because not everyone thinks about sound in the same way they carefully plan and control the visuals. Since sound is “invisible,” it’s harder to “see” the impact of good sound during production and editing, and too often even the sound design created in postproduction is not properly fit to the story. The purpose of this book is to inform users of the potential for good sound to improve motion picture storytelling, whether fiction or nonfiction, and to give filmmakers the tools and information they need to better utilize the storytelling medium in its full expressive capacity. Two famous filmmakers agree: sound is very important. George Lucas and Michael Moore represent a considerable range of filmmaking styles, yet each director insists that sound is vital to the overall motion picture experience.

The range of productions covered by this book is from simple one-person shooting situations (generally using cameras meeting certain minimum audio requirements) up through independent video productions shot for feature film release with larger crews. Included along the way will be tips that are useful for industrial and commercial information videos, webisodes, video blogs, meeting videos, event videos such as wedding videos, documentaries, and scripted videos designed for both film and television release and for distribution via the Internet and mobile platforms. Note that the techniques given here, and inherent in the medium, are neutral with respect to their use, in a political and ethical sense. They could have been used to great effect by Leni Riefenstahl, had she had them, to fawningly document Hitler’s rise, just as they can be used to improve the storytelling and emotive power of the best documentaries and fiction films today. Technique is always only a means to an end.

Digital Video Landscape

The digital video landscape has grown tremendously in the past decade with the proliferation of high-definition video formats and camcorders, alongside increased consumer interest in the video capabilities of digital devices such as phones, tablets, and traditional still image cameras. Whereas ten years ago “digital video” was nearly synonymous with DV tape formats, today newer file-based systems and video compression formats have taken over much of the market. One consequence of the proliferation of formats is increased complexity in ensuring the interoperability of editing systems and interchangeability of recordings made on different devices. For the most part, mobile devices such as phones and tablets have filled the gap left by analog video and older consumer DV formats in the “home video” market for quick and easy capturing and sharing of video with friends and relatives. More professional projects demand higher quality video shot on digital camcorders or, increasingly, the motion picture recording modes of high-end digital SLRs.

Still, the range of features and formats spanned by purpose-built digital video cameras today is staggering. The simplest camcorders have a built-in zoom lens and a single CMOS or CCD image sensor delivering resolution (or sharpness) and a range of color that is less than that of a good conventional video monitor. More expensive digital video cameras may have lens mounts allowing lenses of different focal lengths to be used, and may have one or three high-resolution image sensors delivering much greater sharpness and color fidelity than low-end models. Digital single-lens reflex (SLR) cameras offering high-definition video modes became popular for shooting beginning with the introduction of the Canon 5D Mark II to the market in late 2008. One advantage of shooting on DSLRs is the capability to interchange lenses, using the same high-quality lenses designed for professional still image shooting. Still higher-end digital video cameras may even utilize 35 mm cinema lenses via their standard lens mounts or adapters, and may use larger image sensors to achieve shallow depth of field and more closely approach the look of shooting on film.

Early digital video cameras, both standard-definition and high-definition models, shot on tape, recording digital video and audio streams to 1/4” or larger magnetic tape instead of the older analog signals recorded by conventional tape decks and camcorders. As the field of digital video has grown, both in original production via digital camcorders and in digital distribution of motion pictures via DVD and Blu-ray, new recording media

such as hard drives, optical discs, and flash memory cards have caught up and overtaken tape-based workflows in many sectors of the market. Still, both consumer and prosumer digital video cameras today may use tape-based or file-based recording, or even offer both methods of recording on the same camera. Cameras that record to tape may work with multiple tape formats, including consumer-based and more professional format cassettes, and may have the option of recording different video formats depending on which type of cassette is loaded in them. Cameras offering recording on flash memory devices, such as P2, SxS, SDHC, or even Compact Flash cards, typically offer a range of video bit rates and resolutions as well, and may record in one or more video formats; DV, MPEG-2, and MPEG-4 codecs including AVC/H.264 are common on cameras available today, and new codecs such as HEVC are expected to be forthcoming, and new codecs such as HEVC are expected to be forthcoming. So the range of options in selecting a digital video camera is wide, and a particular camera may itself have a wide range of options such as recording medium, format, and lens choice that must be tailored to the needs of a production.

Across the spectrum of digital video cameras in use today, there is also a range of audio recording capabilities represented, but the capability of a camera to record useable audio for a given production may generally be gauged by looking at a few specific features and specifications. The following minimum standards for audio dictate what is typically needed not just in terms of technical quality, but also in terms of the ability to capture sound in a practical sense, using proper microphone placement and mixing technique as discussed in the chapters on production sound in this text.

Minimum Standards for Audio

Although all camcorders record some kind of audio, just what audio they are recording is another matter altogether.¹ Serious limitations are

¹ The present discussion, and the thrust of this text, is on digital *video* cameras and camcorders, typically recording from standard-definition up to 1080p high-definition video. More expensive digital cinema cameras used on high-budget Hollywood productions may not necessarily record audio, since it is typical for sound to be recorded double-system on high-budget features (see below). Double-system recording is a necessity for projects shot on film, and remains the norm for digital cinema productions because of the high-quality components and features, such as multichannel mixing and recording, present in dedicated audio recorders.

found for camcorders that record only from on-board microphones with no available external inputs, and for those that lack manual level controls. Lesser limitations sometimes occur even with cameras equipped for external inputs and with manual level controls in the form of the types of inputs available and their features. Finally, the audio recording format (PCM versus “lossy” formats) may dictate whether it is appropriate to use the camcorder as the primary audio recorder on a production.

If budgetary or picture considerations necessitate using a camera that doesn’t meet the minimum standards for audio discussed below, it may still be possible to record good audio using a dedicated audio recorder instead of recording audio on camera. This approach, called “double-system” since the sound and picture recording systems are separate, will be discussed in detail in Chapter 3. Although double-system recording is common in professional work, one potential drawback to such a setup is that more time and effort must be expended to sync audio and video when transferring files to an editing system, whereas “single-system” recording offers the benefit of marrying sound to picture in camera as both streams are recorded on the same medium. For that reason, single-system recording is commonly used when available on projects with a large volume of footage such as documentary films, and has traditionally been used on digital video projects of all kinds, including web-based video, because of its ease of use in postproduction.

For single-system recording, a camera should meet the minimum audio standards outlined above. First and foremost, the limitation of using an on-board microphone as the only audio input to the recorder is just too great for practically any pro audio use, although a few things can be done this way that will be described later. While an on-board microphone may be used to record a “guide track” to aid in syncing double-system sound, selecting a camera that only allows the on-board microphone to be recorded is just too great a limitation for most single-system uses described in this book. Cameras falling in this category include both the simplest cameras and some of the more sophisticated high-definition consumer cameras. It is perhaps the first question for the sound-savvy person to ask of a camcorder: Does it have external audio inputs that can be substituted for the on-board microphone, if one is present?

The next most common limitation is lack of manual control over level. While some built-in automatic gain controls work reasonably well today, nonetheless their constantly seeking and changing gain according to the

input level causes problems editorially. With automatic gain riding, one of the most common problems is that the level is turned up when things get soft, such as in long pauses in a speech. Then the automatic gain control (AGC), also called a *compressor*,² turns the level back down when speech starts again. An operator, being smarter, will leave the level control alone when a long pause is being recorded. Later, in postproduction, an edit can be performed to cut from one pause to another with minimal problems if the operator leaves the manual level control alone. If the gain is constantly changing as a result of the use of AGC, the result is likely to be a jump in level at any particular edit, which must at least then be manipulated for smooth and continuous sound. And this problem could be repeated for every edit.

Another problem with compressors is called pumping, wherein the gain changes become audible within a word in speech, although some more modern AGCs seem to have eliminated this problem. Normally, AGCs will limit your ability to make edit points just too much to be used for high-quality work. So it is interesting that finding manual level controls on a camera is a mark of a higher-quality camera; inexpensive ones are more automatic, both in picture and in sound, but higher-end ones allow greater manipulation of the image and sound by the user.

High-end professional cameras usually have balanced audio inputs, for both microphone and line levels. These are usually represented by the presence of XLR connectors (Fig. 1-1) on the camcorder body or on a microphone input accessory. The reason to use balanced lines, with their two principal conductors covered by a common shield conductor, is that this type of wiring is less susceptible to hum caused by magnetic fields associated with all kinds of power wiring. Alfred Hitchcock's first day of shooting the first British sound film was greatly delayed until it was sorted out that the microphone cables could not be run next to the lighting power cables, because otherwise the resulting hum was overwhelming. Balanced wiring tends to reduce such magnetic-field-induced hum, so it is preferred in professional use over the more common unbalanced consumer wiring. Unbalanced

² The word *compressor* as used here and in the next paragraph describes the audio process of dynamic range compression, described more fully in terms of its valuable post-production role in Chapter 10. Another, unrelated meaning of the word "compression" is data compression, a reduction in the number of bits used to store audio or video data. This second type of "compression" will be discussed below.



Figure 1-1 The presence of balanced input XLR connectors is the most readily apparent indication that the camera has separate microphone inputs. Note that other connector types may be used.

cables are usually associated with the RCA phono pin plug such as that found on the back of CD players, although other connectors, such as 1/8" stereo headphone or microphone connectors, also carry unbalanced signals.

Consumer and lower-end prosumer cameras might have balanced microphone inputs, but they are unlikely to have balanced line-level inputs, which are more often provided by unbalanced RCA phono pin plug connectors (Fig. 1-2). Sometimes these connections are even bidirectional: used as input for recording, they become outputs for playback. The Canon XL series, for instance, uses the bidirectional approach. More professional cameras with balanced line-level inputs³ may also offer line-level outputs via RCA connectors. The reason for including balanced line-level inputs is that it allows line-level signals to be sent from an audio mixer to the camera and recorded, giving the sound person the ability to control levels and mix multiple microphones during shooting, while line-level outputs allow the

³ Many cameras offer a single pair of XLR inputs that may be switched between mic and line level.



Figure 1-2 The ubiquitous consumer connector, alternatively called RCA, phono pin, or Cinch plug, and sometimes combinations of these.

sound crew to monitor the audio that's actually being recorded on camera. Although line-level outputs are convenient for monitoring and playback, prosumer cameras often provide only an 1/8" stereo headphone jack for monitoring, which may be used by the camera operator to listen in headphones or as a monitor return.

Once an audio signal has made its way into the camera, it must be digitized by an analog-to-digital (A/D) converter and recorded as a digital bitstream alongside the video. The ability to record uncompressed PCM audio is important in all but the simplest applications, and at any rate the audio recording format and quality of the on-camera audio circuitry must be transparent enough so that recordings play back with no audible distortion caused by the on-camera encoding. Many consumer camcorders—not to mention “home video” devices such as point-and-shoot digital cameras, mobile phones, and tablets with movie modes—use lossy data compression to reduce the bit rate or file size of recorded audio.⁴ The most

⁴ This is the second use of the word “compression” mentioned above. Lossy compression or encoding refers to a process that reduces the bit rate or file size of audio data by

common encoding format for low-end consumer camcorders is 2-channel Dolby Digital encoding, also known as 2-channel AC-3. In order to minimize the effects of compressing the audio data, AC-3 uses perceptual coding, which “throws away” the parts of the sound least likely to be noticed by human listeners and is an effective method of reducing file size for distribution on film, Blu-ray, or DVD.

The problem with recording lossy audio on camera is that using perceptual coding on recordings that will be edited or mixed later on compromises the quality that can be achieved during postproduction. After recording audio with a lossy codec, any further processing necessitates decoding the audio, processing the raw audio waveform, and re-encoding the audio for storage or distribution. Since information is thrown away each time the audio is encoded and for each track of recorded audio that makes its way into the final mix, artifacts that were at first benign quickly accumulate and become noticeable in the same way that “generation loss” occurs when copying analog recordings. Moreover, too much data reduction can produce audible artifacts such as metallic-sounding ringiness or audible distortion at high frequencies (splashy cymbals and sibilants), and without a full-quality original recording there is nothing to fall back on; the audibly distorted sound must be used or the clip re-recorded for use in the final program.

Generally, the ability to record uncompressed waveform audio (usually called PCM, short for “pulse code modulation,” which is a method of representing the analog waveform as a digital signal) is necessary for most of the applications discussed in this book. Higher-end consumer and prosumer camcorders and professional digital video cameras usually record two or more channels of PCM audio. Both the video and audio recording formats of a given camera should be listed in the manufacturer’s specifications for that camera, and the audio format will usually be listed as either linear PCM or a lossy format such as Dolby Digital / AC-3. Most standard- and high-definition tape formats record uncompressed PCM audio, although HDV (sometimes referred to as MiniDV HD) may use lossy encoding, depending on the recording mode.

ignoring certain parts of the data, so that the original audio waveform can be approximated but not precisely reconstructed during playback.

The Four Dimensions of a Sound Track

Normally, the word dimensions refers to the three directions we're familiar with in space, or to distances along those directions (the dimensions of a box). More generally, a dimension is an axis used to pinpoint or identify an object. In the case of sound, there are several characteristics that are useful in thinking about how sound works, how its representation works when recorded on a medium, and how it is perceived. Because these properties span the range from strictly objective measures into purely subjective ones, they provide a valuable method of thinking about sound technology and design at one and the same time. We call them the four dimensions of sound: frequency range, dynamic range, space, and time.

1. *Frequency range.* Perceived as pitch from deep bass to high treble, frequency range has an objective basis—the frequency in cycles of a sound wave per second of time, in the units Hertz (Hz). It is also at the root of much subjective thought about sound: a low-frequency rumble during a clear summer afternoon means there is a threat around the corner, a thunderstorm, that may endanger the characters in a story.

The frequency range of sound perceived by humans runs from about 20 Hz to 20 kHz (20 thousand Hz), although these numbers are taken as averages of large numbers of listeners; some young people can hear out to as high a frequency as 24 kHz (Fig. 1-3). At the other end of the spectrum, sounds below 20 Hz are not usually heard, but can be felt as vibrations if strong enough. Various devices transmit or reproduce only portions of the audible frequency range: a telephone delivers a narrow frequency range, from about 300 Hz to 3 kHz, so lacks deep bass and high treble. This is why the letters “s” and “f” pronounced by a talker are easily confused over a telephone. The “s” sound lies principally at a higher frequency than the “f” sound, and the telephone cuts off the higher frequencies, making the two sounds hard to distinguish.

2. *Dynamic range.* This is usually given as the range in amplitude from the softest audible sound, called 0 dB Sound Pressure Level (SPL), to the loudest sound tolerable, which is usually considered to be 120 dB SPL (Fig. 1-4). Interestingly, older treatments call 120 dB SPL the threshold of pain, whereas because some stadium concerts and raves can operate at even higher levels, such levels are nowadays called the threshold of tickle in the ear; 140 dB SPL is now considered the threshold

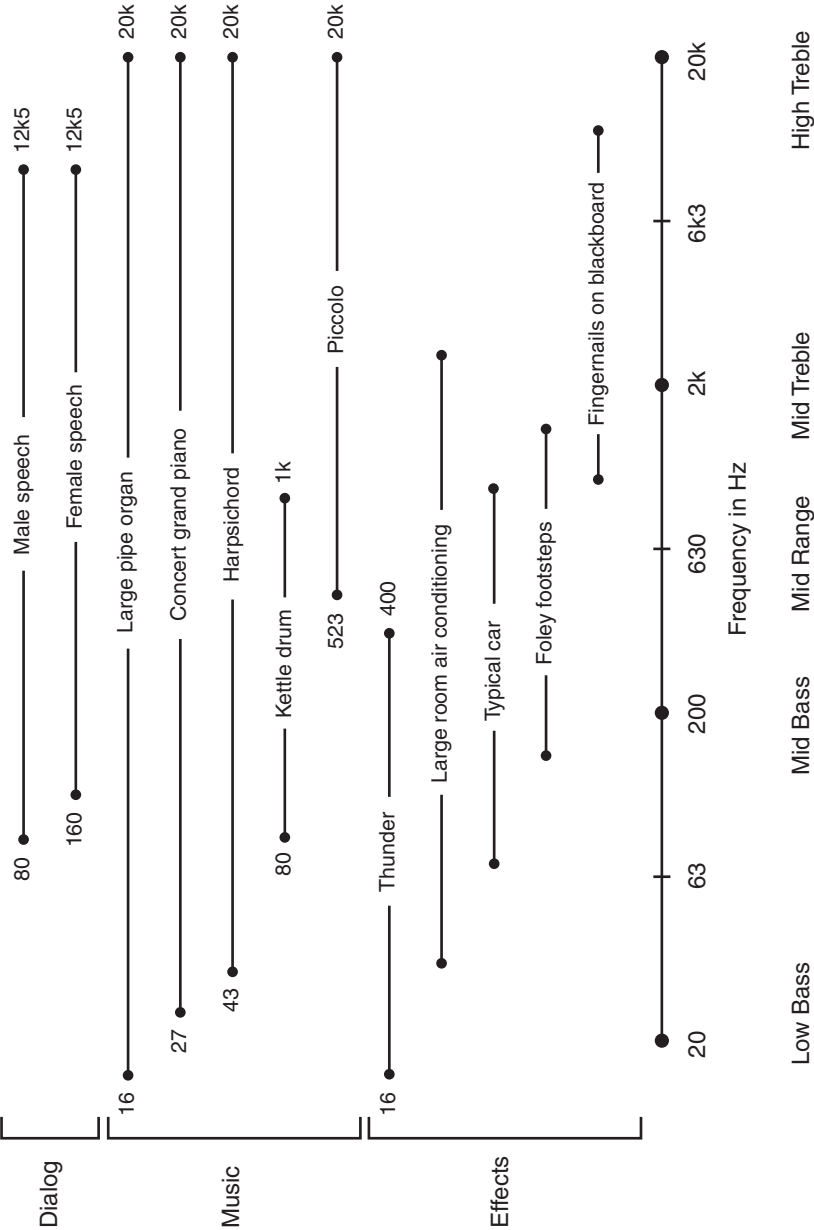


Figure 1-3 Frequency range, and some common sounds occurring within the range.

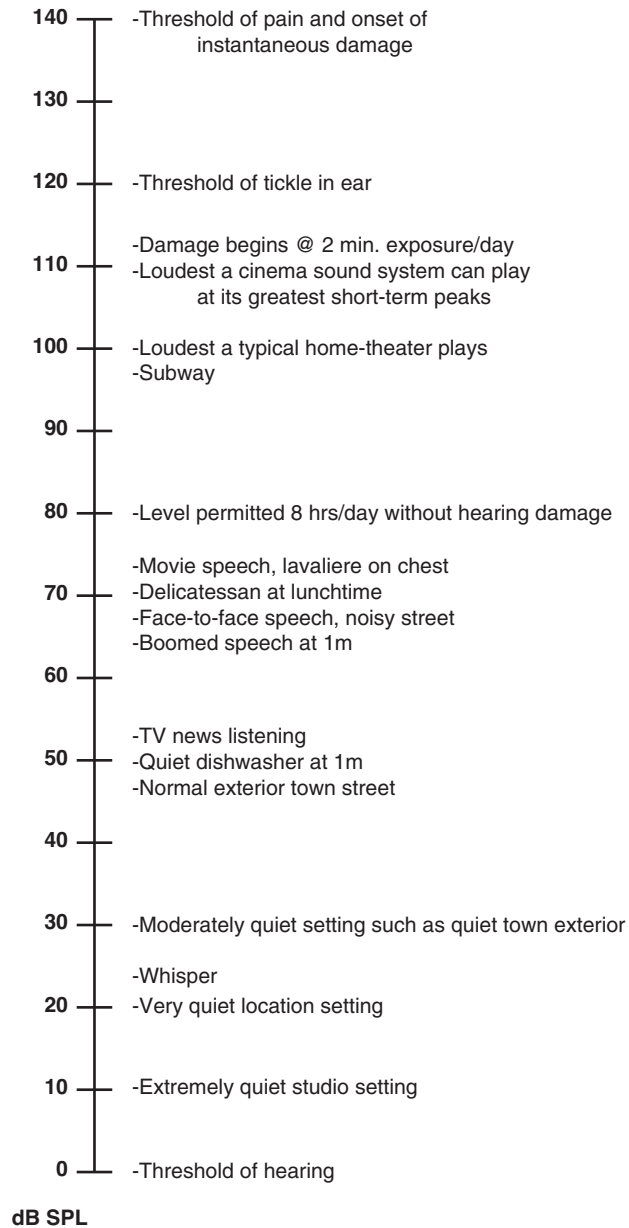


Figure 1-4 Average sound pressure level of typical sounds.

Table 1-1 Permissible daily noise exposure.⁷

dB SPL, L_{Aeq}	Daily Time Tolerable
88	4 hrs
91	2 hrs
94	1 hr
97	30 min.
100	15 min.
103	7.5 min.
106	3.75 min.
109	1.875 min.

of pain, but it is also the point at which permanent damage may occur even for brief exposure. While the SPL scale is designed to represent the range of human hearing, physical devices for transmitting and recording sound may have less dynamic range than the human ear. When used in that sense, dynamic range means the difference between the loudest and softest sounds a device or system can deliver. A telephone loudspeaker, for example, cannot produce a sound pressure level as high as the maximum level attained by a cinema sound system or heard at a loud rock concert.

It's a good thing telephones don't sound as loud as theater or concert speakers, given the potential for hearing damage to occur during exposure to loud sounds. Table 1-1 shows the typical daily exposure to various sound levels that most people can tolerate without significant risk of hearing damage. A level of 85 dB SPL,⁵ about twice as loud as normal speech in a movie theater, is tolerable for eight hours per day without long-term consequences to hearing,⁶ but louder sounds are tolerable only for shorter times.

⁵ Several items affect the sound pressure level readings used here: (1) frequency weighting, which gives more emphasis to frequencies to which the ear is more sensitive; (2) time weighting of one second, which is thought to match hearing damage more than a shorter time constant (faster metering); and (3) averaging of the one-second intervals over time, called equivalent level. Thus, in technical terms, what is described when speaking of hearing damage potential is called L_{Aeq} .

⁶ Except, perhaps, for the 1 percent of the population that is most susceptible to hearing damage from noise exposure. See ISO Standard 1999.

⁷ Ibid.

Decibels (dB) are a logarithmic scaling of the physical amplitude or size of the sound wave used to accommodate the very large dynamic range of hearing. The softest sound one can hear, 0 dB SPL, is a million times smaller in amplitude than the loudest sound most of us can stand, 120 dB, but a scale dealing in numbers from one to a million is inconvenient. By compressing the absolute number scale logarithmically into decibels, one gets easier numbers to deal with. The character played by Walter Pigeon, Morbius, explains a log scale in *Forbidden Planet*, which is worth watching just for this scene. The meters on the wall behind him in the Krell inner sanctum show the power that is being used by the planet-sized machinery beneath him. He says, "Ten times ten times ten times ten, raised almost to infinity," explaining that each meter represents ten times the power of the previous meter. So it is with dB. Each meter on Morbius's wall represents 20 dB (10 times in amplitude), and a 120 dB range (or six meters on the wall) represents the whole of the human hearing range from just audible to the sensation of tickle inside the ear.

Dynamic range is used for artistic purposes by filmmakers all the time. James Cameron uses the extremes to great effect in *Terminator 2*: the playground just before the atomic explosion is made very quiet to exaggerate the extreme loudness of the explosion that follows. The primary consideration in setting and adjusting the level of a signal is to control the level range of the source to fit it within the capability of the medium. Adjusting level continuously on a knob or fader is known as riding the gain of the program to be recorded, and is most common in postproduction mixing, but may be necessary to constrain the recorded sound to the dynamic range capabilities of a medium in situations with large variations between loud and soft sound. One of the worst failures in production sound or re-recording is to record with excessive distortion (too high, over-recorded) or noise (too low, under-recorded) because this factor has been overlooked or the level has been set incorrectly.

3. *Space*. This dimension of sound corresponds to localization in the usual three dimensions of space, but translates to sound reproduction systems as the number of speaker channels used to represent a sound field. Conventionally represented on video as one-channel *mono* versus two-channel *stereo* recording, today that dichotomy is extended to include surround sound, such as 5.1 or higher multichannel recording, which is prominent in cinemas, on DVD and Blu-ray discs, and even

in other media such as video games. Two main effects arise when more than one audio channel is available on a release medium: sound localization and spaciousness. *Localization* means that you can point at the originating direction of a sound in space, and sound can be made to match the position of the corresponding visual source (or not, according to the aesthetic for a particular style of sound). Localization of sound with on-screen visual elements usually gives both sound and picture greater verisimilitude. *Spaciousness* means having a sense of the space in the recording. When reproduced over just two channels, spaciousness derived from stereo can provide a sense of listening into a space, perhaps from the next room with an archway between the originating space and the listener. Surround sound provides a means of extending spaciousness to envelopment, the sense of being surrounded, of being in the same space as the source. Although high-end film and video production employs 5.1 discrete channels of audio to bring about these sensations of localization, spaciousness, and envelopment, nonetheless media with just two tracks can do a good deal of the same thing. For example, even two-channel media can be used to store surround information with the right production techniques and the use of an LtRt decoder such as is found in cinemas (Dolby Stereo and others) and homes (Pro Logic decoding and others) on playback.⁸

On the production side of the equation, though, the use of left-right stereo original recording, while available on many cameras with on-board microphones and digital SLRs with stereo microphone inputs, is not widespread. The two recorded channels are often used instead to record two different perspectives on a scene, such as a boom mic and a radio mic on the principal actor. Such dual-mono is not the same thing as stereo, because the term stereo implies that the channels are related spatially in a particular way, generally as a pair intended for reproduction over left and right loudspeakers during playback.

4. *Time*. How sounds are juxtaposed and evolve in time is the province of editors, who may simply maintain what was obtained at the time of shooting or may manipulate time across multiple tracks to

⁸ Chapter 11 covers the issues of surround 5.1 versus LtRt surround sound.

great effect. Editing and manipulating sounds in the time domain will be taken up later in the text.

Digital Sound

Whether digital sound is good enough for a given purpose is typically determined by the properties of the signal at the time of conversion from analog to digital. Although the quality of the analog signal being converted is certainly all-important and will be discussed later, three factors alone set the maximum quality level that can be achieved with digitized audio, and these factors grow out of the dimensions of sound directly. They are frequency range, dynamic range, and the number of audio channels. These three items are represented on cameras and recorders as (1) the sampling rate of the digital audio, f_s ,⁹ which is proportional to the frequency range; (2) the word length, also known as bit depth or resolution, which is proportional to the dynamic range and is typically given as a figure such as 16- or 24-bits; and (3) the number of channels that can be recorded on the medium simultaneously, usually two or four but occasionally more.

The number of independent channels is the simplest of these three concepts to understand. Most cameras, including tape-based models, record two or four channels, although some may record as many as eight. In double-system recording, two- and four-channel recorders are also common, but there are systems on the market to record six, eight, ten, or other numbers of channels, and with the right technology a modern laptop computer can even be used on set to record several channels of audio simultaneously. On fixed bandwidth media such as DV tape formats, the choice of whether to record two or four channels may trade off against both the sample rate and the word length so that only some combinations of the three parameters are available, but higher quality formats like HDCAM tape are capable of recording four channels of full bandwidth audio at both high bitrates and full frequency sample rates. File-based recording on hard disks or solid-state memory cards is generally not limited to a constant bandwidth the way tape-based recording is, so cameras that record to memory cards generally specify two- or four-channel audio recording

⁹ Also notated (incorrectly) as F_s by some camera manufacturers.

without trading off bit rate and sample rate against the number of channels recorded. The constraint that comes into play with four-channel recording is that many cameras that support it have only two external microphone or line-level inputs, so two of the four channels are limited to recording the on-board microphone. To overcome this limitation, some camera manufacturers supply mountable accessories to add additional XLR inputs and allow four channels of external audio to be recorded on camera.

The sample rate determines the high-frequency limit of sound that can be recorded. The rule is that the sample rate must be twice the highest frequency to be recorded, plus a little bit. Since the frequency range of hearing extends to about 20 kHz (or as high as 24 kHz in some people), the most common sample rate for audio in film and video is 48 kHz.¹⁰ Digital audio recorders and computer audio interfaces today advertise the capability of playing back audio at higher sample rates such as 96 kHz and 192 kHz. Practically speaking, the difference between sampling audio 48,000 times per second and 96,000 or more times per second for video applications is minimal since the frequency range of hearing can already be accurately represented by 48 kHz recording. On the other end of the spectrum, the standard-definition DV tape formats that trade off number of recorded channels versus sample rate and bit rate (such as MiniDV camcorders) may restrict four-channel recording to a sample rate of 32 kHz, which limits the maximum frequency of sound that can be represented to just under 16 kHz. Because some humans hear past 16 kHz, even to as high as 24 kHz, there is an audible limitation for some listeners when sampling at the 32 kHz rate, but because people hear over a broad range of frequencies, losing just a part at the very top of the frequency range rarely causes much trouble. Even though the sample rate is given prominence in advertising recording features, selecting recording modes in camera menus, and so forth, the word length or bit depth a system can record is usually a more important factor to the quality of recorded audio, both in current and future development of high-end systems and in assessing the limitations of low-end recording formats.

¹⁰ Sampling the amplitude of the signal 48,000 times per second.

Table 1-2 Trade off of audio channels versus sample rate and word length for certain digital video tape formats: DV (MiniDV), DVCAM, and Digital 8. (Not all cameras support both modes, and some cameras will allow overdubbing on extra channels but will not record four channels simultaneously. Some cameras will also play back content recorded at 44.1 kHz, 16 bit, the standard sample rate and word length of CD audio.)

No. of Channels	Sample Rate	High-Frequency Limit	Word Length ¹¹	Theoretical Dynamic Range ¹²
2	48 kHz	<24 kHz	16 bit	93 dB
4	32 kHz	<16 kHz	12 bit	69 dB

Represented on media, dynamic range runs from the point of noise on the quiet, small end of the decibel scale to the point of unacceptable distortion at the loud end of the scale. Most audio recording today uses at least a 16-bit digital representation of dynamic range, although the principal disadvantage of operating older DV tape formats in four-channel mode is the limitation to 12-bit recording as shown in Table 1-2. A *bit* is the fundamental unit of digital representation, a 0 or a 1, and is the building block of any piece of data stored and used by computers.¹³ In binary arithmetic, numbers are represented using only the digits 0 and 1 instead of the more familiar 0 through 9 of base-10 arithmetic. Since fewer symbols are available to represent a number in base-2 (binary) than in base-10 (decimal), the binary representation of a number will contain more digits (0s and 1s) than the decimal representation when written out. In digital audio, each bit of word length delivers approximately 6 dB of dynamic range, so you can multiply six times the number of bits to get the dynamic range of a recording format. With each bit delivering about 6 dB of range, 16-bit audio has roughly 96 dB range. This is only an estimate, because a small amount of deliberately added noise called dither is needed to smooth the steps in the digital structure of a waveform that would otherwise cause low-level distortion, so the actual dynamic range capability of 16-bit recording is closer to 93 dB. To cover the full capacity of human hearing

¹¹ Also known as resolution and bit depth.

¹² With dither, explained later.

¹³ The word *bit* is a portmanteau, meaning *binary digit*.

from 0 dB SPL up through 120 dB SPL, about a 20-bit word length is needed, but in most practical recording situations the background noise level is far higher than 0 dB, and the loudest sound does not approach 120 dB, so a 93 dB range (16 bits) is adequate most of the time.

The advantage of moving to greater word length, such as 20- or 24-bit recording, is that once the maximum recordable level is set using volume controls on the camera or recorder, the theoretical noise floor of the recording medium is much lower than it would be for 16-bit recording: the extra 4 or 8 bits add resolution at the bottom end of the scale. In particular, this gives the operator more leeway before running into the digital noise floor of the recorder if the record level is set slightly too low; as long as the acoustical noise floor of the sound being recorded remains above the noise floor of the 20- or 24-bit recording system, the signal can be turned up in postproduction without a perceptible increase in noise on the final soundtrack. However, practical considerations such as the noise floor of microphones, preamplifiers, and other electronics must also be taken into consideration to ensure that a sound is loud enough to be “heard” by a microphone and that the signal is not turned down so low that it blends into the noise floor at any point in the chain before being recorded. Gain staging, or setting levels to fit the dynamic range required for a shot onto the recording medium, is covered in more detail in Chapter 6.

While 24-bit recording has its benefits in various situations, two-channel recordings made at 16-bit word length and delivering a theoretical 96 dB dynamic range (less on practical cameras¹⁴) are still adequate for most shooting situations, even though they fall short of the full dynamic range of hearing. This problem is reduced by the facts that for most productions we rarely encounter sounds so loud as to stress the threshold of pain, and most environments we record in are not nearly as quiet as the softest sound we can hear, limiting the range we have to capture at the source. Thus, for most practical purposes, 16-bit recording is sufficient for production sound accompanying video.

The disadvantage of using a 12-bit word length in four-channel DV tape recording (see Table 1-2) is that 12-bit recording delivers only about 72 dB of theoretical dynamic range, without accounting for the effects of dither.

¹⁴ This is for their line-level inputs, not microphone inputs, and certainly not for on-camera microphones, which will be discussed later.

Thus 12-bit audio falls far short of capturing the full range of audible sound. This means that if the loudest sound present in a program is recorded such that it does not exceed the highest level that can be represented without distortion—in other words, the recording level is set right—and then played back at a standardized playback level, noise in the form of hiss will probably be audible, certainly in a quiet environment. Perhaps for shooting situations where the acoustical background noise level of the environment is high, and the source being recorded is not too loud, this might be adequate, but it remains inadequate for faithfully capturing the dynamic range present on most sets and locations.

The bottom line of this discussion is that most pros would hardly ever record audio at less than 16-bit, 48-kHz sampling, and that the most pronounced gains in quality for certain shooting situations can be achieved by increasing the word length to 20- or 24-bit, 48-kHz recording. Usually, a standard word length and sample rate should be selected for a project and used for each tape or file recorded for the project, since recording audio in multiple formats for a single project necessitates converting all the files to a common format before beginning the sound editing process. So, when recording audio for a new project, the first thing to do is to decide what format to record in and to find the audio settings menu on the camera or audio recorder, ensuring that a 48-kHz sample rate and the appropriate 16-bit or higher word length are selected. If a camera has the ability to record uncompressed PCM waveforms or to encode audio in a lossy format such as MPEG Layer II audio or Dolby Digital, lossless PCM recording should also be selected. Usually two channels of audio will be recorded for digital video, but higher-end formats may record more high-quality channels, mainly because of the space available on the medium or the number of inputs and internal audio processing channels present on the device. The capabilities of several formats and families of cameras are detailed in the following section.

Features of Digital Video Recording Formats

In addition to production sound recording on a variety of media, the channel information for tape formats in Table 1-3 applies to postproduction exports for projects that finish on a particular format. For example, a 5.1 surround mix cannot be recorded to an HDCAM master because only four audio channels can be recorded on an HDCAM tape. The usual method

Table 1-3 Audio capacity of various digital formats.

Channels	Sample Rate	Word Length	Encoding	Format
2	48 kHz	16 bit	Linear PCM	DV (a.k.a. MiniDV)*
4	32 kHz	12 bit	Linear PCM	DVCAM* Digital 8*
2	48 kHz	16 bit	Linear PCM	DVCPRO
4	48 kHz	16 bit	Linear PCM	DVCPRO 50
4	48 kHz	20 bit	Linear PCM	HDCAM
8	48 kHz	16 bit	Linear PCM	DVCPRO HD
2	48 kHz	16 bit	MPEG Layer 2	HDV (a.k.a. MiniDV HD)
8	48 kHz	16 bit	Linear PCM	Sony XDCAM formats
4	48 kHz	24 bit	Linear PCM	
2	48 kHz	16 bit	Dolby Digital AC-3	Typical consumer HD camcorders
2 or 4	48 kHz	16 or 24 bit	Linear PCM	Typical prosumer or professional HD cameras
2	48 kHz	16 bit	Linear PCM	Most digital SLRs
2	44.1, 48, or 96 kHz	16 or 24 bit	Linear PCM or MPEG Layer 3	Small portable audio recorders (Zoom, Tascam, Marantz, and others)
2, 4, 6, 8, 10, or other, depending on device	44.1, 48, 88.2, 96, or 192 kHz	16 or 24 bit	Linear PCM ¹⁶	Professional audio field recorders (Sound Devices, Fostex, Zaxcom, and others)

*These formats can also play back two-channel 44.1-kHz 16-bit audio (the CD standard) when recording is made by way of a FireWire interface.

for recording a surround mix on HDCAM is to record a two-channel LtRt mix¹⁵ on channels 1 and 2 of the HDCAM tape, even if a full 5.1 mix is also made for inclusion on DVD and Blu-ray copies of a film. This and other finishing considerations are discussed in Chapter 11.

¹⁵ See Chapter 11 for more information on mastering and distribution, including a description of LtRt encoding.

¹⁶ Other settings may be available, but uncompressed linear PCM would typically be used for professional applications.

Note that individual equipment may not address all of the possible variations that can be recorded within a format; for instance, some camera models may not have the capability to record all of the channels that the format has space for on the medium. However, as a general matter, the data shown in Table 1-3 applies.

**The Director's Cut
(or what the technician hopes that the director knows)**

- Choose a camera with external microphone input capability and manual control over recorded level.
- Select a camera or recorder that records uncompressed linear PCM audio rather than encoding production audio in a lossy format like Dolby Digital or MPEG formats.
- Don't mix up digital audio settings; choose a format at the beginning of a project and stick to it. Typically this will be 48-kHz, 16-bit or 24-bit, two-channel recording.
- Watch your own exposure to loud noise, especially covering events such as concerts, car races, industrial sites, and the like.

This page intentionally left blank

CHAPTER 2

Introduction to Digital Video

This introduction to digital video technology will describe some of the fundamental common features of all of the formats that make up digital video, as well as some of the distinguishing features among them. Chapter 1 has covered the fundamental sound differences, and to delve more deeply we will have to venture into the picture side of the equation, because the distinguishing features among recording formats are often related to the picture. Although there are some distinguishing sound differences among digital video formats, one would have to say that they are smaller than the range of picture qualities. Thus a practical-sized sound book can cover a lot more formats than could a similarly detailed picture-related book, and learning about just a few distinguishing issues for sound formats makes it possible to cover the whole field.

The rapid development of digital video over the past several years has set the stage for a number of new workflows and recording paradigms to emerge, principally involving the refinement of video compression schemes and the proliferation of tapeless media such as memory cards, optical discs, and hard disks. What these systems share with the older tape-based formats that first made digital video viable is the method of recording digital signals to represent distinctly analog phenomena: traveling waves of sound and light. Moreover, many of the basic techniques, compromises, and audio trade-offs are the same for tape-based and newer file-based cameras. What newer file-based systems add is a non-linear recording paradigm, extending the non-linear, random-access nature of digital editing to the world of production video capture. The process of digital recording, the major characteristics of digital video, and the peculiarities that inform sound recording for digital video projects will be covered in turn.

Basic Digital

The basic process of capturing picture and sound for all digital formats involves conversion of the continuously varying signals produced by the image and sound sensing devices into numbers. Digital video cameras use one or more image sensors, typically charge-coupled devices (CCDs) or complementary metal-oxide silicon (CMOS) sensors, to convert the varying brightness of the three-color components of light striking the sensor into a numerical value for each pixel. For sound, the moment-by-moment amplitude of the sound pressure level versus time is converted by one or more microphones into an electrical voltage versus time, which is then represented as a series of numbers. In each case, it is the amplitude of the signal that changes in time and is recorded as a numerical value at specific moments in time, once per frame for the video and typically thousands of times per second for digital audio. The reliability of storing numbers rather than directly recording the smoothly and continuously varying signals that constitute analog phenomena is a primary justification for digital techniques. The advantage of the digital technique is that copies are literally clones of their masters if certain prerequisites are met: copies are then just as good as the original. This is never true with analog techniques, where each stage of copying exhibits some inevitable generation loss, blurring the variations in detail and tone that made up the original work the way a copy machine introduces slight imperfections to a picture through the generations.

This difference between analog and digital reproduction makes digital techniques well suited to distribution channels, where getting it right for large numbers of copies is important. That is why the impact of digital methods on the sound industry in particular began with the growth of a distribution format, the CD (compact disc). Later it was found that digital techniques were valuable in sound *production*, not just *replication*, once they were made to exhibit all of the qualities necessary. In digital video, the first users were high-end postproduction companies that needed to maintain quality among the generations needed to make video masters. Later in the development of digital techniques came the explosion caused by the near-simultaneous introduction of DV and DVD for capture and distribution. Ultimately, digital video introduced many advantages as it began to replace older analog workflows. For example, digital processing makes it possible to select one particular color in a scene—say that of the grass skirts of Hawaiian dancers—and change it from green to chartreuse without affecting anything else in the scene, something that conventional film chemical and exposure techniques could never achieve.

Another advantage that has helped digital video solidify its position in the marketplace is cost. Digital camcorders and cameras have effectively ousted analog video cameras from the market since the turn of the century because of the tremendous quality that can be achieved by digital means at relatively low cost. Compared to shooting on film or high-end digital cinema formats, high-definition digital video is eminently affordable, offering high-quality image and sound recording at great value. In particular, cameras with professional audio features such as external microphone inputs and manual level controls allow high-quality sound to be recorded alongside the picture, while growth in the digital audio recording market has made double-system recording viable for many independent productions shooting on DSLRs and other cameras without professional audio capture built in. So digital technology, as it develops, continues to bring the possibility of recording professional-looking (and sounding) programs to a wider base of people than ever before.

Some Basic Video for Audio People

Frame Rates

Film and video share a common feature: both produce moving images as a sequence of frames taken at uniform intervals through time that are typically not seen separately, but instead blend together into continuous motion through our persistence of vision. A purely mechanical 19th-century invention, the Zoetrope, shows that blending still images together through time can produce the illusion of motion.

Theatrical films have used a frame rate of 24 fps ever since the coming of sound forced standardization in this area, around 1927. This is a universal standard, used around the globe. Actually, in most instances each frame is shown twice on the cinema screen before being advanced to the next frame, so there are 48 flashes showing 24 unique pictures each second. The reason for this has to do with the perception of flicker. Shown at just 24 fps and at the brightness used in theaters, pictures seem to flicker (the term *flicks* used as slang to describe the movies probably originated from this effect).

U.S.-based NTSC video started out at 30 fps. The reason it could not be made 24 fps was the higher brightness of television compared to theatrical

exhibition. Even showing each frame twice still produced visible flicker at 24 fps, because human perception perceives more flicker at higher brightness, and television sets were routinely operated some two to six times brighter than theater screens. So the black-and-white television standard adopted was 30 fps. Films shown on television used a convention of showing some of the frames twice in some of the video fields (half-frames)¹, so that while the film traveled through the equipment at 24 fps, 30 video frames per second were produced. This is called a *3:2 sequence* (sometimes 2:3) and is illustrated in Figure 2-1.

Then came color. Color television standardization had the requirement that the color picture had to be fit into the same broadcast space on the

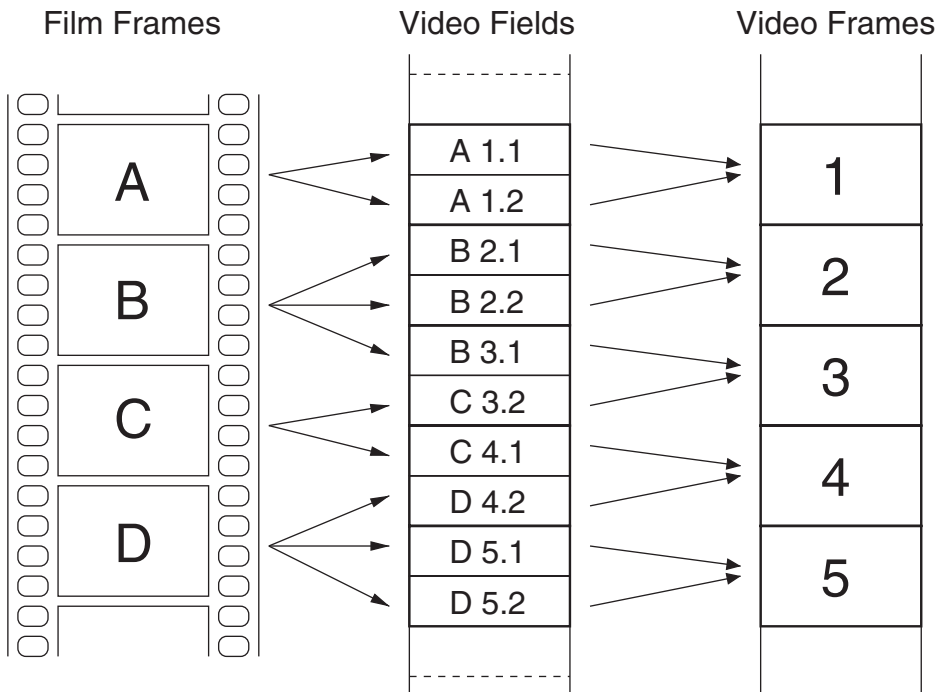


Figure 2-1 Films shown on television or transferred to video in the U.S. are fit to the higher frame rate of NTSC television by a process called 3:2 insertion, resulting in a sequence of images with portions of some film frames doubled.

¹ A field represents half a frame of interlaced video, typically the odd or even scanlines of a single frame. Video fields and the reasons for interlacing video are discussed in more detail below.

air—called *bandwidth*—which black-and-white television had been using. The color component of the broadcast, transmitted as a subcarrier, could not interfere with the black-and-white broadcast, nor with the sound. In order to accommodate all of these requirements, among other things the frame rate was changed slightly, to 29.97 fps,² for color broadcasts. An ordinary black-and-white set had no problem recovering a picture that was slightly off speed, and a color set could recover the color subcarrier and audio with minimal interference, so this rate was standardized.

In Europe, the standard adopted for color television was 25 fps, and U.S. theatrical films are portrayed on European television by being played 4 percent faster, so that 25 film frames are shown on the television each second. Audio played faster to maintain sync with the image sounds higher-pitched if untreated, so film to video transfers in Europe often have an attendant pitch increase in the sound. This effect is noticeable even on actor's voices, so it should be corrected in transfer through the use of a hardware or software device called a *pitch shifter*, but in practice only certain titles are treated this way to recover the correct final pitch. Whereas pitch correction of time-scaled material was formerly thought to introduce artifacts into the audio and was in limited use, recent advances in digital processing have improved the output quality that can be expected from professional equipment.

For film shot for theatrical release at exactly 24 fps, the conversion for postproduction editing to NTSC video involves the 2:3 field sequence of repetition to get to a nominal 30 fps, and a slow down on the telecine or film scanner to 23.976 fps, together called 2:3 (or 3:2) pulldown. This means that sound recorded double-system also has to be slowed down by the same amount to maintain synchronization. This is most often accomplished by simply playing the audio 0.1% slower than it was originally recorded, whether off an analog tape or a digital bitstream. For example, digital audio recorded on set at 48 kHz will only be played back during telecine or editing at a rate of 47,952 samples per second, and will often be resampled to match the 48 kHz sample rate of audio recorded in

² This number, while close to the exact value, is technically an approximation. The exact frame rate of NTSC color television and any derived formats called nominally 29.97 or 23.976 (or, less precisely, 23.98) fps is slower than 30 fps (or 24 fps) by a ratio of 1000:1001, so that 29.97 would be written without rounding as the repeating decimal value 29.970029970029....

postproduction in sync with the pulled-down picture. One way of avoiding the need to resample digital audio when film is pulled down to video speed is to record audio on set at 48,048 Hz (sampling 0.1% more often per second) so that, when copied into the postproduction editing environment and run at 48 kHz, it will maintain sync, having been slowed down the same amount as the picture.³ Unlike the effect of the 4% speedup of film transferred to 25 fps video, the slight change in pitch (0.1%) attendant with pulldown to NTSC video is not noticeable⁴ and need not be corrected.

The previous discussion demonstrates one of the difficulties of dealing with film originals in an environment that will include video postproduction. Video original for video post is easier, but the standards are less universal, and adapting to them is potentially more complicated. For instance, if you shoot NTSC video and wish to release internationally, conversion to PAL can be complicated, expensive, and imperfect. An emerging method of making video more universal is to capture 24 fps (using the 24P mode on a digital video camera) and treat the video like a film original, as explained later.

Interlaced Video

Most video cameras operate at 29.97 fps for NTSC or 25 fps in PAL countries in interlaced format. Interlaced scanning means that each frame of video is split into two fields, or half-frames, where the odd-numbered lines 1, 3, 5, and so on of each frame are scanned by the camera during recording (and by the television during playback) in one pass, followed by the even-numbered lines (2, 4, 6, ...) in a separate pass. Each video field

³ Most high-end portable production recorders offer 48,048 Hz in addition to 48 kHz in their sample rate settings, although it may be referred to as 0.1% pulldown or by some other name. They may also offer a 47,952 Hz or 0.1% pullup sampling rate for video projects that will be pulled up to film speed, although that situation would be far less common. When shooting on film and editing on video, it is important to make the decision whether to record audio at 48,048 Hz or 48 kHz before production and *stick with it*; mixing audio recorded at different sample rates will most likely lead to headaches during postproduction.

⁴ Even by those with perfect pitch. The minimum noticeable pitch difference for the 0.01 percent of the population with perfect pitch on an absolute basis (not comparing side by side) is about 20 percent of a half-step, or 1.2 percent.

is evenly spaced in time so that a camera recording 29.97 fps video is actually recording fields at 59.94 fields per second, and a 25 fps camera captures fields at 50 fields per second.

The reasons for interlacing the picture in early television systems were practical: refreshing the television display 50 or 60 times per second resulted in less visible flicker than only refreshing at 25 or 30 Hz, just as film projected with a double shutter to show an image 48 times a second flickered less than true 24 fps projection in theaters. But while film projectors illuminated an entire frame of film at once, television sets scanned images one row at a time, and the time required for the electron gun to scan each line meant that only half a frame of video could be displayed in 1/60th of a second at the desired frame size. Rather than settling for a lower vertical resolution and scanning each frame twice, the full vertical resolution of what we call standard-definition television was maintained and only half the image (a single video field) was scanned in one 1/60th of a second pass. Then the electron gun returned to the top of the screen and scanned the second field of the image in the next 1/60th of a second. This method of scanning and transmitting video was called *interlaced* because the even and odd scanlines, stored and transmitted separately on the medium, were interlaced back together visually during playback.

When high-definition television (HDTV) was introduced, the standards included both interlaced and progressively scanned (non-interlaced) modes of operation. During its adoption, HDTV existed (and still, to some extent, does exist) alongside traditional standard-definition content and distribution channels, so interlacing remained partly as a historical artifact of standard-definition systems. However, the practical advantage of increasing the effective vertical resolution that can be represented in a given bandwidth is another reason interlaced broadcast remains popular for some high-resolution signals: it is a space-saving measure. With cathode-ray tube displays largely supplanted in the marketplace by newer technologies, such as LCD and plasma displays, which don't rely on an electron scanning gun to illuminate an image pixel by pixel, true interlaced signals must be de-interlaced by modern displays. That fact alone may be a driving factor in progressively scanned video becoming increasingly common; indeed, progressive scanning is also considered desirable as part of the "film look" camera manufacturers advertise when promoting 24P modes on their products—the P stands for "progressive."

If that's the case, why do we say that most cameras still operate in interlaced format? The answer is threefold. First, many new cameras still offer the option of shooting standard-definition video at lower data rates, which is typically recorded in the U.S. as interlaced NTSC video. Second, since high-definition standards include both interlaced and non-interlaced format options, most camera manufacturers offer 60i (technically 59.94 Hz interlaced) or 50i recording in addition to 30p or 25p non-interlaced scanning and, often, a 24p "film" mode.⁵ Finally, even the progressive scan modes on most digital video cameras store images in an interlaced format on the medium, taking the even and odd scanlines captured at a single moment in time and storing them into the two fields of the image as if they were captured separately. This results in an image that does not need to be de-interlaced for display on a computer or modern television even though it is stored in an interlaced container format on tape or disc. When digital video cameras with progressive scanning first came to market, the existing video infrastructure and the DV codec family almost universally in use dictated that interlaced NTSC video be recorded on tapes, so 24p and other progressively scanned modes were fit into the existing infrastructure using techniques such as the 3:2 sequence familiar from film to NTSC transfers. The technique of fitting progressive content into an interlaced format persists even in tapeless devices today because, in addition to improving backward compatibility, it is easier to build a camera with a single recording format—interlaced video—and treat progressively scanned images as if they were interlaced.⁶

"Film" Look

There are a number of reasons video looks different from film, even when filmed images are transferred to video and viewed on television. One is

⁵ Here 50i and 60i refer to interlaced capture, with the number preceding the letter "i" indicating how many fields are captured per second. For NTSC devices, the image capture frequency is usually rounded up to the nearest integer, so that 60i is used when what is actually meant is 59.94i video. Similarly, 30p is usually used to represent 29.97 fps progressively scanned video, the "p" indicating that each image is captured as a complete frame at a single moment in time rather than in two separate fields per frame.

⁶ Treating an interlaced image as progressive would result in a "combing" effect on the image during motion since every pair of adjacent lines would have been captured at a different moment in time. On the other hand, a progressively scanned image will look correct even when treated as interlaced video and viewed on an interlaced display.

the size of the sensor area of a video camera compared to the area exposed on a film negative. The CCD or CMOS sensor on a typical digital video camera is much smaller than a full-frame 35 mm negative, so it is much harder to achieve shallow depth of field effects on video than it is with film. In other words, it is difficult to keep just the subject in focus and throw the background out of focus, a common technique in film cinematography, using video.

Another is the higher frame rate of NTSC video compared to 24 fps film, combined with the fact that video has traditionally been captured in interlaced format whereas film cameras expose the entire frame at once. To compensate for some of these differences, producers have increasingly relied on the “film look” of 24P for entertainment shot on video but intended to have the emotional and dramatic scope of motion pictures traditionally shot on film. As explained above, 24P also offers an emerging pathway to a more universal video master format than 25 fps or 29.97 fps modes, which require conversion for video to be transferred from one system to another.

The 24P mode on a video camera means that 24 frames of video are captured each second using progressive scanning, reading each line from the image sensor in order (1, 2, 3, 4, ...) at a single moment in time, as opposed to interlaced video captured one field at a time. One problem with interlaced scanning that progressive scanning mitigates is a lack of sharpness when the picture moves noticeably between the first and second fields of a frame, which would turn a moving vertical line into a zigzag, for instance. If you look closely, you can sometimes see this effect on interlaced video when the camera pans and then comes to a stop: once the pan is over, vertical edges are sharper than when the camera was moving. The laser disc version of *Jurassic Park*, which uses interlaced video, shows this effect when the bad guy is stealing the genetic material of the dinosaurs from the vault. The camera pans over the vertical lines of bars in the vault, and before and after the pan, the vertical lines are sharp; but while the camera is panning, the vertical lines appear jaggy on a conventional television with interlaced scanning.

Another advantage that makes progressive scanning seem more film-like is the increased density of information each time the picture on a modern progressive display is refreshed. With each interlaced frame split into two fields 1/60th of a second apart in time, the lines written during the first half of an interlaced scan are fading by the time the

second half is written.⁷ We perceive this as lesser sharpness in interlaced compared to progressive scanning. Thought of another way, interlaced scanning produces one-half the scanning lines in 1/60th of a second, but progressive produces all of the lines in the same amount of time, so more information is presented per unit time by progressive scanning, and this is seen as better sharpness.

Even though it is called 24P, the actual rate at which frames are captured is 23.976 fps on most camcorders in the U.S. because the video is encapsulated into a 29.97 fps interlaced signal for recording, as described above. The two fields of interlaced video recorded on tape can be put back together in digital editing systems to reproduce the progressively scanned image as it was originally read off the camera's image sensor, so the benefits of progressive scan video are preserved by this arrangement. With video editing software set to correctly recover the 24 original frames out of every 30 recorded frames, 24P footage can be edited directly at 23.976 fps, and the only practical difference is the 0.1% slowdown from film to NTSC speed (24 fps to effectively 23.976 fps) caused by the camera clock running at NTSC speed.⁸ This speed differential can be reversed by "pulling up" the video to film speed in order to create a true 24 fps film print, with a corresponding time correction to the audio accompanying picture, or the video can be sped up by about 4 percent and output in PAL format for European distribution and to other parts of the world using the PAL system. Of course, the video can also be exported back to NTSC format tapes or discs by repeating fields in a 3:2 sequence like the camcorder would have during capture or a telecine would have for a film transfer to video, and this output can be used with equipment based on the NTSC system traditionally used in the United States, Japan, and other parts of the world. Thus 24P camera originals become source masters useful for

⁷ This was especially a problem in CRT displays using a scanning electron gun to illuminate phosphors, although the principle remains true on newer progressive displays in that progressive content may be displayed all at once (and refreshed every 1/60th of a second or more) whereas interlaced content must be de-interlaced or shown field by field to appear correctly on a progressive scan monitor.

⁸ The alternative to editing at 23.976 fps is editing the 29.97 fps video stream as if it were an NTSC video original, but this eliminates many of the benefits of shooting in 24P since interlacing artifacts introduced by the 3:2 sequencing of frames will then be present in the edited program, and cuts may not occur precisely on 24P frame boundaries. When shooting 24P, it is better to use editing software capable of interpreting and editing 24P material directly.

film release and for transfer to both of the world's major video standards, kind of a Great Grand Master.

While most camcorders and even DSLRs with a 24P mode actually record 23.976 video frames per second,⁹ some professional cameras like the Sony F900 and various offerings from RED Digital Cinema allow true 24 fps recording in addition to 23.976 fps shooting for compatibility with NTSC video. When recording at one of these speeds and playing a tape back at the other, or playing a file back at the other speed with audio pulled up or down in software to match, this difference is not noticeable. However, if the sound is separated from the picture and the assumption about speed is made incorrectly, the audio will no longer sync with picture because of the 0.1 percent speed difference. In that case, audio recorded or reproduced at the rate opposite that of the camera will go out of sync visibly within a little over a minute. This is one of the many sources of potential problems for audio, because any sync problem that occurs is almost invariably considered an audio problem even if it is the picture that is running off the stated speed. So it's important to be aware that consumer and prosumer cameras without separate 23.976 and 24 fps settings may well be operating at 23.976 fps even when they say 24 fps or refer to 24P in menu settings or operating manuals.

Now that 24P recording has become commonly available, the field of digital video is developing to more closely approximate the look of film and high-end digital cinema in other ways as well. One of the driving factors behind the adoption of digital SLRs that shoot video has been the possibility of capturing higher-quality images for a fraction of the cost of high-end digital video cameras. When shooting video on a DSLR, videographers have access to the same range of interchangeable lenses used by still photographers on a particular camera. The image sensors in DSLRs with professional-quality movie modes are also larger than their counterparts in most digital camcorders, allowing greater range of depth of field as discussed earlier and what many consider a more artistic look. Some DSLRs, such as the Canon 5D, have so-called "full frame" sensors matching the size of a 35 mm still frame—quite a bit larger than a 35 mm motion picture frame due to the orientation of the image on the negative. Still others, like the Canon 7D, more closely approximate the size of a 35 mm motion picture negative. And even outside the world of DSLRs,

⁹ $24 \times 1000/1001 = 23.9760239760...$ to be exact.

higher-end digital video cameras like the Sony F3 now offer sensor sizes comparable to traditional motion picture film frames. Many purpose-built digital video cameras in prosumer and professional lines can also be used with a variety of cinema lenses via 35 mm lens adapters, and the features brought to the table by DSLRs promise to spur future development in this area of the prosumer digital video market as the artistic capabilities of digital video continue to grow.

Resolution and Aspect Ratio

For decades, the most obvious distinguishing feature of film shot for theatrical distribution was its widescreen aspect ratio compared to the narrower 4:3 aspect ratio of video and television broadcast. That's one reason the 16:9 widescreen aspect ratio was introduced as an integral part of HDTV and high-definition video standards, and continues to be adopted even in emerging markets like mobile and Internet video platforms. Widescreen images can be produced in several ways. One is simply cutting off the top and bottom of the picture with black bars, but this limits resolution of the remaining picture because of the loss of 25 percent of the image area in pixels at the top and bottom of the frame. A second way is to use a squeeze lens, an anamorphic attachment that squeezes the image horizontally so that a 16:9 (1.78:1) picture fits onto a 4:3 (1.33:1) image sensor (Fig. 2-2), and then is unsqueezed in playback on a 16:9 monitor so that it takes up the entire widescreen frame. Both of these methods are used by various film formats: the 1.85:1 aspect ratio commonly used in the United States results from cropping off the top and bottom portion of each frame of a 35 mm print, whereas wider aspect ratios such as 2.35:1 rely on the use of anamorphic lenses.¹⁰ The third method of capturing widescreen video is to use native 16:9 chips in the camera, which is what is done today in high-definition cameras designed specifically for the 16:9 aspect ratio.

¹⁰ Anamorphic widescreen video is also widely known as a distribution format for wide-screen movies released on DVD, replacing the older method of letterboxing with black bars at the top and bottom of the frame. As in anamorphic capture, the goal in this case is to preserve greater vertical resolution by utilizing every pixel in the frame to store useful picture information.

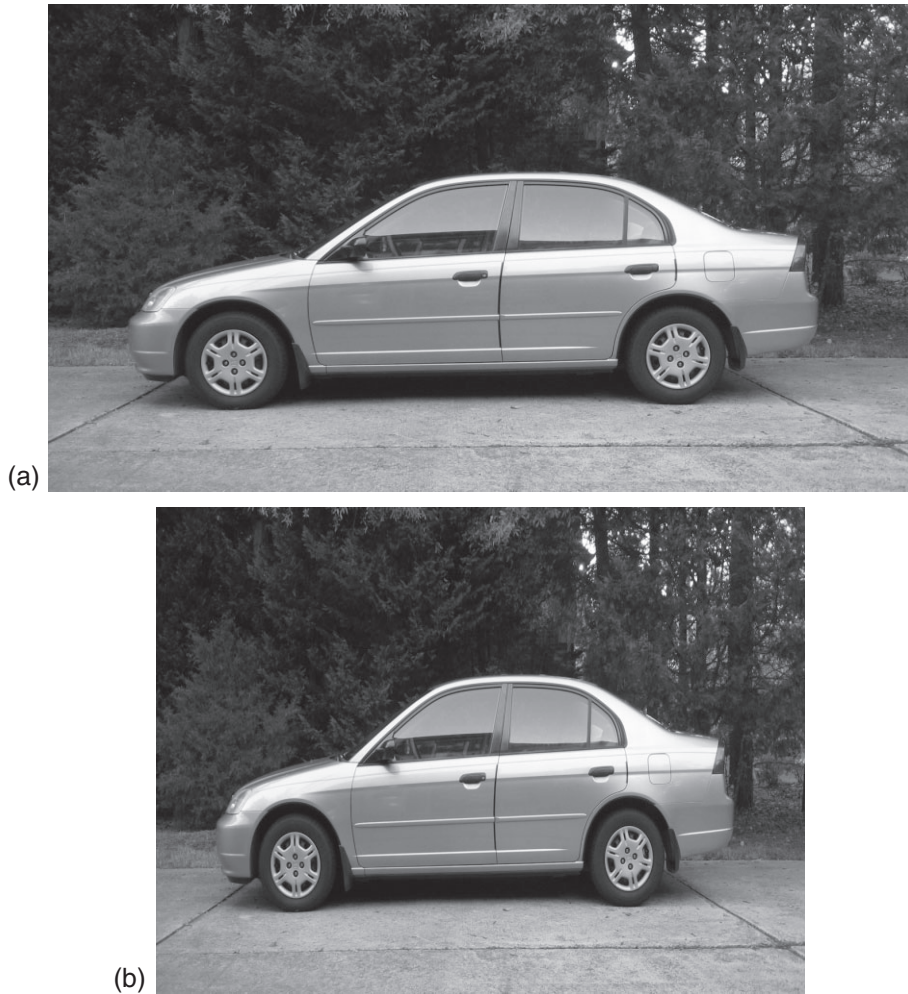


Figure 2-2 (a) An image in the 16:9 widescreen aspect ratio. (b) The same image shrunk horizontally by an anamorphic lens.

The original prosumer digital video formats (DV, DVCAM, DVCPRO) recorded video at a resolution of 720×480 pixels: 480 lines of video with the image sampled horizontally 720 times per line. Although there were more than 1.33 times as many pixel locations on the horizontal axis than on the vertical axis, the image was captured and intended to be displayed in the 4:3 aspect ratio; you could think of the pixels represented on a DV tape as non-square pixels, as shown in Figure 2-3. To shoot in anamorphic

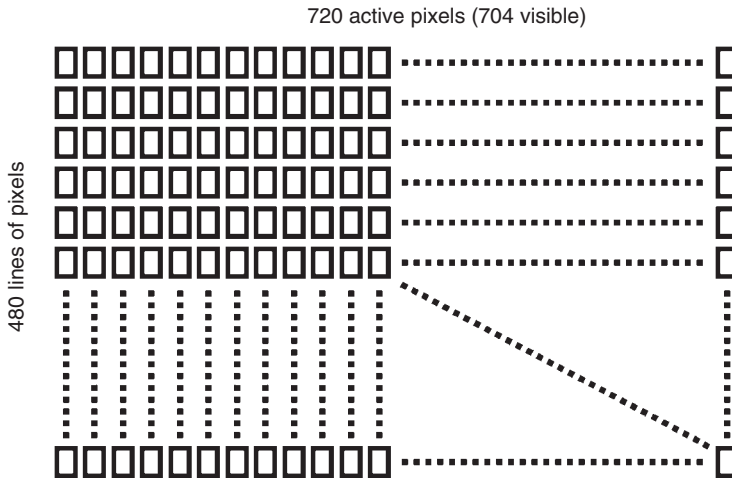


Figure 2-3 The arrangement of pixels in a 720×480 standard-definition DV frame. Notice that the density of pixels is greater along the horizontal axis than along the vertical axis.

widescreen on prosumer DV cameras, anamorphic lenses or, in some cameras like the Canon XL-2, 16:9 CCDs and digital processing were needed to fit the 16:9 widescreen image into the 720×480 pixels of standard-definition DV formats.

While standard-definition digital video uses 480 active lines, high-definition video may use 720 or 1080 lines, either progressively scanned or interlaced. Table 2-1 shows the different high-definition video standards in use in North America. While 1280×720 video can be stored or broadcast at 60 full frames per second in progressive scan mode, a 1920×1080 image contains more than twice as many pixels, so, to fit into the same bandwidth at comparable image quality, 1080-line video must be interlaced into 60 fields (half-frames) per second or captured at half the frame rate (30 fps) in progressive scan mode. For this reason, 1080 is not necessarily always better than 720; it is on fixed picture, but moving pictures are complicated by the interplay of interlacing and allowable frame rates. In addition to 30 and 60 fps modes, the high-definition standards also allow for 24 fps (and 23.976 fps) video natively, without the need for the 3:2 sequencing used by early NTSC-based 24P systems. So, although many cameras still record 24P images in a 59.94 fps interlaced container

Table 2-1 High-definition and digital standard-definition broadcast formats defined by the Advanced Television Systems Committee (ATSC), which correspond to formats in use for current and future digital video recording systems in North America.

Resolution (Pixels)	Aspect Ratio	Frame Rate	Scanning	Nickname
1920 × 1080	16:9	23.976 fps, 24 fps	progressive	1080p24
		29.97 fps, 30 fps	progressive	1080p30
		29.97 fps (59.94 fields/s), 30 fps (60 fields/s)	interlaced	1080i60
1280 × 720	16:9	23.976 fps, 24 fps	progressive	720p24
		29.97 fps, 30 fps	progressive	720p30
		59.94 fps, 60 fps	progressive	720p60
704 × 480 ¹¹	4:3 (standard), 16:9 (anamorphic)	23.976 fps, 24 fps	progressive	480p24
		29.97 fps, 30 fps	progressive	480p30
		59.94 fps, 60 fps	progressive	480p60
		29.97 fps (59.94 fields/s), 30 fps (60 fields/s)	interlaced	480i30
640 × 480	4:3	23.976 fps, 24 fps	progressive	480p24
		29.97 fps, 30 fps	progressive	480p30
		59.94 fps, 60 fps	progressive	480p60
		29.97 fps (59.94 fields/s), 30 fps (60 fields/s)	interlaced	480i30

¹¹ The ATSC standard specifies 704 × 480 image size using non-square pixels, which corresponds to 720 × 480 video used by MiniDV, DVD, and other standard-definition digital video recording formats (see Fig. 2-3). Although 720 horizontal pixels are recorded on the medium per line of digital video in a 720 × 480 format, the leftmost and rightmost 8 pixels are not part of the 16:9 or 4:3 image and are discarded on playback. Using 720 instead of 704 allows for a 3:2 nominal ratio of horizontal to vertical line size in DV recording systems.

format, future development will be in the direction of native 24 fps recording.¹²

The standardization of multiple high-definition video formats has bearings on sound because of the different frame rates permissible and typically used with each mode, including both NTSC-based and integer frame rates. Although many combinations of resolution and frame rate are possible, the most common formats for high-definition digital video capture are 1080i60, 1080p30, 1080p24, and 720p60,¹³ usually at NTSC-based (non-integer) frame rates (23.976, 29.97, and 59.94 fps). Generally, if audio is not recorded on camera single-system, an audio recorder would be set to 29.97 fps time code in any of these situations, but to 30 fps time code for material recorded at film-based (integer) frame rates. Time code settings, including the difference between drop-frame and non-drop-frame time code, are discussed below.

Under- and Over-Cranked Camera

Motion-picture cameras may be equipped with variable-speed motors, allowing them to shoot at a precise 24 fps most of the time, but also allowing the operator to run them off speed over a wide speed range for special effects. Until recently, these effects were hard to come by with video, because all the timing signals and so forth in a camera are based on the frame rate—many things have to change together to vary the frame rate. Today, some high-end video cameras offer variable speed recording, but usually only in specific increments. Thus it may be possible with some setups to capture slo-mo or sped-up action by over- or under-cranking the camera, respectively. Audio for over- or under-cranked situations is usually invented in postproduction, although one of the sources used may be a separate recording of the same event recorded in real time and then

¹² One of the earliest cameras to offer “native” 23.976 fps recording was the Panasonic HVX200. Recording 24P video natively rather than embedded in a 30 or 60 fps video stream saved space on the recording medium since redundant frames were no longer stored.

¹³ The chief advantage of 720p over 1080-line formats is that each frame of 720-line video contains less data, so more frames can be stored per second, and the higher frame-rate of 60 full frames per second allows for smoother motion. That’s why it is particularly useful for sports or action programs with a great deal of motion on the screen. That’s also why use of a lower frame rate such as 720p24 is uncommon, since at that frame rate the bandwidth limitations on 1080-line video do not come into play.

sped up or slowed down by the same amount as the camera to maintain synchronization.

Operational Matters

Quality Modes and Codecs

Recording raw pixel values uncompressed would take an enormous amount of space on recording media, so digital video today is stored and subsequently transmitted using a data compression scheme known as a *codec* (short for *compressor–decompressor*). The DV family of codecs (including DVCAM and DVCPRO) provided a means of recording standard-definition video at bit rates between 25 Mbit/s for consumer variants and 100 Mbit/s for the highest bit rate professional formats, providing about a 2:1 to 5:1 reduction in bit rate compared to uncompressed standard-definition video. The recording times and capabilities of DV tape formats varied depending not only on the features and bit rates of the codecs in use, but also on the particular cassette formats in use and the settings selected on the camera. Today, high-definition cameras continue to offer users a trade-off between video quality and recording time, although newer video codecs are in use on high-definition cameras.

Uncompressed high-definition video takes more space than standard-definition material, for obvious reasons, so the compression ratios involved in storing HD video are generally greater than the 5:1 ratio achieved by digital video in its infancy. The most common codecs in use today are MPEG-2 and MPEG-4 variants, including AVC/H.264, and typical compression ratios are on the order of 20:1. The digital video recording landscape mirrors the distribution world of Blu-ray and DVD in that respect, since standard-definition DVDs encode video using the MPEG-2 codec, and both MPEG-2 and AVC/H.264 are standard Blu-ray encoding formats. An important difference between the worlds of capture and distribution lies in how audio is encoded on the medium. The AC-3 audio codec, a perceptual coding scheme also known as Dolby Digital, and similar offerings from DTS are commonly used to compress audio for authoring to DVD, and to a lesser extent Blu-ray, where lossless “high-definition audio” formats like Dolby TrueHD and DTS-HD are now commonly offered as well (see Chapter 11 for information on encoding final audio for

distribution). Perceptual coding is well suited to compressing media for distribution, where it will not have to be re-encoded or edited down the line. In capturing video, on the other hand, recording full-quality, uncompressed PCM audio is important in ensuring that audio can be edited and processed in postproduction while maintaining the full fidelity of the original recordings.

Similar to SP and LP recordings on older tape formats, many cameras offer a choice between standard play (SP) and long play (LP), standard quality (SQ) and high-quality (HQ), or other similarly named recording modes. By using a lower video bit rate, low-quality or long-playing modes fit more video onto a recording medium, but at the expense of image quality and often sound quality as well. Accepting the trade-off in picture quality for increased recording time may be useful in recording home videos or simple videos for the web, but with the space availability on high-capacity discs and memory cards today, low bit rate modes are normally not needed in professional or industrial video production.

The caveat for sound is that, on consumer camcorders in particular, compressed video modes may be tied to a lossy audio format, the rationale being that the SQ or LP is designed to save space on the memory card and that more space can be saved by heavily compressing audio as well as video. For home video recordings designed to be played back directly with no editing or postprocessing, this may be fine, depending on the particular codec implementation and bit rate used. But for professional purposes or situations in which audio may be processed and combined with other elements during postproduction, lossy audio recordings can introduce audible artifacts into a program when the audio is decoded (and potentially re-encoded) in the editing room. Indeed, more than a few student and even independent films have been compromised by low-quality, “ringy” audio caused by MPEG encoding on an incorrectly set-up recorder. The important thing when selecting a camera is to make sure that a video mode exists that will record uncompressed PCM audio, and to use it during recording.

Interchangeability

In the early days of video, tape formats were designed to allow video to be recorded on any camera using a particular cassette format and

played back on any deck handling that format. In digital video, material could still be interchanged between two Mini DV or two DVCAM cameras, although complications could arise when recording on a camera made by one manufacturer and playing back off another manufacturer's device. On the other hand, certain DV cassette formats were used by both consumer and professional systems, leading to the question of whether a tape recorded in one digital format would play back on another format of camera. Generally, the answer was that higher-end format decks or cameras could play lower-end tapes, but not the other way around. That was when everyone was using the same codec family (DV/DVCAM) and recording medium (8 mm magnetic tape).

Today, there are so many variables involved in recording an image on tape or, more commonly, memory card or hard disk file systems that it's unlikely a card recorded on one camera will be readable even on another camera that uses the same media type. Within a single manufacturer, camera, or model line, it may be possible to share cards on a production, but generally each card will be read out to one or more hard drives for backup and ingestion into editing software once it is full, or at a natural break in shooting, and reformatted to start from scratch in whatever camera it ends up in.¹⁴ Memory cards, optical discs, and hard disks can all be read by computer¹⁵ and backed up there; the camera remains an instrument for capturing video, but is no longer needed as a vehicle for playback and ingestion the way it was with tape-based workflows.

So interchangeability has decreased with growth in the number of camera formats on the market, but on the other hand the necessity of playback from a camera or deck during postproduction has been removed as tape originals no longer represent the archival format they once did. Instead of ensuring that the tapes in use on a production are all interchangeable and can be played back on a particular deck or camera during postproduction, the burden of interchangeability in a tapeless workflow becomes making sure that the files recorded all play nicely in the editing room. And for that, it remains important not to mix wildly different formats and expect them to work together seamlessly. The best advice for both picture and sound recording is to select a method and stick with it. For sound in particular, be

¹⁴ The nuances of working in a tapeless production environment and managing media will be discussed in more detail in Chapter 7.

¹⁵ With the appropriate software if a proprietary video file format or file system is in use.

sure to select a recording platform that records uncompressed PCM audio that can be used transparently for editing and processing in postproduction.

Off-Line/Online

In one conventional method of working, video is first copied from source tapes at low resolution into an off-line editing system. The lowered resolution allows seeing what is happening, but also permits keeping a great deal of content available simultaneously because of the trade-off between picture quality and running time.¹⁶ The picture is edited off-line (without accessing the full quality originals) and the result is an Edit Decision List (EDL), a file containing the starting and ending time codes of each clip included in the program. This EDL, in electronic form, is then taken to an online facility along with the original camera or other source tapes. The online system rapidly re-creates the cuts by selective recording from the source tapes to an Edit Master at full resolution. Off-line time is much cheaper per hour than online suites, which offer not only higher quality but also special effects that have traditionally not been available in basic editing systems.

In contrast, the digitization of video at the camera level and tremendous growth in hard disk storage capacity have today enabled the transfer of full resolution video into non-linear editing systems, giving the editor a clone of the camera original throughout the editing process. Full-resolution editing is possible even when a great deal of content is kept online. The enabling technology is the digital video compression employed between the CCD and the recording medium to reduce the bit rate of even high-definition video to practical levels for recording onto memory cards and storing on hard disk, along with the remarkable gains in storage capacity and falling cost of storage over the last several decades. Thirty years ago, I (Tom Holman) bought a hard drive that cost \$2.43 per megabit, whereas today you can buy a drive for \$0.00001

¹⁶ "S..t Tom, you gotta tell George about this," said Stanley Kubrick to author Holman in a phone call. Kubrick went on to say he'd memorized the film dailies, so he didn't need high quality picture in the editing system, and that the flexibility offered by a non-linear editor opened up a whole new experience cutting for him.

per megabit, an improvement in cost-performance of a factor of 240,000:1! This fact demonstrates why editing systems today can afford to handle full-resolution content even for large projects, whereas formerly they could not.

Time Code

Almost all prosumer digital video camcorders, many consumer models, and some DSLRs, have the capability of recording time code in the SMPTE format. The time code 05:15:22:23 means 5 hours, 15 minutes, 22 seconds, and 23 frames, marking a single frame with a number that may be used as an index address to that specific frame. The numbers start at the beginning of a tape or disc at 00:00:00:00 on many consumer models, but on more professional devices the starting number is presettable.

Time code is simple in the hours, minutes, and seconds digits, incrementing upward serially just as a clock would, rolling over from 59 seconds to zero and incrementing the minutes counter, for instance. However, when it comes to the frames counter, things are a little more complicated because of the various frame rates in use in digital video. For NTSC television and NTSC-based formats, the counter increments from zero through 29, then back to zero, incrementing the seconds counter, while for PAL-based formats the frames counter increments from zero through 24, then back to zero at the beginning of each new second. This process thus yields a nominal 30 frames per second (fps) for video in NTSC countries and 25 fps in PAL countries. For true 24P material, the frames counter increments from zero through 23 and then starts from zero once more, incrementing the seconds counter. However, for 24P material embedded in a 50i or 60i video stream as described above, 25 fps or nominal 30 fps time code would be used. One reason for this approach is that material recorded this way is backwards compatible with existing video infrastructure and can be easily played back on any television.

Now let us turn our attention to the word *nominal* in the foregoing discussion, when 30 fps was used as a term. Although 29.97 fps video in NTSC countries plays fewer than 30 frames on average every second, we don't count time code in increments smaller than a frame, so the 30 fps time code sequence is used even for 29.97 fps video.

However, because frames are now very slightly longer in playing time than those played at precisely 30 fps, an error accumulates over time. The time code is running slow compared to a wall clock, so playing out frames at this rate will result in the program running slightly long. One hour of time code will actually run one hour plus 108 frames of clock time. For many purposes, this error is unimportant. For such uses, the counting sequence that includes all numbers in the normal, incrementing sequence is used, and programs run very slightly long in comparing their time code to a wall clock. The name for this, rather unfortunately, is non-drop-frame (NDF) time code. We'll come back to why we say the name is unfortunate in a moment.

Alternatively, if you are delivering to a broadcast operation that runs on clock time, 108 frames per hour of error is very important. For time-sensitive operations, a means of keeping the time code in sync with a wall clock is needed, but it can't be done if all of the frame numbers are used—quite the conundrum. The solution to the problem is to skip 108 frame *numbers* per hour, jumping over them completely. No frames are in fact skipped, but rather the number sequence skips certain numbers; if video frames were in fact skipped, you would often perceive the jump as a mistake. This is why we consider the name unfortunate, because this code sequence is called drop frame (DF) time code when, in fact, no frames are dropped!

In drop frame (DF) time code, frames 00 and 01 are skipped once every minute (which would yield a difference of 120 frames per hour), except in the minutes 00, 10, 20, 30, 40, and 50 (for a total of 108 frames per hour). So the actual displayed time is not precise at most points in time during an hour but adds up to one displayed hour of time code at the end of one actual hour. You can tell if video has been recorded DF or NDF by seeing whether these particular frame numbers are skipped. A common method of indicating DF code is to change one or more separators of the time code parts from a colon (:) to a semicolon (;).

The choice of NDF or DF code to be made at the time of recording depends on the editing system and the requirements of the delivery master. Most conventional broadcast uses require DF time-coded master tapes, so real-time television operations can be timed by looking at the time code. Thus this requirement backs up into editorial and even, sometimes, into the camera tape specification. At one time it was difficult for off-line

editing systems to make the internal calculations necessary for counting frames with DF code, when some numbers were jumped over, and thus the most common way of working was to record NDF code on the camera, work with it editorially, and then make the final delivery master with DF code. In some cases, the choice of video format may affect the type of time code available for use on a particular camera. For example, even on cameras that allow switching between DF and NDF time code, selecting a 24P mode may limit the camera to recording NDF time code.

Today more editing systems can handle DF code internally, so it is possible to shoot it on many cameras and keep it throughout postproduction. As a sound person, this is something that you should bring up to producers: “What type of time code should we use, non-drop-frame or drop frame?”¹⁷ The answer to this question is too infrequently known during shooting because most of the consequences occur in postproduction. One of the biggest problems occurs if the answer is not known during production and different parts of a production randomly choose DF or NDF code. In such a case, editing is at least greatly complicated, and sometimes made impossible on certain editing systems, by the mixed code.

The simplest code generators start at 00:00:00:00 each time the camera is powered on or a new disc or tape is inserted into the camcorder. In fact, some camcorders will lose track of the last time code recorded if taken out of standby mode or if a tape is rewound for playback and then cued up for subsequent recording. In that case, the same time code sequence may be recorded a second time on the same tape—a potential difficulty for the editing system to sort out if trying to go back and find a particular clip off a source tape. Other cameras are able to read the code of an existing tape or from files on the medium up to the point where no recording has been made and then pick up recording the time code from where they left off.

An important difference lies in the two possible modes of generating and recording time code on the camera: FREE RUN and REC RUN

¹⁷ Unless the answer is prescribed by the format. Some consumer camcorders, including most early DV cameras, for example, only support dropframe time code. Most 24P modes only support non-dropframe.

time code. For DV cameras, typically the time code generator runs when the tape rolls and pauses when the tape is paused. This is called Record Run (REC RUN, R RUN) and it produces continuous code throughout a tape even with camera starts and stops, so long as the power isn't interrupted or the tape removed and put back in. Free Run (FREE RUN, F RUN) means that the time code generator is typically set to clock time and runs continuously, storing discontinuous code with each start of the camera, but providing evidence of the actual time of an event, sometimes needed for court requirements, for instance. Cameras recording to more professional formats than DV may be switched to REC RUN or to FREE RUN, although REC RUN is used far more often. REC RUN is generally preferred because editing systems are then presented with continuous time code, and this is more easily dealt with than discontinuous code.

Setting the time in FREE RUN mode can be accomplished by a function called JAM SYNC, which sets the camera's internal time code generator from an external source. Jam sync is used to synchronize camera time code generators when multiple cameras are used to shoot the same event, or to synchronize one or more camera time code generators to the time code generated by an audio recorder when double-system sound is in use. Higher-end cameras may also provide an input for a Genlock video input that permits sources shot on separate cameras to be matched precisely in time (the first line of video will start at the same time on each camera). Genlocking multiple cameras permits overlaying a graphic while switching between cameras in real time, for instance, whereas just jam sync alone is not enough to accomplish this.

Many prosumer and professional cameras have the capacity for the user to set a starting time code even in REC RUN mode, as do a few consumer models. In tape-based situations with large amounts of footage—including documentary programs—the resetting to 00:00:00:00 each time a new tape or disc is inserted is a nuisance for logging media, because the time code then can't be used to distinguish one out of a multiplicity of tapes for a given project. Running time code continuously from one disc or tape to the next can reduce or eliminate the appearance of repeated source time codes over the length of a project. In one conventional method of working, the minutes and seconds of time code are reset each time a new tape or disc is inserted, while the hour counter is incremented by hand to match the tape number (or, more generally, the digital reel number) in use.

In an off-line/online workflow (described above), the EDL would then contain information about which source tape a clip comes from embedded in the time code, in addition to any notation elsewhere in the metadata for a clip. Without that extra information, very good records must be kept of which tape is meant for each picture and/or sound edit, because the online editing system will ask for a specific tape by an identifier such as a six-digit code to be mounted into a playback machine and then will shuttle to the time code indicated on whatever tape is provided. If it is not known what tape a particular time code event comes from, then the wrong tape may be used and the wrong picture or sound inserted into the program. Moreover, if a given tape or disc contains passages with identical time codes at multiple locations on the medium, it is especially ambiguous which clip is meant.

Luckily, the need for fulfilling the bookkeeping requirements of the off-line/online method of editing is not as great as it once was because newer cameras and editing systems allow full-resolution digital video editing in most situations, precluding the need to go back to camera original tapes or media during an online session since source video files on the editing system are clones of the original recordings. In most systems in use for simple digital video editing today, the distinction between off-line and online quality does not exist.¹⁸ Thus there is no need to repeat the import of audio and video for an online session, and keeping track of sources is not quite so vital. In fact, source media are frequently re-used in solid-state recording systems once footage has been transferred to a primary hard drive and one or more backup drives, so the original sources may no longer exist once editing has commenced in earnest. Still, for backup purposes, it is always a good idea to know where the source video came from, in the case of an editing disaster or a question about syncing sound and picture or matching multi-cam picture, so labeling and logging tapes with a distinguishing code such as at least A, B, C,... Z, AA, AB, AC... is good practice. Even in the case of reusable media being recorded over several times on a production, the file structure of each individual card is

¹⁸ Although there may still be a finishing step in an online suite that can offer more extensive video processing than is available in off-line systems. In this case, an export of a file containing the video is made from the off-line system to the online one. In the conventional case, all that is exported from the off-line system is an Edit Decision List (EDL), and the edits must be reproduced by the online system. Editing at full resolution in a non-linear editing system relieves the requirement of recopying from source tapes to the master during an online session and possibly eliminates the need for a session altogether.

typically preserved when the card is initially transferred, so notes on which card a clip originated on are still useful in case of emergency.

In recording time code on an audio recorder separate from the camera, one nuance is that 30 fps or 29.97 fps time code is used on NTSC material even for cameras operating in 24P modes. Usually 29.97 fps time code is used for material shot at 29.97, 59.94, or 23.976 fps. When recording audio at 48,048 Hz for a project shot on film (or at precisely 24 fps) that will be pulled down to video speed, 30.00 fps NDF time code is used, as it will also be slowed down to 29.97 fps when the audio is clocked out at 48 kHz to match the video derived from the film. Once again, note that even where the camera is operating at 24.000 fps, the audio recorder is set to 30.000 fps time code.

User Bits

Within the SMPTE specification for time code is a place for the user to note additional information, called user bits or UB (sometimes wrongly and hilariously called the user's bit by manufacturers). This can be information regarding scene and take, time of day, particular filters in use, method of recording the frames (whether interlaced or progressive), and so forth. Because these are literally the user bits, they are not subject to strict standardization, and so mean different things on different cameras. Consumer camcorders generally do not allow the user to write user bits, whereas prosumer and professional digital video cameras may. Capabilities in this area vary by format, manufacturer, and even specific camera model.

PAL Formats

Generally PAL formats, those used in Europe and other 50-Hz power line countries, use 25 fps code and do not have the complications associated with 29.97 fps time code described previously. Frames are numbered with time code 00 to 24, and then increment the seconds counter and reset to 00. The time code runs according to a wall clock, so there is no notion of dropframe versus non-drop-frame time code in PAL-based systems. One difficulty, noted above, is that in order to transfer to film for theatrical

exhibition, the 25 fps video must be slowed down by some 4 percent to achieve film standard 24 fps projection. Sound accompanying PAL video will also be slowed down accordingly in order to stay in sync, and without correction this will result in all the pitches being 4 percent lower—the opposite effect of the 4 percent upward pitch shift on film or 24P originals transferred to 25 fps video. Once again, this is noticeable even on actor's voices, so it should be corrected in transfer through the use of a hardware or software *pitch shifter*.

Locked versus Unlocked Audio Sample Rate

Camcorders contain internal oscillators, much like those used in quartz crystal watches, to set various timings. They may use one master clock that is then divided as necessary down to various rates for audio and video, or they may employ two separate clocks for the audio and the video. The separation into two clocks makes the design somewhat simpler. On the other hand, this requires the audio and video clocks, those doing the audio sampling and making the video frames, to run in a synchronized relationship to each other, called a locked audio sample rate, to avoid the possibility of sound and picture drifting out of sync over the course of a long recording. For NTSC standard U.S. video at 29.97 fps (including 59.94 Hz interlaced), every five video frames contain exactly 8008 audio samples.¹⁹

On tape-based formats, where large sync drifts could occur over the length of a continuously recorded tape if audio and video were separated and then brought back together in an editing system, this was a particularly important consideration. The consumer DV format (also known as Mini DV) used an unlocked audio sample rate, in which no special hardware or firmware was included in the camera to ensure that the audio and video clocks remained in sync over the length of a tape. Note that tapes played back would not exhibit drift since audio and video were interleaved on the tape, but since the total number of audio samples recorded per video frame was not necessarily the same for every frame, programs would drift in some editing systems that strictly reassigned 48,000 samples of audio to

¹⁹ Note that no smaller number of frames contains an integer number of audio samples.

each second of video.²⁰ Other formats such as DVCAM and DVCPRO used locked audio so that no matter what happened down the line, even with audio and video transferred separately, so long as certain conditions were met,²¹ *lip sync* (audio–video synchronization) would be preserved.

Not having a strict requirement on audio sample rate makes cameras somewhat simpler and cheaper, so cameras in the early DV format and certain other models employ unlocked audio in the interest of cost savings. While prosumer cameras generally should employ synchronized (locked) audio and video clocks, it is still a good idea to check that audio and video on a new camera maintain sync through to the editing system over the course of a long recording during preproduction, as described in Chapter 3. In an hour-long interview recorded on a single disc or memory card, or an hour of tape playback, a typical amount of drift that unlocked audio may cause in an editing system (with sound and picture transferred digitally) is 20 frames out of sync. Because one frame out of sync is visible, and two frames is quite noticeable, 20 frames, or about two-thirds of a second, is quite bad. Most takes are not 60 minutes long, and a six-minute take would be two frames out of sync and would be on the edge of becoming unacceptable. However, transferring full-length tapes into an editing system all at once can result in later takes being badly out of sync.

Instant Playback

One tremendous benefit of file-based (as opposed to tape-based) recording on modern digital video cameras is the ability to instantly play back a take through the on-camera screen or an externally connected monitor without rewinding the tape. In particular, this reduces the danger of recording over material after rewinding a tape and eliminates the need to cue the tape back up to the end of a recording after playback. Non-linear access to material has long been a hallmark of digital editing systems, and now the availability of random access media in camera original recording has

²⁰ What is done to maintain sync while playing back the tape in real time is that the audio sample rate is sped up or slowed down by a small amount during playback of the tape to keep the audio in step with the video.

²¹ Namely, that the clocks doing the playback of the audio and video have the same time base, like setting watches to the same time before a mission (“synchronize your watches, gentlemen”) in all those World War II movies.

introduced the non-linear paradigm to on-set production as well. While perhaps not as far-reaching a transformation as the genesis of digital photo cameras capable of displaying a photograph on screen instantaneously, this capability of modern digital video cameras nonetheless adds ease of use to the process of reviewing video on set.

Interconnecting Video

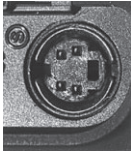
In addition to viewing video on the LCD screen or through the viewfinder of a camera, video can be output for live monitoring and playback to an external monitor on set or in a studio or office to view dailies after production.²² Video consists of three color signals, but these may be portrayed on from one to five wires, with various degrees of quality and complexity as outlined in Table 2-2.

Conclusion

The principal differences among the various digital video formats are frame rates, codecs in use (which can include undesirable audio codecs on some cameras or in some camera modes), resolution, and aspect ratio. Particular camera characteristics that can affect sound also include having a locked versus an unlocked sample rate, the ability to handle simple or more complex time code requirements, and the capability for over- or under-cranking, which introduces special sound considerations that should be considered in preproduction. Except for in low-quality or long-playing modes, though, the basic audio quality remains fairly consistent in cameras that offer professional audio inputs. Normally, for professional production, audio is recorded on camera as two-channel, 48-kHz sampling rate, 16-bit audio (or in some cases 24-bit), or with an external recorder capturing two or more channels of 48-kHz, 16- or 24-bit audio.

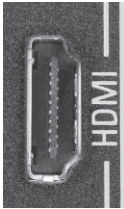
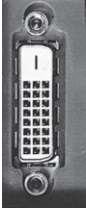
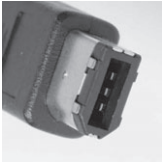
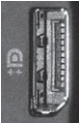
²² Video can also be transferred into an editing system or dubbed to disk or tape with standard video cables or Firewire cables, although that was more common in the era of tape-based video formats, including both analog and DV formats. Today, direct file transfer from solid state media to hard disks is more common, as described in Chapter 7.

Table 2-2 Video interconnects.

Name	Connectors	Description	Up/downside
Composite Video	Yellow phono connector Single BNC connector on pro gear	Complete color video signal on one wire	Simple, but combining color components together causes artifacts; useful for field monitoring, but never used if possible for production purposes 
S-Video	Four-pin mini DIN connectors	Separate luminance (the black-and-white part of the picture) and chrominance (the color part) on two conductors	Better than composite and still simple to connect, but limited bandwidth and thus resolution of color channels compared to component connections. Good for monitoring, but not production standard where subsequent matting is necessary, for instance 
Analog Component Video	Three to five phono or BNC connectors: three colors plus potentially vertical and horizontal sync signals	Three separate color signals plus possible sync signals; progressive or interlaced	Best for bandwidth of each component color; useful for postproduction. Progressive scan useful for eliminating interlace artifacts 

Continues

Table 2-2 Continued

Name	Connectors	Description	Up/downside
HDMI	Multi-pin connector used on HDTV, Blu-ray, and other consumer electronics	Digital video (progressive or interlaced) transmitted alongside digital audio on a single interface	High-bandwidth digital video transmission. No need for a separate audio connector
			
DVI-Digital Video Interface	Customized multi-pin connector used primarily for PC video	Separate digital color component video, potentially progressively scanned	Like HDMI, but incapable of transmitting audio and video on a single connector
			
FireWire	Digital signal on custom connector	Used for data transfer to hard drive or for playback on older tape-based DV formats	Fast for file-based video transfer. Allows for real-time playback of SD video, but not HD video
			
DisplayPort	Custom digital video connector	Used primarily in the computer industry for connecting PC displays	Also available as a modified Mini DisplayPort connector on many laptop computers
			
Thunderbolt	Peripheral connector identical to Mini DisplayPort version of the above interface	Digital interface carrying a combination of video and data	Designed to be backwards compatible with DisplayPort video sources. Can be adapted to DVI, HDMI, and other interfaces for video transfer purposes
			

Director's Cut

- Be careful to choose and stick to a single choice of standard format, because intermixing types is next to impossible. NTSC is U.S. standard video; PAL is European standard video. NTSC records at 29.97 fps, and thus it is difficult, although possible, to get a film output at 24 fps. PAL records at 25 fps, and may simply be slowed down 4 percent to make a 24 fps film output. Even though PAL and NTSC are technically standard-definition video systems, the frame rates derived from PAL and NTSC television carry over to American and European high-definition video systems, including 24P video modes.
- Aspect ratio: while standard-definition video is usually 4:3, high-definition is 16:9. (Some specific anamorphic video modes such as widescreen video on DVD are intended for 16:9 playback even though recorded in a standard-definition 4:3 frame.)
- Understand interlaced versus progressive scanning. Most traditional video is interlaced; most computer work is done progressive. High-definition video may be interlaced or progressive depending on frame size (1080 or 720 active lines), frame rate, and bandwidth requirements. Many cameras store 24P video as interlaced video, although the picture itself is progressively scanned and can be put back together into a single progressive frame in editing, which leads to better film outputs if necessary. 24P in effect mimics the universal nature of 24 fps film that may be released as film or transferred to PAL video at 25 fps by a 4.166 percent speed-up, and NTSC video by a slowdown to 23.976 fps and the introduction of a 2:3 repeating field sequence, resulting in a video frame rate of 29.97 fps.
- Use high-quality video modes to ensure both picture *and* sound are recorded high-quality. Don't use long-playing (LP) or "standard"-quality video modes except in absolute emergencies.
- Understand time code terms such as Non-Drop-Frame (NDF), Drop Frame (DF), Record Run, Free Run, and Jam Sync, described in the text. Choose one method of working and stick to it throughout a production. For multicamera shoots, use Genlock if possible.
- Know that audio may go out of sync in some editing systems for DV tapes, or even on discs or longer files on memory cards, because of the use of unlocked audio and video clocks. Test for sync drift before recording a program.

Director's Cut *(Continued)*

- Interconnecting video ranges from one-wire composite video, with compromises, through S-Video, to component video, each step with less of a compromise. In addition to playing video via a connected monitor on set, it is also possible to review digital video with on-camera LCD screens and take advantage of the non-linear nature of modern digital video recording.

This page intentionally left blank

CHAPTER 3

Preproduction Planning

For many programs, the quality of the sound picked up on location is the principal determining factor in the overall sound quality. There are a number of decisions that go into producing a program, many of which must be made before production starts in earnest. The most obvious are things like which camera to use, whether to record sound single-system if the camera capabilities allow it, when to record sound double-system and how to set up a dedicated audio recorder, and how to assign crew positions or responsibilities on set so that each job that needs to be done is taken care of. However, other decisions that affect the sound quality of a program are often made with no thought given to their implications on sound quality down the line, such as decisions about where to shoot made during location scouting. Moreover, testing and setting up equipment properly, including interconnection of microphones, mixers, and recorders, is vital to the technical quality of recorded content and should be worked out in advance to avoid the heartache of losing a priceless performance on set. Of all the information in this book, this chapter and the chapters that follow on Production Sound are probably the most important, because once sound has been picked up badly, it is difficult and expensive if not impossible to rescue.

The First Big Step: A Dedicated Sound Person

In many instances, DV videographers have to be one-man bands, doing their own producing, directing, camera operation, and sound recording. However, when just a second person is added to the crew, that person almost naturally becomes the sound person. It is common in documentary production, for instance, for a two-person crew to split the responsibilities of camera and sound recording, the director taking on the responsibility of sound person on set. The videographer lights the scene, or chooses to

manipulate available light, sets up the tripod and camera, white balances it, and so forth. Meanwhile, the director is getting to know the documentary subject and setting up the sound. If necessary, he or she will operate a boom microphone during the shot while the videographer controls the camera. Fiction filmmaking with minimum crew may follow this breakdown of jobs as well.

It is fair to say that the needs of picture often dominate on location, just because a visual image of the location is being captured and the location was most likely selected for a look that would be hard to reproduce elsewhere. But although production teams often think that sound will be easy to reproduce anywhere or they can “fix it in post,” it is often true on low-budget productions and especially in documentary filmmaking that the sound recorded on set cannot be effectively replaced in postproduction. While pictures tend to dominate thinking over sound, sound still accounts for 50 percent of the experience. With a second person responsible for sound on set, it is less likely to be overlooked or compromised.

That said, it is still important, even in a two-person crew, for the camera-person to be able to aurally monitor the ongoing events. If necessary, this means equipping the camera operator with headphones,¹ because decisions about framing the picture, determining when to zoom in, and so on, particularly in documentary films, depend on what is being said by the characters in a scene. As the scene grows more intense, it is usually desirable to push in to a close-up, for instance, but the camera operator can't know that the intensity is rising unless he or she can hear and respond to the action. Of course, the sound person should also pay attention to camera operation to work effectively in run-and-gun shooting situations. With practice, observing the focal length of the zoom lens by looking at a zoom stick² will tell the sound person with practice how wide or tight the frame is and where the boom mic may go without entering the frame. Alternatively, the boom pole can have a small video screen attached and wired to the camera output, so that the boom operator can see the shot and

¹ Using a Y-adaptor for the connectors in use to extend monitoring to both the camera and sound persons.

² A stubby stick that screws into the zoom ring on the lens, indicating by its angle the current focal length. If your lens lacks such a provision, then white marker tape can be used, set in a triangle to point at the current focal length.

know how close the boom can be held (and it is usually best to get it as close as possible to being in the frame).

In working together to keep the boom out of the frame, there is one additional gotcha in video compared to film: the viewfinder of a video camera provides no area above the recorded picture to see the potential for intrusion of the boom before it actually enters the frame. In a film camera, the eyepiece optics display more area than the picture aperture, and the eyepiece ground glass is marked with the frame line for the format in use. Thus the cameraperson has an early warning of when the boom may intrude into the picture. There is no such capability with video, because the cameraperson is looking at a video monitor through the eyepiece, and every scanned line of video is actually being recorded as part of the picture. Thus video camera people get good at looking through both eyes: one at the viewfinder for the composition, framing, focus, and even in some cases audio level, and through the other eye at the overall scene, to know where items are about to intrude into the picture—especially the boom mic.

In assigning crew responsibilities, then, it is usually most beneficial to have a dedicated person responsible for sound on set in addition to the camera operator, and for the two to work together to capture the best possible material. Nonetheless, even if the crew can only be one person, certain measures should be taken to get good sound, which will be discussed below.

Location Scouting

Feature films always scout locations before shooting, partly because of the complexity of handling their large crews and equipment. However, location selection is too often based just on what a location looks like: they are often chosen by selecting from stills made by a location manager or even from a database with no one from the production visiting the site. Although a camera can pan off an offending modern structure in an otherwise quaint neighborhood, sound does not have the hard frame line that picture enjoys. Microphone selection, placement, and aiming can only produce tendencies, not a hard boundary for inclusion or exclusion like a frame line. Thus, if a jet flies overhead while you are shooting in a quiet desert, the picture may be fine, but the sound will likely be ruined.

The two main items to watch out for in scouting locations for sound are noise and reverberation. Once we had a student film shoot in a perfect location for picture—the shot contained a lot of graffiti, an important story element for the film—but the location was essentially impossible for sound—it was underneath a freeway. While the dialogue was intelligible, it was far from ideal and had to compete the entire time with loud traffic passing overhead and with traffic noise reverberating around the underpass. One technique for explaining, but not mitigating, the noise encroaching on such a recording is to introduce the scene with an establishing shot of the freeway so that the audience knows the source of the noise and doesn't spend time wondering where it's coming from. That can be a crutch for keeping unwanted noise from becoming too distracting, and may be necessary in some documentary situations, but it does nothing to improve the actual quality of the soundtrack. It's much better to shoot in a quiet location and get a clean recording in the first place.

It could be asked: what is the single most important element in capturing good-quality production sound? Without proper microphone technique or working equipment, a sound recording can easily be ruined, but even paying thousands of dollars for a sound crew or special equipment can't rescue a production from shooting in too noisy a space. So in that sense, finding a quiet location free from unnecessary noise and excessive reverberation—and controlling noise sources on set to ensure clean tracks can be recorded—is arguably the most important piece in the puzzle. Unfortunately, it is also one of the most overlooked, either because budgetary considerations dictate shooting in a noisy location, because picture concerns dominate the location scouting process, or because producers don't realize the importance of location selection in determining the quality of sound that can be acquired on set.

A visualization experiment that can help in selecting locations for sound is to think not just about where a scene takes place, but about what will be shown on camera. Suppose a scene calls for two characters to have a conversation outside a beach house while watching the ocean. The sound of ocean waves crashing onto the beach will be quite loud relative to the dialogue if the conversation is shot on the beach, but if the characters are sitting in beach chairs facing the ocean, most of the sync dialogue will be shot with the camera facing the beach house—away from the shore! It would help matters tremendously if the “beach house” front selected for production were not next to the beach at all, but in some other location where crashing waves wouldn't intrude on the soundtrack. Shooting the

scene at the beach, even with body mics, would undoubtedly cause editorial difficulties and add unwanted noise into the recording.

Any time the visual image seen on camera during an interview or dialogue scene can easily be recreated somewhere else, a production crew has the opportunity to ask: Is this an appropriate location for recording good sound *and* picture? Shooting in a garden with a busy highway behind the garden hedge is poor planning if all that will be seen in the frame background is the garden hedge. Any hedge in a quieter neighborhood could be selected in that situation, or a hedge put up as a studio backdrop for close-up shots, and the sound recording would be much improved. Similarly, shooting in a freeway underpass against a concrete wall should raise the question of whether any other concrete wall without the freeway overhead would have an equally appropriate visual effect.

For the same reason that it doesn't make sense to record dialogue beneath a freeway, shooting in a working restaurant or bar is unadvisable in most situations. The build-up of background noise from dishes, movement, and conversations can drown out dialogue in a restaurant or busy room, and many public venues play music during operating hours, which presents a rights issue in addition to the editorial issues background music can cause. Even when shooting in a restaurant while it's closed, it can be difficult to eliminate all the noise sources on set, such as refrigerators, dishwashers, ventilation units, and even restaurant staff cleaning up after hours, so careful planning during preproduction is advised.

Even in an otherwise quiet location, it can be beneficial to consider the reverberation time of the space you'll be shooting in. Sound can take a long time to decay, particularly in large interiors with hard surfaces, and listening to how long sound takes to decay while scouting can help prevent problems later. Does the sound of a single loud hand clap linger in the air, or decay quickly? Do voices sound much more reverberant (often incorrectly described as "echoey")³ than normal? If so, it may be difficult to record clean sound in that space. For example, while shooting a large museum interior, one wants to start out showing the scope of the

³ An echo is a distinct repetition of a sound, such as the repetition of a word shouted across a large canyon when it bounces back toward the speaker, whereas reverberation describes the buildup of sound characteristic of a large cathedral or concert hall where sound may take up to several seconds to decay to inaudibility.

place with a wide shot, but the corresponding reverberation of the large, hard space causes sound problems. Noise can be controlled if filming occurs while the location is closed to the public, but reverberation remains a problem. In addition to threatening the intelligibility of spoken dialogue, excessive reverberation on location can be a limiting factor in deciding where edits can be made within a scene and an unwanted aesthetic consideration in dictating how reverberant the entire scene sounds in the final mix.

A key point to consider is that reverberation can always be added to a sound element, but is practically impossible to remove from a sound once it's been recorded. So common sense dictates that if an irreplaceable event is going to occur in a large reverberant space, like a wedding to be covered from the back of the church, then special sound considerations apply. One solution for this is radio mics, but their peculiarities also lead to some problems, which will be discussed in the following two chapters on Production Sound.

Choosing a Sound and Camera Workflow

Single-System Recording

The black-and-white clapperboard slate has become an icon representing cinema in popular culture. On movie sets, it is clapped shut at the beginning of each shot to provide a sync reference for picture and sound because the sound for a project shot on film is recorded separately from picture: the film camera exposes a series of photographic negatives and has no way of recording sound alongside picture on the medium. This setup is known as *double-system recording*. When video tape was introduced as a recording medium, it became possible to store audio and video together as magnetic signals on a single medium, and to record sound and picture simultaneously in a single device: the video camera. This arrangement is called *single-system recording* because a single system is used to record both sound and picture, and this is how sound has been recorded for most video projects over the lifetime of the medium.

Single-system recording is still used for digital video when the camera in use on a project has the requisite features, such as external audio inputs,

manual level control, and quality audio circuitry, as described in Chapter 1. It gives content producers the benefit of not having to spend time syncing sound and picture together in postproduction, because the two are already in sync, having been recorded alongside each other in camera. Moreover, it eliminates some of the variables involved in matching frame rates using time code (described in Chapter 2) or guessing at frame rates for 23.976 versus 24 fps material when picture and sound are passed separately down the chain to postproduction, for instance. If they can settle on a camera capable of recording quality audio through a professional microphone input, and in the absence of a compelling reason to try for a double-system setup, most simple productions opt to make use of the benefits of single-system recording.

Today, audio and video may not be laid down with each other along the length of a tape because most camera manufacturers have transitioned from tape to tapeless media, but by interleaving audio and video in a single file on a memory card or hard drive, or at least by starting and stopping audio and video simultaneously with each other so that the stream lengths match and sync up, single-system recording still provides the easiest path to ingesting material into post. Single-system setups are not limited to a single microphone plugged into the camera, although that could be the cheapest and simplest method for a one-man interview crew or similar low-profile situation. Among the components involved in single-system recording could be:

1. One or more microphones. In addition to a boom microphone, one or more radio microphones may be in use at any time, or two microphones may be used to record two subjects during an interview. If two microphones are used, they may be plugged directly into the camera, or first fed through an over-the-shoulder mixer to give the sound person more accessible level controls.
2. A mixer. With one or two microphones in use, an over-the-shoulder mixer with high-quality microphone preamps allows the sound person to set and adjust levels on the fly as needed without disturbing the camera operator to access on-camera level controls. If three or more microphones are in use for a particular setup, a mixer will typically be needed to mix the signals down to two tracks since most cameras have only two external inputs and/or two audio record channels.



Figure 3-1 A typical over-the-shoulder field mixer for small productions, the Sound Devices 302. It has three microphone inputs and two channels of output, with dedicated level controls for each input as well as headphone level controls and other features.

In addition to providing controls for mixing one or more microphones on the fly, even simple over-the-shoulder mixers such as the one shown in Figure 3-1 provide headphone outputs for the sound person, or (via two outputs) to the mixer⁴ and a boom operator in a two-person sound crew. If only one crew member is responsible for sound, the second headphone output could be used, perhaps with a headphone extension cable, to provide an audio feed to the director.

3. The camera. Either the microphone must be connected to the camera directly, or the output of the mixer sent via audio cables to the camera inputs for sound to be recorded on camera. Since there is no dedicated audio recorder in a single-system workflow, the connection to the camera must be maintained at all times for audio to be recorded with picture.

⁴ The term mixer is ambiguous since it may refer to a device for mixing audio, often known as a *field mixer* if of the variety used on most digital video productions, or to the person responsible for mixing the audio, as it does here. Usually the intended meaning can be inferred from context.

4. Cables connecting microphones to the camera or field mixer, the mixer to the camera, and/or providing a headphone monitoring return from the camera back to the mixer. Headphone returns from the camera's headphone jack or audio output allow the sound mixer to hear the actual audio feed being recorded by the camera and ensure that it is not being distorted by level controls on the camera or lost en route.

The setups in Figure 3-2 represent some of the most common single-system recording configurations.

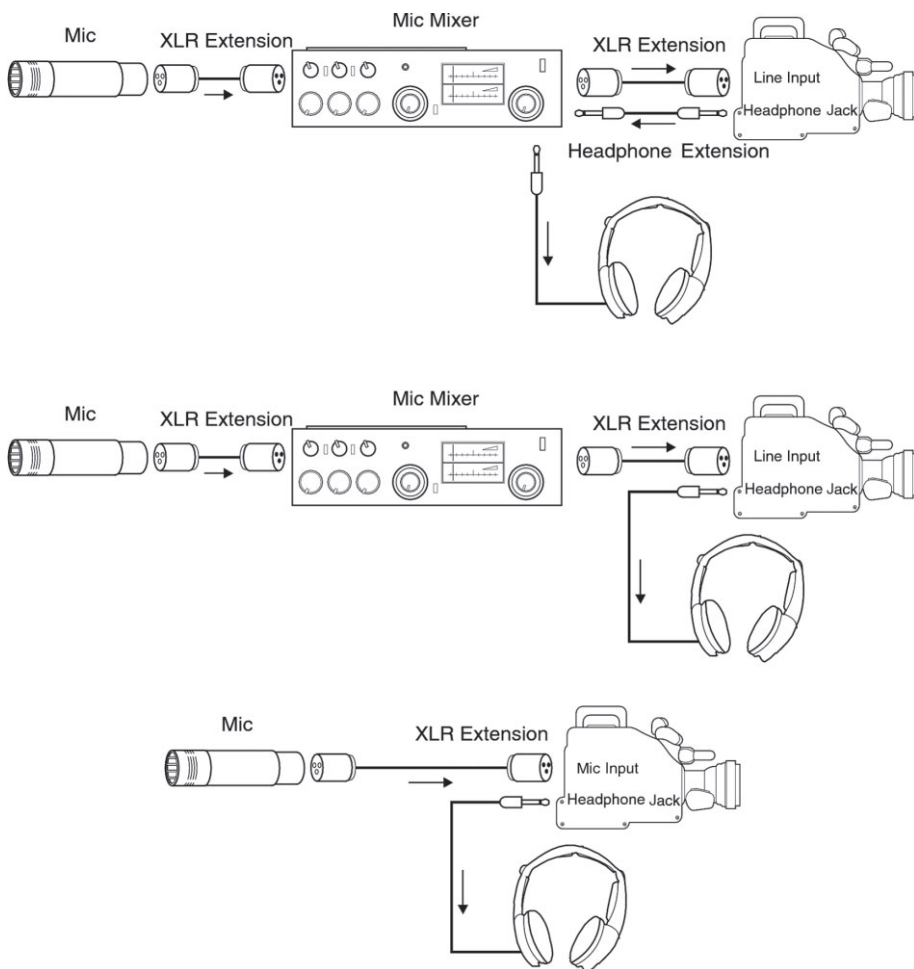


Figure 3-2 Several single-system signal paths, with and without the use of an external mixer inline before audio reaches the camera.

Double-System Recording

While the introduction of video cameras made modern single-system recording feasible, not all digital video camcorders have the capabilities outlined above for recording audio in camera. Some consumer models may have only an on-camera microphone or lack manual level control, while others may not have the quality audio components (microphone preamps, analog-to-digital converters, and so on) needed to produce clean-sounding recordings. Ironically, it is not only the low-end models that have these problems. Both very low-end camcorders and very high-end professional cameras may omit the features or quality components desired in an audio recording system, because on the low end the price range of a camera may preclude investment in “extra” features, whereas on the high end the camera manufacturer will likely expect professional productions to follow the film model of shooting double-system anyway: if a production can afford a high-end professional video camera, the thinking is they will be able to invest in a dedicated audio recorder and probably a larger sound crew as well. So it is on midrange prosumer (high-end consumer or low-end professional) video cameras that one most often finds the nuts and bolts necessary for quality single-system audio recording. On other models, and on DSLRs in particular, *double-system recording* may be a better option, or in some cases the only option.

As on a film set, the use of double-system recording calls for careful planning and execution to make sure audio and video can be synced up in the editing room. It may be possible to sync audio and video by means of a common time code signal or by jam syncing the camera time code to the audio time code in free run mode, as described in Chapter 2. Even with a time code link between audio and video, it is a good idea to utilize a clapperboard slate to provide a sync reference at the beginning of each shot, in case anything goes wrong with the time code sync. And for low-budget or small productions, a clapperboard slate can certainly be used as the primary sync reference for the production.⁵

⁵ One reason time code sync is preferred by editors is that it reduces the time required to sync audio and video by hand when content is ingested into software for postproduction. Finding the frame where the slate closes and the corresponding peak in the audio waveform by hand for each take can be tedious if there is a large number of shots in a program. On the other hand, adding time code sync requires more planning in preproduction, takes more effort on set, and puts additional requirements on production equipment, so may not be possible even with mid-level prosumer equipment.

The procedure for slating shots on set is as follows.⁶ Before each take or camera roll, the sound recordist starts recording sound and may give a verbal identification stating the scene and take or a description of the action about to be shot. Once the sound recorder is rolling, the camera operator starts recording video while a designated crew member holds the slate in frame with the sticks open. If the scene and take were not announced by the sound recordist earlier on, the person holding the slate will give a verbal identification. He or she then says, loudly and clearly, “marker,” and claps the sticks together in one swift motion. Starting the audio recorder before rolling the camera ensures that everything captured on video has a corresponding sound recording, and saves space on video recording media by not recording video while waiting for the audio system to start recording.⁷ Announcing “marker” just before clapping the sticks ensures that the correct peak in the audio waveform can be found when the audio is played separately from picture during synchronization. Otherwise, a similar sound such as a tape measure closing or a sharp clap off-screen may be mistaken for the sound of the sticks closing on camera. If the sticks bounce or sound indistinct when clapped, the sound person may ask for “second sticks,” in which case the person holding the slate announces “second sticks” and claps them cleanly a second time.

The components making up a double-system recording setup generally include (Fig. 3-3):

1. Microphones, including multiple boom or radio microphones mixed together for two-channel recording or intended to be recorded on individual channels by a multi-channel recorder.
2. A mixer, especially if many microphones are in use and a mixdown is desired as a primary audio mix, to fit several microphone feeds onto a smaller number of channels, or as a guidetrack. In simple setups or with advanced mixer-recorders, it is possible to combine the functionality of the mixer and recorder by plugging microphones directly into the audio recorder and setting the record levels with gain and trim controls on the recorder (Fig. 3-4). This limits the sound

⁶ This is a general description of how slating works on a film or video production. For a more detailed description of slating operations on professional motion picture sets, refer to the description in *Sound for Film and Television*.

⁷ Since video takes up more space on a medium than audio.

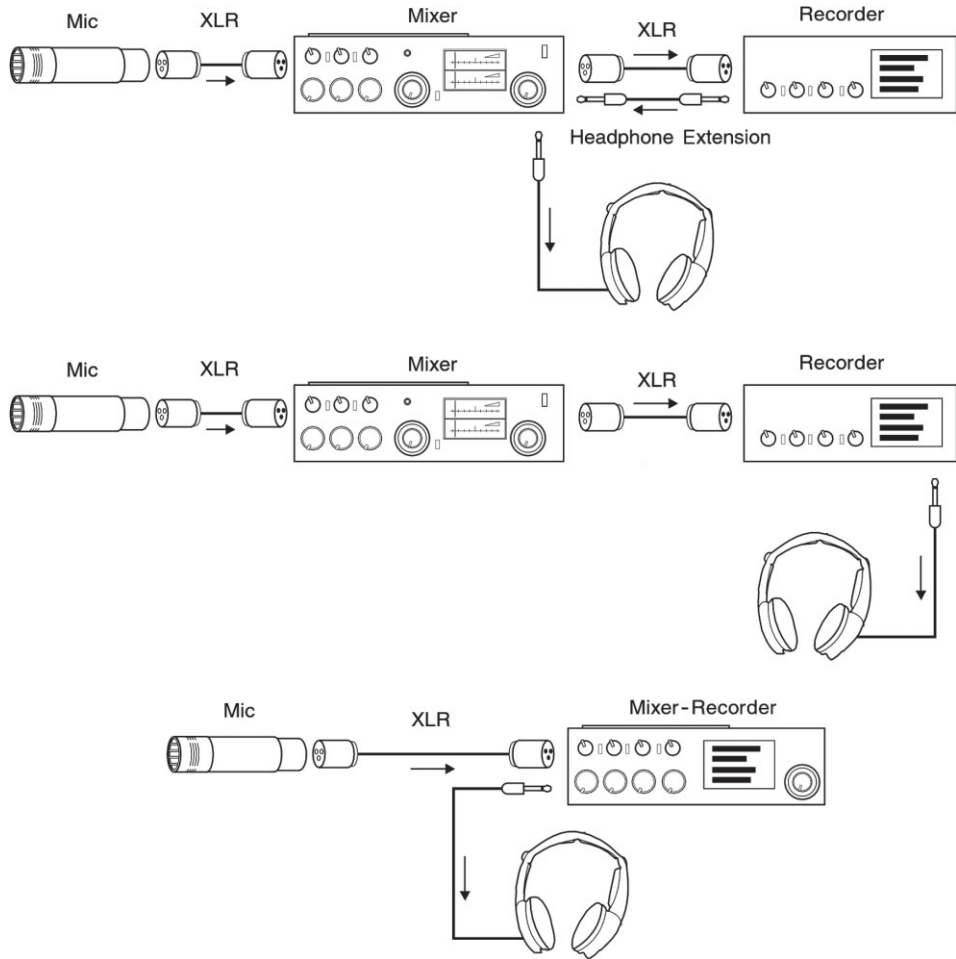


Figure 3-3 A variety of double-system signal paths.

person to recording each microphone on its own channel, to be mixed or isolated for use later on in postproduction, but it can be useful since isolated recordings give the greatest freedom in making audio decisions in a studio environment later on in the process.

3. A recorder. In studio setups or under controlled conditions, this may be a computer or laptop with an audio interface recording directly to hard drive, although more commonly an audio field recorder is used to record 2, 4, 6, 8, or rarely more channels of audio on a hard drive or flash memory card.



Figure 3-4 A typical high-end mixer-recorder, the 8-channel Sound Devices 788T, which adds eight-channel mixing and recording to the capabilities of similar 2-channel and 4-channel digital field recorders.

4. Cables connecting microphones to the mixer, mixer outputs or direct outs to the recorder, and potentially a monitor return from the recorder to the mixer for headphone monitoring. The sound recordist can also monitor audio being recorded from a headphone output directly on the recorder, but mixers may provide more monitoring functionality such as listening to individual microphone channels pre-fader without affecting the mix being sent to tape. Monitoring through the mixer can therefore be useful in conjunction with a return feed from the standalone audio recorder.
5. The camera may also be included in the audio setup if it is capable of recording some form of audio and is sent a direct feed or mixdown of the master audio for backup or sync purposes. In this case, the camera may receive a backup of the actual mix or, particularly if more than two channels are being recorded on the sound recorder, a guidetrack for use in checking sync. This may be a wireless feed, especially if it is only intended to be used as a guidetrack and the possibility of a dropout occurring is considered unimportant, although wired connections are still far more common.

Recording a guide track either sent from the audio mixer or picked up by an on-camera microphone is always a good idea in case it is needed to check synchronization of sound and picture down the line. With a good guide track and the right software, it may even be possible to automate the synchronization of double-system audio to video from one or more cameras. PluralEyes, one software solution developed to synchronize audio from a double-system or multi-camera environment using the waveform from a camera guidetrack, is discussed in Chapter 7.

Testing Sync

In double-system shooting situations, it is especially important to test sync *in advance* using the camera and sound equipment planned for the shoot to preempt any problems that may occur in postproduction. A typical sync test involves rolling camera and sound, slating with a clapperboard slate at the beginning of the shot and again at one- to ten-minute intervals for the length of a tape or file, marrying sound and picture in the editing system that will be used by the production, and checking that, once synced, the sound and picture remain in sync for the duration of the shot. The sync test should cover at least the amount of time that the longest shot will last on a production. In addition to checking sync with a clapperboard slate (and repeated sync points throughout the shot to check for drift), any time code synchronization planned for a show should be implemented during the sync test to verify that it will work correctly and give frame-accurate sync on production dailies. One frame errors are common, and sloppy. While tolerable, they easily add up to the two frame limit that most people can see.

If sync drifts during a shot, so that picture and sound are initially in sync but one lags behind the other as the shot progresses, there may be various causes, and it will be necessary to check the most likely culprits until the problem is found. The most common cause for a slight sync drift on the order of two frames per minute (actually closer to one frame every 33 seconds) is a missing pulldown if either audio or video (but not both) was slowed down from film to NTSC speed, or pulled up from NTSC speed to a true integer frame rate. On double-system setups, the same effect could result from sound being recorded at a 48,048 Hz or 47,952 Hz sampling rate (a setting discussed in Chapter 1) but then played back at 48,000 Hz even though the video original did not require a speed change on transfer into telecine. For larger or variable drifts, it is possible the clock oscillator in the camera and sound recorder are running at different speeds, and the next question to ask would be which is running off speed so that it can be fixed before production. When digital video capabilities were first introduced to the Canon 5D, a firmware bug caused audio recorded double-system to fail to sync properly, and the solution was to update the camera firmware, which was also fixed in newer model DSLRs in Canon's line. So while there are certain scenarios such as the lack of a required audio or video pulldown that have historically caused sync to drift, sync issues can occur for a number of reasons, and the best advice for avoiding them is to test sync early and ensure that it will be preserved during the entire workflow from production through postproduction.

Mic, Line, and Speaker Level

Analog interconnections have many variations, which can get you into trouble if they are mixed within one system. The first notion is the level regime. The three basic level domains in analog audio are:

1. *Mic level.* Typically rather small in the millivolt range (thousandths of volts), although at very high acoustical levels these signals can approach a volt or more.
2. *Line level.* There are two common standards for line level, called colloquially -10 and $+4$. The -10 level is found as a reference on consumer-level equipment, often on phono connectors or mini phone connectors (3.5 mm). The $+4$ level is found as a reference on professional-level equipment, most often on XLR connectors. The meaning of these levels is described in the following text.
3. *Speaker level.* This is a nominally higher level than line level but is also backed by a lower impedance power amplifier output. The voltage depends on the loudness, and a typical one is 2.83 V for 85–95 dB SPL, depending on the loudspeaker.

Among these three the first item to note is the difference between mic level and line level, because the connectors for these may be the same, and it would be unusual to find speaker-level connections on the same connector. The consequence of connecting a line-level output of a unit, such as a microphone mixer, into a mic-level input, such as that of a camera, is almost certainly great distortion. That's because you are connecting something putting out around a volt or so for normal signals into an input intended for millivolt levels: the much higher line-level overloads the mic-level input, and the input clips the signal, usually badly. The corresponding consequence of connecting a mic-level signal into a line-level input is having a weak signal. If you turn the level up enough to hear it, then it may be noisy, because this millivolt range signal is being connected to an input with a capability in the range of volts, and raising the gain will also raise the noise.

The mic/line problem is a common one in video production, and solving it is not difficult. It occurs because each piece of equipment is designed to face a harsh world in which it has to adapt to all possible conditions. Thus a microphone mixer will have both line-level and microphone-level

outputs because there is no a priori knowledge on the part of the designer as to what it might be connected to next in the chain, and there may only be one or the other type of input available. Many, if not most, professional cameras have input XLR connectors that can be switched between mic and line levels. So because both ends of a single cable between a mic mixer and a camera are switchable, several bad combinations can occur. Luckily, to overcome this potential source of distortion, it is simply necessary to set both the mixer output and the camera inputs to the same level, preferably line level.

Reference Levels and Headroom

To connect and set up recording equipment on set, it is important to understand the concept of reference levels and headroom, especially in digital recording where hard clipping caused by an electrical level mismatch or incorrect audio level setting is the most common type of gross distortion. The first thing to understand about reference levels is that they don't represent any particular piece of program material very well. They are the levels of line-up tones that are sent between pieces of equipment so that interchange can be performed without adding noise or distortion, but relying only on them may cause trouble. The reason is that the program material doesn't, in many cases, bear much of a relationship to the reference. What is actually important is that the highest peak level of the program should never exceed the highest-level capability of the recording medium. The meter scale for most digital audio metering has a maximum of 0 dBFS at the point where overload (in this case, digital clipping) would occur. All other levels are below that, such as -6, -10, -20 dB, and so forth (even if there is not enough room on a meter for the minus sign to be displayed, in which case it is implied). The standard is set this way because the maximum level recordable without distortion is the most unambiguous point of all, and various items of program material use headroom in different ways, implying different reference levels, so while a case can be made for a variety of lower reference levels, 0 dBFS is the most obvious one.

However, for interchange between digital and analog systems, a reference in the analog domain that is at the top of the permissible levels is never used. The primary reason for this is that VU meters were commonplace in the United States. VU meters are sluggish devices, taking some 300 ms,

nearly one-third of a second, to reach full reading with a suddenly applied signal. This means that for shorter signals, the meter does not read their full value and they are underestimated. Because we can hear distortion in as little as 2 ms, the VU meter could easily miss content that is audibly distorted. This fact was accommodated in years of analog recording by setting the reference level for 0 VU some number of decibels below the maximum capability of the medium and allowing the medium's headroom⁸ to accommodate the peaks.

So, for exchange between digital and analog media, a reference level that is some number of decibels below 0 dBFS is useful. How many decibels below depends on who you are and what you are doing, because there is no complete agreement on digital reference level. SMPTE specifies -20 dBFS, and many professional digital video recorders observe this reference. It was developed so that most of the content of existing analog film masters could be transferred without adjustment, although the fact that the analog media actually has about 23 dB headroom makes life a little difficult. One Hollywood studio transfers from the film reference level tone⁹ to -24 dBFS and then turns the level up in the digital domain by 4 dB to get back on the reference, encountering potential distortion in the process that is then tamed by using audio limiting, described in Chapter 9.

At the other extreme are those users who have little headroom in the analog path they are driving, and they may use a reference level as high as -12 dBFS, assuming that the analog equipment they are driving only has 12 dB headroom. Examples are audio for VHS tape delivery and certain analog satellite and microwave transmission systems. A third reference level of -18 dBFS is used by broadcasters in Europe for digital recordings.

Figure 3-5 shows various digital reference levels, how they match up with standard analog line levels (see below), and illustrates the relationship between reference level, nominal headroom above the reference level, and the dynamic range of a system. Any piece of audio equipment, including a recorder, has some minimum level it can represent, called the *noise floor*, and some maximum level it can represent without overload causing

⁸ The amount by which a signal can exceed a specified reference level—such as 0 VU on an analog meter or -20 dBFS on a digital video camera—before overload occurs, causing distortion.

⁹ Which is at the analog magnetic reference level of 185 nWb/m.

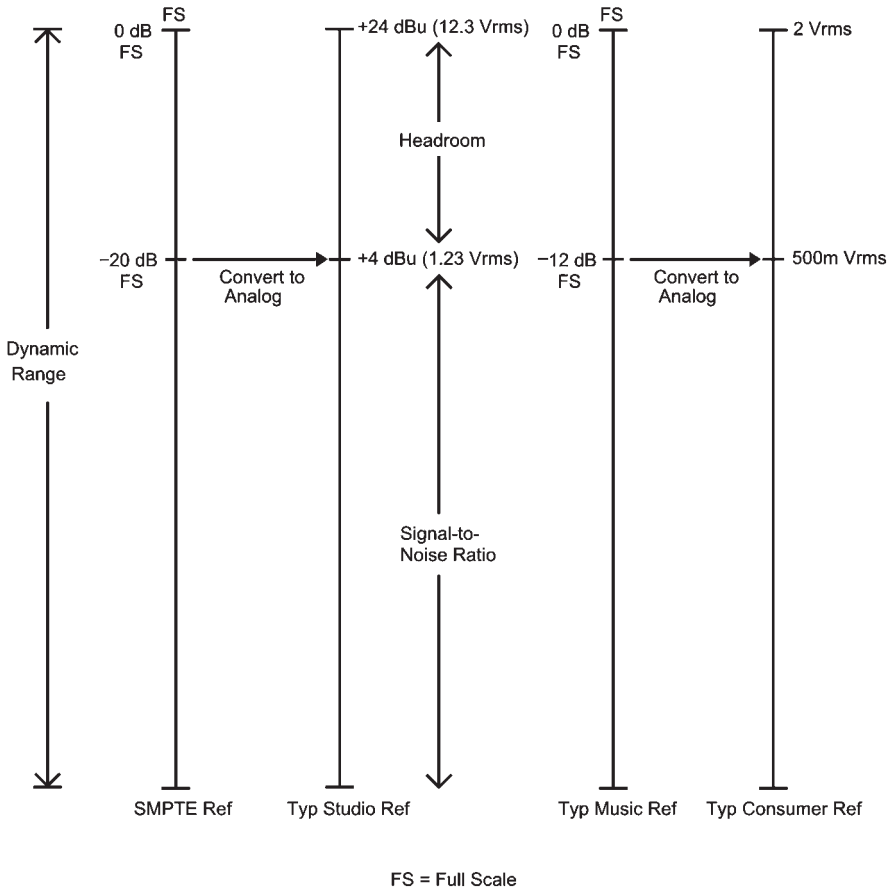


Figure 3-5 Diagram of various digital reference levels.

audible distortion.¹⁰ The level range from the noise floor to the maximum undistorted level is called the *dynamic range* of the device or medium, measured in decibels. Since the nominal reference level is some decibels below full scale, the *signal-to-noise* ratio of a medium is defined to be the difference between the reference level and the noise floor of the medium, which is equal to the full dynamic range minus the headroom, as seen in the figure.

¹⁰ In analog systems in particular, the maximum level may be frequency dependent.

While tones at these nominal reference levels are used for line up, particularly accompanying color bars at the heads of program tapes, what is most important is that the maximum peak of the program material be known and accommodated in transfers.

Setting Up the Recording Chain

In setting up equipment to be used on set, the usual method for connecting a mixer to the camera is to set the output of the mixer and the camera's XLR inputs to line level if possible. For each microphone plugged into the audio mixer, the input channel should be set to mic level unless the "microphone" is really a wireless receiver or similar device which outputs audio at line level. Many wireless receivers operate at mic level or can be set to output signals at a range of levels covering both mic and line levels, so for complicated setups it will be necessary to check the specifications of the particular device in use, and it is always useful to record and listen to some actual program material to verify that it has been recorded without distortion.

Even when connecting a mixer to the camera at line level, most video cameras have an audio level knob that will raise or lower the recorded level, to allow on-camera mixing and compensate for the various differences in reference level between different types of equipment. So, when connecting up, a reference tone should be sent from the mixer (if possible) to the camera and the record level set by hand so that the audio meter on the camera registers a nominal reference level (usually -20 dBFS). On some cameras, the on-screen audio meters are not labeled with specific numbers, but will show white bars proportional to a signal received and red bars for signals far above the reference level, nearing 0 dBFS. In this case, there is usually some indication of a reference level within the scale, but even if there is not, most signals being recorded should show up on the meter without entering the red "warning area" before 0 dBFS. Again, what is most important is to make sure that the actual recorded program material fits within the allowed headroom.

Obviously, a microphone connected directly to camera without the use of an audio mixer will feed the camera with a mic level signal, and the camera's XLR input(s) should be set to mic instead of line in this case. On-set interconnections and the process of fitting a signal into the available

dynamic range of a medium are discussed in greater detail in Chapter 6, but this brief discussion is provided here to highlight the importance of analog signal levels and reference levels in setting up to record clean audio.

Line Levels

While the most common dichotomy on set is between mic and line level, there are further variations within each domain. Recall that line level describes a signal level used to connect one piece of studio or recording equipment to another. There are two prominent electrical reference levels in use to exchange signals between pieces of equipment, called +4 and -10, which are normally professional and consumer-grade, respectively. You might wonder where prosumer equipment falls in this area, and in fact it is ambiguous. For instance, a piece of prosumer gear might be equipped with XLR connectors, implying a +4 reference, but then use consumer-type levels. CD players are on yet a third standard, which will be explained later.

Whatever digital reference level is chosen between -12 dBFS and -20 dBFS, a digital-to-analog converter will convert that digital level to an analog voltage level at its output. This is where the notion of +4 or -10 comes into play. The terms +4 and -10 are shorthand for the following:

+4 dBu, which is a voltage of 1.23 V. The *u* refers to that voltage which corresponds to 1 mW of power dissipated in a 600-ohm resistor, namely 0.775 V, and +4 dB above that voltage is 1.23 V.¹¹

-10 dBV, which is a voltage of 0.316 V. The *V* refers to a voltage referenced to 1 volt. -10 dB below 1 V is 0.316 V.

Each of these may be used to correspond to digital reference levels between -20 and -12 dBFS. Note that because the references are different, as explained above, these two levels are 12.2 dB apart, not 14.

¹¹ The *u* is a more modern usage than the *m* that preceded it, which was dBm for 1 mW reference. Because actual 600-ohm loads are not used nowadays, and milliwatts are not actually dissipated, the updating of the term from *m* to *u* makes sense.

With -20 dBFS digital reference set to correspond to $+4$ dBu, then 0 dBFS will be $+24$ dBu which equals 12.3 V, a quite high voltage level. This is why it is found only on professional equipment, and not all of that!

With -12 dBFS digital reference set to correspond to -10 dBV, the other extreme from the last case, then 0 dBFS will be $+2$ dBV which equals 1.26 V, nearly 20 dB below the pro reference and a level easy to obtain, even on most battery-operated portable equipment.

This 20 dB range discovered is the reason why there are sometimes interface problems with so-called line-level signals. These problems are compounded when CD players, computer sound cards, and other sources are added to the mix. For CD players, the rather loosely applied standard is 0 dBFS equals 2 V (± 3 dB, or from 1.42 to 2.83 V). For computer sound cards, all bets are off because there are so many potential level controls between a track being played back and the output of the DAC. These include potentially the gain setting in the track within the editing system, the gain control of the sound card software, and a gain control in the operating system. With three level controls between the track and the outside world, it is easy to see why there is little standardization of the output.

Fortunately, in most cases these differences can be calibrated out. Let us say that we have a professional DAC, a CD player, an S-VHS tape machine, and a computer sound card, all as sources for our monitoring system (Table 3-1). We put each pair of source channels into pairs of input channels of a small monitor mixer, such as a Mackie 1202. We could manipulate the various levels with the main faders, but this is not the best way to operate, because they will have very different settings. The best way to perform gain staging is to set the channel level faders to a standard, often marked 0 dB or "U" for unity gain: that level at which what is presented at the input is reproduced at the output. Then set the main faders also for 0 dB. Observing the LED-style VU meter, set the input trim control for each source, playing a standard level tone so that the meter reads 0 VU. The input trim controls do the work of adjusting for the variety of reference voltage levels to be found with these disparate sources.

To connect all of these up, it may be necessary to use some audio adapters. Besides XLR connectors and RCA phono plugs, each described previously, phone plugs or mini (3.5 mm diameter) phone plugs may be used. Stereo phone or mini phone plugs—those having distinctive tip, ring, and sleeve

Table 3-1 Reference levels.

Source Device	Reference Level on Medium	Reference Output Voltage	Reference Level dBu	Maximum Level dBu
Pro DAC	–20 dBFS	1.23 V	+4	+24
CD Player	–12 dBFS (example; there is no one widely observed standard)	0.399 V	–5.7	+8.2
S-VHS	International Tape Association Reference Level	0.316 V	–7.8	+8.2
Computer sound card	–20 dBFS (example only, there are wide variations among cards, and their output is influenced by multiple software level controls)	Varies	Varies	Varies

conductors¹²—may be used in one of two ways (Fig. 3-6). In the most common form they are used for stereo with the tip carrying the left channel signal, the ring the right channel signal, while the sleeve is a common ground for the two channels. In a variation, only one signal is carried, but in a balanced mode, with the tip being the positive signal, the ring the negative signal, and the sleeve serving as ground.

Mixing Balanced and Unbalanced Connections

In the olden days of audio, life was easy. Professional equipment was equipped with input and output transformers, and these produced a balanced, usually floating (no ground reference on the lines) system. Impedances were matched, so every output expected to be looking into a 600-ohm load provided by every input. No fan out was possible without a distribution amplifier; that is, each output was only capable of feeding one input. If you wanted to drive more than one input, an extra distribution amplifier was necessary. All of this was derived from telephone company

¹² Giving the distinctive nickname “TRS connectors.”



Figure 3-6 Miniature phone plug, 3.5 mm diameter, with tip, ring, and sleeve connections permitting balanced microphone or unbalanced stereo headphone connections.

practice, which had to be of very high quality for its day to deliver the goods across long telephone lines, and, depending on the quality of the (rather expensive) transformers, hum was low, frequency response and distortion were good, headroom was high, and so on.

Professional and consumer worlds were quite different and at opposite ends of the spectrum when it came to signal distribution. Consumer-grade equipment used the unbalanced system, with outputs represented on one conductor plus ground connectors wired to similarly equipped inputs. In addition, impedances in consumer equipment were bridging; that is, they worked just like your electrical outlets: you plug in as many bulbs as you want to one circuit, and the system does not care how many bulbs are plugged in until you reach a maximum determined by how much current you are drawing. Thus audio signals could be applied to a Y-adaptor for distribution to multiple inputs.

Although unbalanced audio was not as bulletproof for hum as the professional system—something that its lower reference voltage levels also did

not help—the unbalanced system was good enough for small, stand-alone systems in general. However, integrating consumer equipment into larger professional systems required additional interface equipment. Thus the IHF Match Box or +4/−10 box came on the scene. Providing a way to use cheaper consumer equipment in a professional studio, this device made the level and balanced/unbalanced transition necessary to enable the use of such equipment.

So far the distinction between professional balanced lines and consumer unbalanced ones has been straightforward. If you needed to go in between the two domains, then a match box would do the job. However, this is where things get more complicated. Good transformers are expensive, leading electronic designers to try to emulate their isolation properties with circuits that attempt to achieve the same end more cheaply. The salient thing about a transformer is that the two sets of wires around the core, called primary and secondary, are not connected to each other electrically—they are said to be galvanically isolated. This is how transformers provide isolation: the electrical voltage input to the primary wiring is turned into a magnetic field, which is then relayed to the secondary magnetically, and in its wiring produces the output voltage.

Unfortunately, circuits to provide such isolation range across a very wide span of quality, and some of them can be downright bad in some circumstances—clipping prematurely, for instance—so if you should happen to have a signal peak at 0 dBFS in the digital domain, it may well clip electronically some 6 dB sooner, at −6 dBFS, depending on what the output circuit is driving. And this electronic gimmickry is used across many levels of equipment, with more expensive professional gear sometimes worse than less expensive. Unfortunate too is the fact that you can rarely tell anything about this situation from the specifications of the unit. One DTRS-format eight-track digital tape machine, an “upgrade” from a former model, performs much worse than the earlier model when confronted with an unbalanced input connected to its output, as it would be in many semi-professional environments. So even higher-end equipment is not immune to potential complications in matching reference levels across devices.

One way to test for this particular problem is the following: Use a battery-powered voltmeter to measure the voltage at reference level out of a piece of equipment. Then connect it to the input of the following piece of equipment, keeping the voltmeter in parallel with the input. If the level doesn't

change, it is likely that the interface port will be all right. If it changes by -6 dB (half the voltage), it is likely you are in trouble.

Table 3-2 lists a variety of audio connectors in widespread use. Generally, balanced connectors such as XLR are used both on set and in the studio for professional work today, although equipment such as CD players, tape players, or sound cards which output audio via phono (RCA) connectors may necessitate the use of unbalanced audio in postproduction or for live

Table 3-2 Audio connectors.

Name	Uses	Notes
RCA, Cinch, Phono, Pin Plug	Analog audio, digital audio, analog video	Most ubiquitous consumer plug found on most nonportable consumer equipment; unbalanced
3.5 mm Mini plug, mono and stereo versions	Headphones, some consumer microphones	Mono version has a tip and a sleeve; stereo version has a tip, a ring, and a sleeve
2.5 mm Micro plug, mono and stereo versions	Microphone inputs of microcassette recorders	Be careful not to confuse with 3.5 mm plugs/sockets
1/4" TRS connector	Headphones, line level inputs on mixing consoles	Stereo or balanced mono signal for line level inputs; stereo for headphones
BNC	Analog and digital video, digital audio (especially to AES3id standard)	Locking connector
XLR	Analog audio, digital audio	3-pin for balanced connections
Banana plugs	Loudspeakers, test equipment, Nagra line inputs/outputs	
Tascam format DB-25	Multichannel AES connector	Same connector as old parallel port computer cables, but wired for multichannel digital audio transmission

production purposes. In editing audio today, it must be said that transferring audio digitally from a CD or other digital audio source would be much more common than re-recording the audio over an analog connection, so for postproduction purposes the need to transfer audio via an unbalanced audio connection, even in a semi-professional setting, is far less common than it once was.

Director's Cut

- If at all possible, designate a sound person for production whose principal responsibility is to record and monitor sound.
- Scout locations for sound as well as for look. You can pan off an offending sign, but you can't pan off an offending noise nearly as well.
- Decide on a *single-system* or *double-system* workflow for capturing picture and sound in advance, including settings like audio bit depth (16 or 24 bit) and sample rate (normally 48 kHz, or 48,048 Hz for film content shot for video release), and *stick with it*. Don't change settings and try to mix different file types partway through a production.
- Test camera and sound gear in advance and know the correct settings for mixing and recording equipment, including Mic/Line level settings and nominal recording levels on audio meters.
- Use balanced cables and connectors when possible to minimize the possibility of electromagnetic interference affecting sound recordings. Know what kind of connectors your camera or recorder uses and what type of signal each device expects.

CHAPTER 4

Production Sound I: General Considerations

Production sound is the principal source of dialogue and a basis for the sound design of most digital video projects, so making every effort to record good sound on set is one of the most important steps in creating the soundtrack. Good technique can make a program, while bad technique can, at the very least, cost a lot of extra time, effort, and money in postproduction. Any reality-based program is unlikely to be able to replace dialogue in postproduction, so microphone technique and recording is all-important for these programs. In some cases, vital performances or reality situations that cannot be repeated may be lost. This chapter and the chapters that follow on production sound will help you make the most of the equipment at your disposal to improve both the technical quality and the practical value of the sound you record on set.

Coverage

In filmmaking the term *coverage* refers to the number of different ways that a scene is photographed, from wide master shots, through two-shots, to over-the-shoulder and full-face close-ups. When I (Tom Holman) first saw the flying wing sequence of *Raiders of the Lost Ark* during postproduction on the dubbing stage, I was sitting next to Michael Kahn, Steven Spielberg's long-time editor. I was astonished at the variety of cuts and the naturalness of the edits in the scene, and I said to him, "Michael, that's brilliant cutting." His reply was, "It's nothing—coverage, coverage, coverage." A little like "location, location, location" in real estate, this mantra is good guidance for filmmaking in general.

In the case of fiction filmmaking where the action of the scene can be repeated a number of times, this usually takes on a fairly formal structure, with a scene being divided into setups each encompassing particular

camera positions, framing, lighting, and so on. In documentary filmmaking, no such formal structure exists, but nonetheless the job of the cinematographer is to cover the scene with a variety of shots for making editable footage. For instance, having some visual to cut away to provides a means to compress an otherwise real-time interview without suffering from a jump cut, a visual discontinuity in a basic on-camera interview. Sound carries the story in such a scene, because the sound is being cut so the story is told linearly, while the picture is cut to support the sound, and not the other way around. Such supplementary visual material is sometimes called B roll in the jargon of news coverage.

Here we want to take the term coverage and apply it to sound. How to sort out what is the desired sound moment-by-moment within a scene, and how to concentrate on that sound, while reducing less important or even annoying sound, is the job of the location recordist. It is always possible to add reverberation and background ambient sound later in postproduction, but it is extremely difficult to reduce excess reverberation in particular, and noise in general, that accompanies original recordings. All too often, excesses spoil production sound recordings in one or both of these factors. The job of microphone technique is to reduce both of these to manageable amounts that make sense in the context of the scene. This requirement is what leads to our wishing to move the microphone off the camera body whenever possible, and get it up in the air over the principal action, and/or place microphones on the talkers or hide them on the set close to the talkers.

Not only is it the responsibility of the sound person to minimize noise and reverb, but on fiction films almost everyone in the crew gets involved one way or another:

- Set dressers in hanging blankets, say on C-stands, out of camera view, to soak up excessive reverb.
- Grips in making dolly motion quiet.
- Costume people in providing cloth booties for actors to use in scenes where their feet don't show, and in providing places for radio mics in costumes.
- Gennie operators in siting generators for minimum noise.
- Assistant directors in maintaining a quiet and disciplined set.

Inexperienced crews will not necessarily know the importance of capturing good production sound, and one of the primary skills of the sound people is to make their needs known to others who may not understand the effect they have on sound, without being obtrusive. The previous examples just begin to explore the range that might be required as suggestions, but there are surely many more. Regardless of how many people are working together to make a program, remembering that the job of production shooting is capturing good picture *and* sound goes a long way toward making better sound.

The obvious thing to capture in a scene with dialogue is the speech of the actors or documentary subjects. Usually we want the sound quality to be on mic, that is, having not much recorded reverberation or noise compared to the direct sound of an actor's voice, and miked in front of the actor, most often from overhead (this being the position that most sounds like the actor's natural voice, for reasons that will be explained later). On the other hand, there are certain relatively rare times when we want to hear an actor sound off-mic. This might occur when we concentrate on the face of one actor in a scene listening to another character speaking, for instance, and may be an indication that we are on the way into the thoughts of the actor who is being seen, not the one being heard. So alternative microphone perspectives can be used as a storytelling element. However, principal on-screen dialogue deserves to be covered sounding clear and on mic, and this most often means that an on-camera microphone won't suffice, although there are a few exceptions, described later.

In original recording accompanying a picture, there are many possibilities for deciding what is important and what is not. Filmmakers who have gained some experience come to realize that the important elements to capture, whether for a fiction or a nonfiction film, are whatever bits and pieces can be best edited together into a seamless whole. This means, perhaps paradoxically, that the best training for production personnel—camera and sound people—is picture and sound editing. This actually occurs on the smallest productions because it is the same person doing all the jobs, and feedback from editing and mixing will inform production sound, at least on your second show!

Normally, good production technique calls for edits within a scene to be smooth and not jarring. Continuous sound across picture edits indicates to the audience that the scene is continuous, too: while we may have changed point of view, we are anchored in the same space before and after

the edit. This is complicated by the fact that we may not be cutting between similar shots, but from a wide shot to a close-up, for example. While the sound recording can exhibit some difference between these two perspectives, if we were just to use the on-camera mic for a wide shot and a close-up, the difference in sound perspective at the picture and sound cuts between the two would be jarring, even though the picture cut would make sense. A much smaller difference in perspective, such as that achieved by covering both shots with a boom mic but with the wide shot a little looser in perspective and the close-up somewhat tighter, makes for a better sound track.

Another part of getting useful sound recordings is to acquire not just the sound of the principal action, but also room tone that can be used to bridge gaps in the action, cover up momentary noises, and generally smooth out the edits. In shooting years ago, a pause would be taken by the cast and crew to record room tone specifically between shots or setups. This is rare today because digital audio workstations can be used to find fragments between the speech elements and other set noises where the background is said to be clean, and these can often be looped and thus extended to the length necessary. However, it can be difficult to find clean room tone long enough to loop cleanly (about 20 frames minimum is a good rule of thumb), so keeping quiet at the beginning and end of scenes or recording some room tone immediately after a particular take can still be quite useful.

In shooting sound for picture, image usually takes precedence over sound, so the requirements needed for clean picture, such as not having a microphone visible in a shot, usually take precedence over sound capture considerations. The job becomes recording the best sound on location that can be smoothly blended together in postproduction to form a cohesive whole, with minimal outside interference, under a set of given conditions of camera angle and other factors. This means that microphones are rarely allowed within the field of view of the camera, among other things.

Scene Coverage and Microphone Technique

Conventional dramatic television shows are a good thing to watch to understand scene coverage in the conventional sense of the term. Start with the sound down, and note how scenes often progress, from a wide

shot possibly tracking the actors into a locale, to closer shots, over the course of a scene. The idea is to establish the geometry of the space in which the action occurs for the audience first, and then to use a certain set of grammatical rules about film shooting and editing to produce a continuous-appearing whole scene that progresses to a climax.

The three main microphone techniques in use to produce coverage are boom, lavalier on the actor/subject, and planted mic on the set/location. The relative strengths and weaknesses of the various techniques are given in Table 4-1. Normally, the boom mic is preferred, if it can be used. This is commonly done for feature films and even for documentaries such as *Spellbound* (2003), for which more than 90 percent of the production sound was from a boom microphone.¹ For situations where the boom simply can't be used, lavalier microphones often have to be resorted to. This might occur if there were such a wide shot, for instance, that for any usable placement of the boom microphone the recorded sound would simply be too noisy and/or reverberant.

Planted microphones have their place too. In situations where an actor or subject is moving through a set or location, there may be positions where a boom can't follow, such as through a doorway where it would have to duck into the picture. In such a case, a skillful production sound recordist may be able to smoothly cross-fade from the boom mic on one side of the door, to a planted mic over the door in order to get through it, and then back to the boom. Another example of the use of a planted microphone is where the action occurs far upstage from the camera, and the boom just can't reach the area needed.

For content that is boomed, matching the microphone perspective to the action is something that occurs rather naturally, because the microphone is farther away in the wide shots and can be closer in the close-ups. This is one area where lavalier radio mics are at a disadvantage because they cannot match the perspective of a shot. If an actor wearing a lavalier turns his body and faces away from the camera, the sound is largely unchanged, when it really should change. Matching microphone perspective to camera perspective is usually a good thing. This is why microphone placement is

¹ The producer Sean Welch was boom operator, while the director Jeffrey Blitz operated the camera in production. Postproduction editing and mixing were by Peter Brown and Joe Dzuban.

Table 4-1 Common microphone techniques for production sound.

Method	Pros	Cons	Special Considerations
Boom	Best sound in most situations—most representative voice timbre and ability to match camera perspective	Obtrusive: can make documentary subjects and even actors nervous; potential exists for boom shadow; more susceptible to reverb and noise than the lav	Needs operator separate from camera operator and coordination between camera and sound personnel
Lavaliere (wired), on chest	Closest to mouth, thus typically less influenced by room noise and reverberation than a boom (unless compared to a mic held up to the mouth or a headset mic); mic can perhaps cover more than one subject/actor if they are close together	Mic can't see mouth; that and mounting on chest both change timbre for the worse. Lacks match to camera perspective if subject moves; susceptible to clothing noise, subject touching chest; invasive to place the microphone on the actor/subject*	Wire tether restricts movement but is a more reliable link than radio mic
Lavaliere (radio mic), on chest	Same as above	Same as above	More freedom of movement at expense of less reliable connection; more information is to be found in Chapter 5 under the heading "Use radio mics"
Lavaliere in hair or hat	Commonly used Broadway method; potentially less costume interference than on chest. Perspective stays the same when actor turns head	Difficult to place and wire. Timbre requires equalization. More subject to perspiration than chest mounted	
Planted (hidden mic on set)	Works well for certain situations	Inflexible	

*Even an actor as experienced as Tony Curtis said in an interview that he didn't like being touched by people doing hair, makeup, costume, and sound.

preferred above the frame line, and not to one side or the other of the side frame lines. If the mic was on the left side of the frame, and the actor turned during the scene and spoke to the right, they would be going too much off-mic, revealing the mismatch in microphone and camera perspectives. The position above the frame is preferred to below the frame because voices sound better on or above a horizontal plane through the head than they do below that plane, where they tend to sound chesty, with a mid-bass emphasis. Sometimes, however, it can be necessary to compromise and shoot from below the bottom frame line and make up for it in the mix. Such an instance might occur if the camera is set low with a wide-angle perspective, showing lots of headroom (literally) above the actor.

Shot-to-shot matching of sound perspective is usually a good thing to do. This means that if we are shooting coverage of two people speaking in alternating over-the-shoulder medium close-ups, we will record the actor whose face is being seen on-mic, and the actor not being seen will be recorded off-mic. The editor doing sound can then choose to construct the sound from all of the on-mic lines, covering the sound for the shot over the back of an actor's head with their on-camera lines (the lips should not be seen or otherwise sync problems will arise). A problem for this technique is dialogue overlaps, which are hard or impossible for actors to control equally on each take. In the case of overlaps, the recordist or boom operator then has to choose to weight the on-mic versus off-mic actor at up to 50/50 in the microphone perspective, ultimately by recording from the side of the frame favoring the face view of the actor. This is an example of where a rule given above (don't record from the side because if the actor turns it won't match camera perspective) should be violated for the better good (matching both on- and off-camera actors so that dialogue overlaps can be used).

When the shooting must be unobtrusive, such as for a documentary where the subjects may be spooked by the presence of a camera or boom close to them, then a new set of rules apply. For example, reality TV today relies greatly on radio microphones, with one on each subject. This captures their dialogue no matter where they are, but does have certain problems:

- *Clothing noise often arises.* This problem can be minimized by using tape to surround the lavalier microphone, then taping it to the chest, and using a second double-stick tape to adhere the mounted assembly to the clothing, all the while not obstructing the sound ports on the mic (see Fig. 4-4).

- *People often fiddle with their clothes near the microphone, or with the microphone itself.* This could include chest beating, which too often occurs right on the microphone. The worst case ever seen was of a new baby in the arms of her mother, who proceeded to reach up and play with that shiny object during the morning television news show *Good Morning America*.
- *People embrace.* When they do, the voice sound becomes at least muffled and often noisy through contact, a dead giveaway that a radio mic is in use.
- *The subjects usually remember they're wired.* Hidden, on-set microphones overcome this problem.

What Can Be Done with an On-Camera Microphone?

This is a great limitation, as has been outlined previously. Shooting a wide shot from the back of a church and expecting the on-camera mic to record the wedding vows is just crazy. However, some useful things can be done even with an on-camera microphone:

- If all that happens within a shot is to establish a locale, for instance, and no principal dialogue occurs, then an on-camera mic may be okay. Such *establishing shots* give a pause or bridge in the ongoing action, helping to anchor placement in time and space. Such shots are helped by the sync between picture and sound that makes them seem more real. For instance, my students once shot a documentary about an artist who worked with large Magic Marker-style pens. Choice of a quiet location and use of a directional mic allowed them to pick up the sound of the marker on paper. The sound effect, especially its in-sync nature, added a real sense of being there for that scene. Hearing it as the only sound emphasized the lonely nature of the task of creating art; on watching the film, the audience thought that if they could hear that sound, it must be quiet!

Feature filmmakers call this type of sound Production Effects, and they cut one or more tracks specifically for them. This is as opposed to ambience or backgrounds, which refers to general background sound of an appropriate space laid over a shot or, more likely, a series of shots constituting a scene. Ambience provides an aural space for the

movie to live in, but production effects are potentially even more powerful because of their sync nature to the actual picture. In order to make an ambience seem more real, hard effects can be cut over the top of it on other editorial tracks, portraying particular events in the picture through sound. Foley effects, too—those made synchronously in a studio while watching the picture—have the same effect: sync sound sells. So feature films go to a lot of trouble to build a sonic environment that matches what you see in the picture: ambience, hard effects, and Foley may all play a role. Still, the cheapest route to obtaining the advantages created by all of these layers is often production effects, if they can be captured cleanly at the time of shooting.

- If a documentary subject can be isolated from others and shot in close-up, with the camera only two to three feet from the subject, then a high-quality, directional, on-camera microphone may do. This means limiting the shooting style in order to accommodate the method of sound recording, but it is effective. I (Tom Holman) made a short documentary for the 66th reunion of the Mt Morris (Illinois) High School class of 1936. The lunch was around a big table with lots of gabbing. While I shot that for coverage, I also arranged to take each of the 12 alumni into a quieter corner and asked them questions in close-up. The close-up shots provided not only a part of the picture coverage, but also the sound that was used as principal voice-over, covering the wider shots of the group at lunch. It worked because I limited the kind of shooting I would do in the interviews to being quite close up, not only in framing, but in actually being close to the subjects, so that I could capture decent sound.
- On a fiction film with good actors, it is possible to shoot the picture with accompanying bad sound just for sync purposes (called a scratch track) and then go back and shoot the lines again for sound only, with the camera and its microphone in a preferred position for sound recording. Called *wild track*, this method requires tedious work in postproduction to resync all of the performances in the wild track to the scratch track, but it is possible to do. A key to it is getting the actors to perform in much the same way for the microphone as they did for the picture.
- It is generally a bad idea to shoot MOS (for “mit out sound,” a term coined by early Hollywood sound recordists, who tended to be

German), because even the on-camera mic sound may prove useful in postproduction, if for nothing else, then for a guide track as to exactly what was said. For scripted productions there are deviations from scripts that make looping or replacing dialogue difficult when there is no precise record of what was said. For documentary films, having some sync sound available instead of putting in a general ambience makes editing easier.

- The camera is a sound recorder too, and often the only one on the location. It can be used as a sound recorder only, with the picture just along for the ride, which may be used to document where the sound was taken. With the low cost of media, and decent sound recording quality, this option is potentially much better than using a separate MP3 audio recorder, for instance.

Another problem with on-camera microphones on tape-based cameras is that the mechanism of the spinning drum that writes and reads the dense audio/video data on tape causes mechanical vibration that is inevitably transmitted through the camera body to the microphone and induces noise there. You might not hear this in a quiet room from two feet away, but if you put your ear in contact with the camera body, you will hear it. This noise is what the microphone hears, and various microphones are more or less isolated from their camera bodies to influence how much is picked up. In one case, the Canon XL-1 series, several accessory devices were sold in the aftermarket to isolate the microphone better from the camera body than the stock equipment and reduce the noise. This vibration is usually heard only in very quiet shooting situations, though.

Louder camera noises can also occur intermittently with zoom lens motors and other camera functions. Because these relatively soft noises are located very close to on-camera microphones, they can be a real nuisance. The largest help for overcoming them is to space the microphone away from the noise source.

How to Use the Two Channels

While multi-channel recording is possible with some dedicated audio recorders and some camera formats, two-channel stereo is by far the most

common audio recording format. When recording production sound, there are several ways to use the two channels:

- *Record the same signal from one microphone—or possibly more mixed together, usually via an external mixer—to both channels, providing redundancy.* This leads to only a little potential confusion in postproduction, and it is simple, but it doesn't provide the most flexibility. We might call it redundant track recording.

One advantage of recording the same signal to two channels on some model cameras is the ability to stagger the levels between the channels—call it redundancy with a twist (of the knob!). The input levels can be adjusted for perhaps a 15 dB difference in peak levels, and if the channel recorded to higher levels becomes distorted, the editor can cut to the lower gain channel (with appropriate level compensation) in the editing system and save the recording from bad distortion. This may require an external Y-adapter.

- *Record a left-right stereo pair from a stereo microphone.* This is the method used by many consumer camcorders with on-camera microphones. A single microphone body contains two pickup elements—in this case, one facing forward and one facing sideways—of particular types. Through electronic manipulation, these are turned into a left-right pair. Stereo recording can potentially sound more real than mono recording, because sounds are reproduced in space in a way that matches the picture. Comparing stereo original recording with mono often reveals an advantage in openness as well as angular matching between picture and sound. However, the complications are that editors must be aware that they must keep the two channels together and treat them as a left-right pair, and there can be subjective complications at edits, where, for instance, the sound jumps around spatially when we have come to expect dialogue to generally be delivered from center screen.

One way to use the two channels from a particular stereo microphone type, called MS stereo for mid-side, is to do further manipulation of them in postproduction to obtain signals corresponding to the original microphone elements' forward and sideways facing sound fields. Then we have the principal sound in the direction that the camera is looking and a second channel that represents the ambience of the space—that which the camera is not looking at. The first channel emphasizes the sound of an actor that the camera is framing, say,

while the second channel literally rejects the actor in favor of hearing the room reverberation and ambience. This can be useful later in postproduction mixing, to be able to manipulate how close or far away the actor sounds by changing the relative balance of the now center-front and surround instead of left-right channels. For example, constructing an aural zoom is possible by cross-fading between the side and front elements to make sound seem far away at first and then bring it closer, or vice versa.

However, having discussed stereo production sound recording, it must be said that typical Hollywood feature films long ago gave up on a stereo approach in favor of the dual mono approach discussed next, largely because of editing and mixing concerns. This may be rather unfortunate, because well-cared-for stereo sound tracks do have things to offer, but they also add some complication, so editors and mixers must know what to expect.

- *Record dual mono.* Two separate microphones are recorded on channel 1 and channel 2. This system allows greater potential editorial freedom than recording the simpler single (redundant) mono signal. However, the picture and sound editor(s) need to be clued in as to what the coverage plan is so that they don't reject something based on what they hear on just one channel, usually the first, which is normally what editors concentrate on. In a variation, two separate *types* of microphones may be mixed together to form one channel, such as putting a boom on channel 1 and two lavs mixed together on channel 2.

Various cases can be considered. If there is only one principal source to cover, it can be covered by two techniques simultaneously. This double coverage of the same activity provides greater redundancy than even dual mono by and of itself, because if one of the techniques should fail, for example through the subject beating his chest while wearing a radio mic, then the other mic can be relied on for a clean signal. Second, the editor is given the choice as to which combination produces the best sound: one, the other, or both used together. The boom approach typically provides the best timbre and perspective. The lavalier approach typically provides the tightest sound, with the least reverberation and room noise, which is both good and bad: good because it is probably the most intelligible sound, and bad because of the limitations caused by the lavalier placement on timbre and naturalness, the potential for handling noise, and the potential for sound to be muffled due to placement underneath costumes.

The best sound might result from using a little of each microphone in the final mix, although because there is likely to be a difference in the time of arrival of the direct sound at the two microphones, a frequency response notch or notches can arise, which may make the recording sound as if it was made in a barrel. With dual mono, though, all of these choices are postponed until postproduction, when there is more time to contemplate and listen to them with this approach than if the microphones were mixed on the fly and combined to a single mono channel in production.

An example of dual mono usage is coverage of a wedding. Conventional thinking might use two radio mics, say one on the bride and one on the officiator, recorded on the two channels, the groom being picked up by the bride's mic—after all, it's her day. The problems with this coverage are that there is no perspective of the wedding space—the sound is rather dull, although the words can be heard. Additionally, the wedding gown's layers make noise. Better practice is to record two radio mics, one on the officiator and one on the groom, whose tuxedo jacket is a lot less noisy than a wedding dress and on which we can pick up the bride adequately, to channel 1, and record a boom mic, or even the on-camera mic, on channel 2. The principle is to group mics having the same type of perspective on separate channels, for best use in postproduction. Another principle illustrated here is to put the best sound on channel 1, because many video editors will typically only listen to the first channel. These two perspectives allow for postproduction mixing to determine how loose or tight sounding the final product is; without the boom or on-camera mic on channel 2, there would be no way to loosen up the overly tight perspective of the lavs alone, especially considering there are other elements of sound, such as the music, to record.

If there are two main sources of sound, then one source may be recorded on each channel. An example is a documentary subject and a questioner. Although a documentary director/questioner will often say "don't bother to record me, we're going to cut out the questions," you'd be surprised at how many times this turns out not to be true, and even if the questions are all cut out, it is still useful to have a clean record of them for transcriptions to know exactly what the questions were. Another example is scene coverage used for a fiction film that has multiple speaking parts. A single boom mic might not be able to cover all of the action simultaneously, because of various distances and angles from the camera. Two boom mics could well be the best

answer in such a situation, rather than trying to stick to the one boom/one lav scenario that is quite common. The freedom this offers editorially is that the two mic channels do not need to be cut simultaneously, and this permits better continuity by allowing greater potential to conceal sound edits.

Coverage with two spaced microphones such as at a podium so that the talker is covered leads to problems. If they are spaced, it is likely that for at least some of the time the distance to each microphone will be different. Adding together signals with different times results in comb filtering, selective enhancement and diminution of related frequencies, which can even result in a swishing sound as the time changes. A better approach is to use two mics close together, for redundancy and to reduce combing.

Other Items Recorded during Production Sound

Until now we have concentrated on capturing voice in production sound. The reason for this is that the story most often depends on it, and dialogue recording is usually what separates the good from the bad in location sound. Another reason comes by way of an analogy. Color television engineers say that you only have to get the colors of skin, grass, and sky right; nothing else matters nearly as much. That's because skin, grass, and sky have familiar colors, whereas you aren't as familiar with your shirt color. These three elements are used to form absolute color judgments because we have day-to-day experience with them. Dialogue is rather like skin, grass, and sky, so familiar that changes to it are easily perceptible. Such changes include variations in level, timbre, location, and amount of reverberation. Thus, typically, the job of the sound crew is to minimize these changes, although some differences may be useful. Nonetheless, if they are quite different, changes in the sound of a voice may distract an observer from the ongoing story.

Production sound effects can show wider variations from dialogue recordings that are correct in the sense of being the best recorded, and still be very useful. That is, they may have compromised timbre because the microphone isn't in the correct place, or be a little too dry (lacking reverb) or too wet (with excessive reverb), without suffering these distortions as much as voice would. Thus it is worthwhile to roll sound on virtually everything, because you don't know what will be useful for building

tracks in postproduction. Sound effects recorded on location are sometimes of unrepeatable events, like blowing up a building. While in a Hollywood feature such an explosion might well be enhanced with sound effects from libraries or other recordings, sometimes the event being recorded may be so rare as to be unique. For expected loud sounds, the information given in Chapter 10 will be useful.

Microphone Accessories

Booms/Fishpoles

Hollywood studios traditionally used large, heavy, floor-mounted booms, principally Fisher booms. These aren't typically used in production nowadays, because so much of it is done on location, and moving such large equipment is too difficult. Their use has mostly given way to fishpoles, which are typically operated from the floor by a boom operator. There are a great many fishpoles on the market, with cheaper ones made of aluminum tubing and more expensive ones of carbon fiber for lightness. Some considerations in the selection of a boom are needed length and corresponding weight to hold up (coarsely speaking, $\text{weight} \times \text{length} = \text{torque}$, a more important matter than simple weight), internal or external microphone cable, potential articulation (poles with a joint that can be rotated), possible counterweights, and ruggedness.

Good fishpole operation means getting the microphone to where it needs to be, aimed correctly, and silently (Fig. 4-1). Professional boom operators are often overlooked for their contribution to sound, but production sound mixers of Hollywood movies know and appreciate their extreme mechanical versatility, their ability to memorize the script quickly and follow the action, their ability to get the microphone where it belongs to match the camera perspective and have the right amount of reverberation, their ability to remain completely professional in the line of fire because they are the closest to the actors and must not distract them, and so forth.

One common problem with booms is operating them so that the cable, either internal or external, does not move relative to the boom and create noise while in action. Internal curly cords like those on telephone headsets may work well. If an external cable is used, it is generally wrapped around the boom for a few turns and held taut with the operator's hand nearest the bottom of the boom.



Figure 4-1 Pictures of good and bad boom technique.

(a) Good position and technique.



(b) Poor position: under the frame line sounds generally sound worse by being chesty, although the recording can be rescued with appropriate re-recording equalization.



(c) Poor position: off the axis of the talker is generally poor recording and probably cannot be rescued.



(d) Be careful not to hit the talent on the head with the boom mic; it is easy to do this if you are distracted.

Shock Mounts

Shock mounting of microphones to be used on booms or other places where mechanical vibration can be transmitted to them through their mounting is essential (Fig. 4-2). Even tapping a little finger on a fishpole results in loud output from the microphone if the shock mounting is poor, because of the size that this vibration represents at the microphone compared to sound. A typical problem with shock mounting is to short-circuit the effect of the shock mount by having too stiff a microphone cable pulled too tightly. The microphone should ride in the shock mount in a completely springy way. Also, the compliance or springiness of the spring, which is usually made of rubber or plastic parts, needs to be correct for the weight of the microphone in use. If the rubber is too stiff, the effect of the shock mount will be minimal; if it's too loose, then the mic may bounce too much and actually hit other suspension parts when moved.

It is important to maintain shock mounts (and windscreens) so that all mechanical connections are tight. Even slightly loose small parts can cause audible rattles, because the components are so close to the mic or even in physical contact with it.



Figure 4-2a A shock mount showing the flexible rubber parts responsible for suspending the mic freely.



Figure 4-2b The lead dress allows the cables to be freely suspended in air—they do not touch each other nor the windscreen when the mic is in place. This particular mount uses two different cable diameters: one connected to the mic for best flexibility and one out of the assembly for ruggedness.

Windscreens

Windscreens are essential in virtually all video applications because just moving the mic, even indoors, creates wind that can cause obvious low-frequency disturbances (Fig. 4-3a). Wind disturbance is usually heard first as low-frequency noise, which progresses up through being so bad that the signal from the mic may intermittently bobble or be cut off, as the wind blows the diaphragm into extreme positions. There are several degrees of windscreening and some require matches between microphone types and windscreens for best effectiveness. Even drafts caused by air handlers indoors can cause small amounts of low-frequency noise to be added to microphones, so there are just not many uses one can have for an unwindscreened mic. The degree of windscreening needed depends on the wind velocity expected. All windscreens alter the sound somewhat, with more aggressive ones offering more protection but altering it more. Thus you should use the minimum amount of windscreening that works.

Because of their principle of operation, omnidirectional mics may use a foam windscreen that encapsulates the microphone fully. A larger-diameter screen will be more effective. However, this type of foam windscreen is typically unsuitable for directional mics. Note that camcorders with directional mics are often supplied with a simple foam windscreen; these are certainly far from being good, and the aftermarket supplies many different solutions (Fig. 4-3b).

For all directional mics, the best idea is to encapsulate a volume of air inside the windscreen, where the velocity of the air may be minimized. Such ball or zeppelin-style windscreens are often made of silk to form an enclosed place to protect the mic. Larger-diameter ones may produce better results, with consequent increases to size, weight, and price.

In more extreme cases outdoors, the silk windscreen can be covered with a Windjammer, a furry covering that progressively reduces the air velocity as it encounters denser fur toward the inside of the device, slowing down the air velocity with which the internal windscreen must deal.



Figure 4-3a Windscreen types. Shown are differing basket-type windscreens with an internal free air volume for best performance—such types are a virtual necessity for directional mics. A Windjammer, used to cover one of the basket-type windscreens for ultimate wind protection, is shown as well.



Figure 4-3b Aftermarket windscreen and added shock mount for the Canon XL series cameras offer greater reduction of camera and wind noise than the stock solution.

Pop Suppression

For close miking, particularly using a directional mic like a cardioid in a music studio miking a vocal, a stretched silk diaphragm, commonly called a Popper Stopper, placed between the vocalist and the mic can reduce or eliminate popping the mic on plosives.

For instances of coverage of people talking at podiums, a good idea is to use an omnidirectional microphone close to the talker, with a small foam windscreen. These will be unobtrusive compared to a full rig with a directional mic and a popper stopper. The reduced susceptibility of the omni to popping, and having no bass boost when used close, as a directional mic would, make up for the fact that this mic type is not directional. While you can't convince many so-called public address experts of this, who insist on using directional mics up close for feedback suppression, nonetheless it is true. One can simply work an omni mic closer than a directional one without problems.

Case Studies

Mounting Lavs

In mounting a lavalier microphone on a subject several issues need to be considered. First is the basic position: too low and it's too far away from the mouth, but too high and the microphone is in the acoustic shadow of the jaw. The optimum place is typically in the range of six to nine inches below the jaw line. It is desirable to place the lavalier somewhere around the centerline of the chest if the talker is facing the camera, because then head turns to one side or the other will take the sound off-mic. If it is a two-person interview situation, then the lavalier can be cheated offcenter toward the other person, so that, speaking straight ahead or toward the interviewee, the interviewer stays on-mic.

It is desirable to have the microphone see clear air and not be buried under dense clothing, and some extremely small types are available that can be pushed through a buttonhole and remain unseen. With a little larger type, it is desirable to place them under the minimum amount of clothing. A test is to blow air through the cloth; the easier it is, the more sound is transmitted.

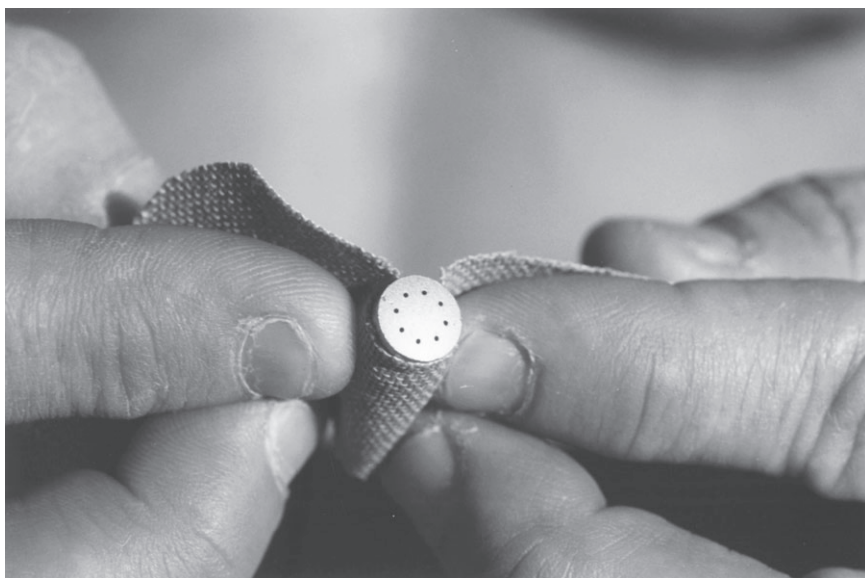


Figure 4-4a Wrapping a bandage around a lavalier provides secure mounting.

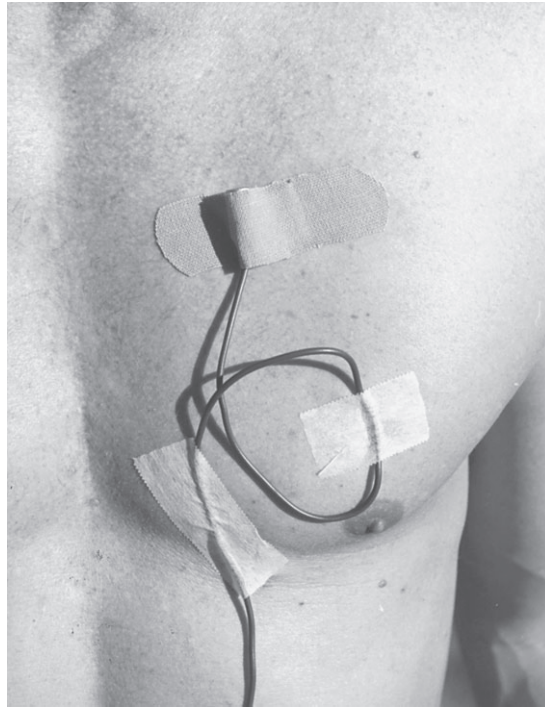


Figure 4-4b The mic is taped to the chest in the position shown (higher to the chin more shadows the sound; lower, the direct sound falls off). The loop in the wire is to provide flexibility so that the assembly is not strained as the actor moves.



Figure 4-4c Even under a T-shirt, the microphone is not very visible.



Figure 4-4d In cross light with poor cable dress, the assembly can become obvious.

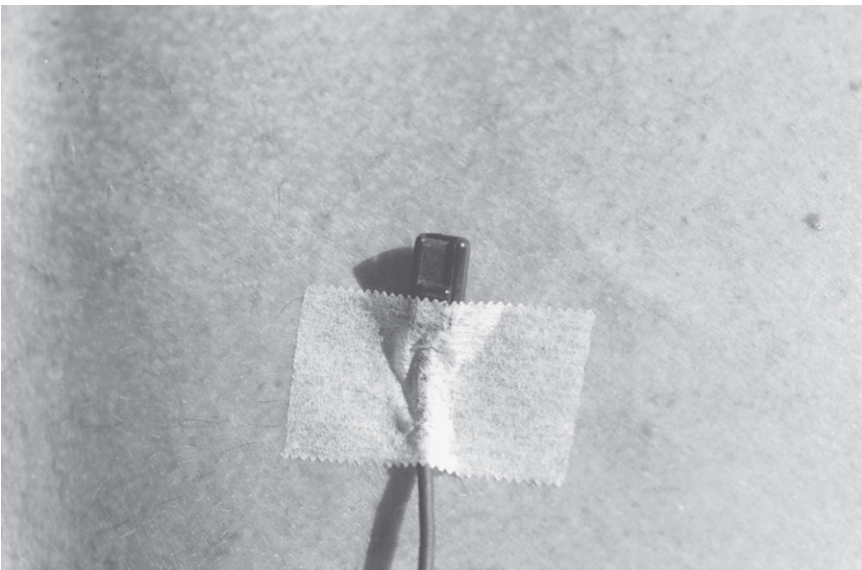


Figure 4-4e Side address microphones are somewhat easier to conceal, being flat.

There are probably several ways to tape the lavalier to the talker and to minimize the effect of clothing rustle. One such way is shown in Figure 4-4.

Radio Mic Usage

Several sections in this chapter have emphasized the pros and cons of lavalier microphones, how to mount them, and discussed considerations regarding the radio-frequency part of the system. All that background having been said, typically the biggest problems most often heard in film dialogue are caused by problems with radio mics. So Table 4-2 is a guide to solving the audible problems.

Disposable Mics

Occasionally you see a shot in a movie where it just seems that the camera must be destroyed as some large object attacks it (and the audience!). These shots may be made with a sacrifice camera, a cheap one that may be destroyed in getting a unique shot. There is a need for being able to plant a microphone in an animal's lair, for instance, to get the super close-up sound that is going to convince us of the reality of a shot, and yet potentially sacrifice the microphone. With the microphones discussed so far, this would be an expensive proposition. For this purpose, small electret capsules are available for just several dollars that can be fashioned into a complete microphone with a battery, a resistor, and a capacitor, as shown on their data sheets. These can even be used successfully underwater, wrapped in a condom and well sealed.

Production Sound Example

It is very useful for those studying how production works to be able to hear what it sounds like at various stages. Finished films only tell so much because we don't know what scenes have been ADRed if it's done well, for instance. Some studio Blu-ray and DVD releases give us a chance to hear what things sound like after picture editing has been performed, but before sound postproduction has taken over, in their deleted scenes. It is

Table 4-2 Radio mic problems.

Audible Problem	Possible Explanation/Solution
<i>Problems originating in the audio domain</i>	
Sound is muffled, indistinct	• Lav placed underneath too much clothing; try thinning out the clothing over the microphone. A test is that if you can blow air through it easily, the clothing is acoustically transparent, and the harder you have to blow air through, the harder it is for high frequencies to get through. If visibility of the mic is the problem, try using one of the smallest types, such as the Countryman B6, which is small enough that it can be stuck through a buttonhole and not be seen in an ordinary shot, and it comes in a variety of colors. • The lav may be in a poor position, such as at the top of a turtleneck pullover, where the sound is shadowed by the chin; needs equalization in addition to the high-frequency boost equalization supplied by most lavs. • If the sound track has already been recorded, then equalization can be applied, which can range from improving the sound somewhat all the way up to making the sound good.
Sound is chesty	Virtually all radio mic recordings benefit from the equalization described in Chapter 10.
Sound is distorted when actor shouts	Gain and limiter settings have to be right for the loudest section of the performance. If the radio mic channel is overloaded by shouts, then at the least limiting will take place, putting a cap on the maximum level but potentially producing distortion of its own.
Sound has clothing rubbing noise	Use 3M Surgical Clear double-stick hairpiece tape* on skin and around the mic, as per Fig. 4-4. Stick it to skin and clothing. Stick multiple layers of clothing together.
<i>Problems originating in the radio domain (further treatment of this topic is in Chapter 5)</i>	
Sound has intermittent electronic swishy noise or dropouts	Multipath is the cause. Locate and/or orient receiver antenna for better reception. Try vertical versus horizontal orientation. Use a Yagi receiving antenna cut for the frequency of the transmitter.** (Boom operator can operate antenna in scenes where only radio mics are in use.)

Continues

Table 4-2 *Continued*

Audible Problem	Possible Explanation/Solution
You hear signals other than the microphone, such as tones, speech, clicks, or other noises	Radio-frequency interference is the cause. With a frequency-agile system, try a new frequency. With a fixed-frequency system, sometimes moving the receiver to be closer to the transmitter can overwhelm the interference. Remember, you need not necessarily put the receiver where the camera is; if the camera is fixed, a long mic cable can be run from the receiver, which can be close to the transmitter.

*Or moleskin available from drugstores.
**Books are available on how to do this at www.arrl.org.

commonplace to shoot more scenes than are finally needed and to remove some after screening for the filmmakers or studio. However, because these scenes are removed from the film before the meticulous sound postproduction is done, they provide us with a means to hear the sound at essentially the dailies stage of the process. A particularly good example is the set of deleted scenes on the DVD of *To Wong Foo, Thanks for Everything! Julie Newmar*, a 1995 Universal theatrical release recorded by production sound mixer Michael Barosky and boom operator Linda Murphy. The scenes contain only production sound as it came off the set or location and edited by the picture editors. There is a mix of several types of scenes, including both interiors and exteriors. In some cases, you can compare the production sound with the finished track by listening to the main feature in contrast to the Bonus Feature presentation (Table 4-3).

Listening to these scenes, and the corresponding ones in the final film for those that are present, shows just what production sound can and cannot do when enhanced with postproduction techniques. Most of the scenes work, with equalization for lavalieres, added ambience and in some cases music, and occasionally added reverberation. On the other hand, there are scenes that are just about impossible to cover with production sound alone, because the background noise level is just too high. Take the car lot exterior. It's real, but it needs a lot of work in sound postproduction. In older Hollywood productions, this would have been shot on a controlled back lot. In today's world, no one has the time and budget for such meticulous staging, so it is shot on location, with the sound consequences heard in the bonus features segment.

Table 4-3 An analysis of production sound.

Bonus Features Scene Title	Notes
Extended Restaurant Scene	Interior. Boom mic Schoeps Mk41. Two booms in use on some wide shots, requiring two operators on final scene that is in the film, due to Robin Williams' ad libs, among other things. (Chapter 4.) Hiss due to multi-generation analog postproduction dubbing, not representative of the original.
Style versus Substance	Exterior, day, Queens, New York City. Radio lavalier mics made necessary by the very noisy location. In New York, if police hold traffic, then drivers honk their horns! The first line "This will make it to California, no problem" muffled due to radio mic placement (fixable with equalization in postproduction). This scene probably required at least some ADR due mostly to the noisy location. There are ticks and clicks, but their source is unknown.
The First Step to Queenliness	Exterior, day, quiet mountain setting. Lavalier radio mics on the three planted in their ample fake bosoms. Wide opening of shot and wind noise prevent use of boom. With equalization and ambience, this would be fine.
Chi Chi's Charms	Interior. Room sound of small room. Hiss probably from post dubbing. Note level of footsteps on stairs is high relative to voice.
A Day with the Girls	Muffled radio mic sound due to placement, fixable with equalization. Excessive footsteps on gravel around the line "Miss Noxie..." This was a preproduction discussion, with the problem pointed out by the location manager and key grip in scouting (sound didn't scout this remote location, but others were looking out for it). The road was raked, and rubber mats put down out of sight, but if the scene evolved to where they would show, they couldn't be used. Some of the footsteps heard are those of the camera operator running the Steadicam, not just the actors.

Continues

Table 4-3 *Continued*

Bonus Features Scene Title	Notes
Salade Nicoise	Interior, kitchen, day. Stockard Channing on her knees scrubbing the floor. Boomed. She sounds reverberant, whereas Patrick Swayze is closer on a second boom. Might need ADR. As scene evolves and Stockard crosses to the sink, it gets better.
Beatrice's "Red and Wild"	Ext. evolving to interior, day. Lavs sound a little dead. Need equalization and ambience, and possibly some added reverberation in the barn. These are good raw tracks for postproduction.
A Woman's Place	Exterior, dusk. Two guys in a wide shot, radio miked. A one-er (scene shot in one master shot). If close-ups were shot, their sound on a boom could be used over the master, by syncing up editorially. "Career" is distorted, but could possibly be fixed with pencil tool in Pro Tools. This may be a multi-generation dub problem and not be present in the original.
Taking a Chance	Interior, day. Good scene shot all on boom. One floor creak from camera move, possibly editable. Production sound crew helped by the quiet location, the shot choice, the placement of the generator, and so forth.
Extended Goodbye Scene	Footsteps cover the line "you take care honey." First shot complicated by camera moves. Radio mics when used need equalization. Crowd wide shot as car exits shows why you should always shoot sync, as it is useful. Some wind sound and distortion on radio mics of three of them leaving.

Boom Operator's Job

The boom operator has an important job. While not as technical as the production sound mixer/recordist, it nonetheless contributes greatly to the track. Among the considerations for the job are the following:

- Hold up the fishpole for extended periods of time and move it silently.
- Learn the dialogue for fiction films as well as anyone on the set, so that you can emphasize the right thing moment by moment.

For documentary films, stay on the ball to cover the person speaking and anticipate the next one to speak.

- Match the microphone perspective to the camera shot. This means knowing the focal length of the lens in use, or, if a zoom, where it is set; knowing where the frame line is and avoiding it; and keeping the distance to the actor appropriate for the acoustics of the space for the shot.
- Be the front-line person for the sound crew, because they are closest to the actors. They will often be involved in positioning radio mics.
- Do not distract the actors/subjects by making eye contact.
- Learn the translation from headphone listening to loudspeaker listening. Headphones exaggerate background noise, making almost any intrusive noise audible, even if it may not be much trouble to reduce in postproduction. There is a long learning curve with this skill. Unfortunately you cannot trust most editorial setups to tell you whether there is good or bad sound. It doesn't emerge until later in the process, with wide frequency range calibrated level monitoring, and most boom operators are long gone from the show by this time, so a valuable feedback loop is lost.

Video has one problem for camerapersons that film doesn't have that affects boom operators. The viewfinder shows more than the photographed area on a film camera, but exactly the captured area (if that—there could be overscan in the monitor) for video. This means there is no warning before the boom enters a shot in video. One possibility is for the fishpole to have mounted to it a small LCD monitor, showing the frame. This involves one cable from the camera's composite output² to the monitor and clamps to hold it to the fishpole. In this way the boom operator can be certain of the frame line.

Common Problems

Some gotchas are as follows:

- High-end cameras often combine channels in their monitoring circuits. Thus you may send a camera two channels and expect two

² Usually a yellow phono jack.

discrete channels back for monitoring with an external mixer, but instead get a return of the two channels mixed. Interestingly, this is not a problem with some semi-pro models.

- Level setting can be complicated by the use of different types of meters in use on mixers versus cameras. Moreover, professional cameras will indicate -20 dBFS as the reference level for tone on a digital peak meter, whereas prosumer cameras may show -12 as reference level (this is described in detail in Chapter 3).
- Cameras with nonlocked audio and video clocks, including almost all Mini DV cameras, may produce no sync problem played from tape, or even from the first-stage editing system, but, depending on how subsequent processes handle the audio-video sync, could come apart later. Systems that interleave picture and sound bits will normally be okay, whereas ones that separate the files and then try to put them back together later may suffer from sync problems. One way to ensure sync is to test the full path you are going to use in advance, from camera capture, through picture editorial, to sound editorial and layback to the picture master, and even to the final output, such as a DVD or Blu-ray disc. If you make a tape with a clapperboard slate every minute for the length of your program, then put it through all of the processes you are going to use, you can be assured that the final result will stay in sync.

Logging

In conventional film production, logs were quite important to keep because picture and sound were separate and had to be sunk up in post, and it was a waste to print takes that were no good. In single-system digital video production this is less of a strict requirement, but with many hours of footage, it is a good idea to know where scenes are located, how many takes of each are on a card or disk, and so on. Thus a camera assistant, producer, or script person can keep track of the sound file number, camera roll number, and general time code for specific scenes.

Sound Kit Accessories

- Location sound logs.
- Tape for lavalier mounting: 3M Surgical double-stick, hairpiece tape (from costumers), or moleskin.
- Tape, Gaffer's, cable tie down.
- Bungee cord, for holding cables.
- Carabiners, mountain climbing clamps for holding cables.
- Spare mechanical parts for boom and mic suspensions.
- Greenie, a small Xcelite screwdriver (model P3321) used for XLR connectors, small adjustments, etc.
- Tweaker, an adjustment screwdriver tool especially used for 10-turn potentiometers.
- Needle-nose pliers, small.
- Diagonal cutter, small.
- Soldering iron, small, and solder.
- First-aid kit.

Director's Cut

- On all shows, make the crew aware of the need to get good production sound.
- Obtain coverage. That is, shoot production sound of the event well, but fill in by recording room tone, and in fiction filmmaking record wild lines if possible for any dialogue that can't be covered well while the camera is rolling.
- Boom microphones are typically greatly preferred. Radio mics are possible but have several drawbacks.

Director's Cut *(Continued)*

- A few things can be done with an on-camera mic: recording ambience or certain production effects, for example.
- Use the two channels on most cameras and recorders for redundancy, staggered levels so you get the widest clean dynamic range, or to record two differing mic perspectives.
- Use shock mounts, windscreens, and fishpoles effectively to produce silent microphone moves.
- The boom operator is much more than “a big kid out of high school with strong arms,” as one producer exclaimed to me. Professional production sound mixers know that skilled boom operators make their job infinitely simpler by learning the scene, tracking the action, and so on.

CHAPTER 5

Production Sound II: Microphones

There are many different types of microphones in the world today, from sophisticated and expensive ones used for high-end film production and music to millions of simple ones used in telephones and answering machines. While the range is large, only a few types are prominent in film and video production, and this chapter will concentrate on these types.

There are two main factors that characterize microphones, beyond their simple visible characteristics of size, mounting, and so forth. The first factor is the method by which the microphone changes acoustical energy, sound, into electrical energy, represented by changing voltage versus time at the electrical output connector of the microphone. The generic name for a microphone is *transducer*, meaning a device that changes energy from one domain to another, in this case acoustical energy into electrical energy. A loudspeaker is also a transducer, at the other end of the audio chain, converting electrical energy back into sound.

The second factor that is important to users is the directional characteristic of the microphone: whether it accepts sound equally from all angles (practically speaking), or has a directional preference, picking up sound from one range of directions while discriminating against sound from other directions. This directional property, called a microphone's *polar pattern*, is in many cases the most important property in choosing a microphone.

Many other factors can be important in microphone selection, but these two factors are usually the first to be thought about. Other factors include the susceptibility of particular microphone models to wind and handling generated noise, how well particular models of microphones cover the frequency range from deep bass to high treble without discrimination, how well the polar pattern is maintained across frequency, power requirements, environmental susceptibility to temperature and humidity, ability to handle loud and soft sounds, ruggedness, and more.

The main method of transduction used in microphones for film and television production today is called *electrostatic*, and microphones built on this principle are commonly referred to as capacitor or condenser microphones. In this method of transduction, a stretched diaphragm of flexible material like a drumhead is moved in and out by impinging sound energy, usually by an incredibly small distance. Typically, the moving diaphragm hovers over a fixed internal back plate. The changing distance between the diaphragm and the back plate is analyzed by measuring the changing electrical capacitance between them, using one of several means.¹ There are several things for users to know well:

- All electrostatic (condenser, capacitor) microphones require a power source somewhere because of their basic principle of operation.
- All electrostatic microphones should be kept dry, because moisture can short circuit the insulation on which the ability to measure the capacitance between the diaphragm and back plate depends and cause crackling in the output or even gross failure.

Power

Power is supplied to electrostatic microphones in several ways, depending on the model of microphone in use. Microphones physically built into the body of a camera will have their power supplied as part of the general power supply of the camera, but those mounted externally fall into two categories: those supplied with the camera and those separately marketed as accessories to cameras. For those sold with the camera, normally the camera will supply the power; but for microphones treated as accessories the camera may supply the power, or one of the methods described as follows may be used.

The most direct way of supplying power to an external microphone is by a battery in the microphone body, or (often with lavalier microphones) in the shell of the connector supplied by the microphone manufacturer that

¹ *Capacitance* is a measure of the ability of a device to hold an electrical charge. In this case, an insulator, such as air, separates two conductors: the diaphragm, which may be metal or plastic coated with metal, and a metal back plate.

resides at the other end of a cable from the microphone. The type of battery used varies widely, and it is a necessity to have spares available, especially when working away from sources of supplies of particular and sometimes-rare battery types. Also, it is easy to overlook such batteries at the end of the day and leave them running microphones overnight. While microphones can operate a long time on some of the battery types, they may be found to be fully discharged just when they are needed if left on while not in use. A related way of powering electrostatic microphones is by use of a small separate box that houses the battery; the microphone plugs into the box, and the output of the box then plugs into another cable destined for the microphone input of a camera or mixer.

Batteries lose voltage as they discharge, slowly at first and then faster toward the end of their life. The consequence of diminishing battery voltage at the end of battery life is potential distortion, and perhaps less overall output level from the microphone. Distortion usually affects loud sounds first, and, as the battery voltage goes down, the effect will be heard with lower and lower levels of sound, until the battery dies completely. It is a good idea to time batteries in use and replace them before distortion becomes potentially audible. A curiosity that is important to know is that battery life may range from 20 hours for one particular model of stereo microphone using an AA battery (Sony ECM-MS5) to 5000 hours for a lavalier (Sony ECM-55B) *using the same battery*. Thus it is important to know the rated battery life of any particular microphone you use.

A second method of powering is by a direct power connection, separate from the wires carrying the audio signal. The microphone supplied with the Canon XL series, for instance, uses a connector with two adjacent plugs: one for the signals from the microphone, segmented into three parts called tip, ring, and sleeve for left signal, right signal, and ground, respectively, and one for power, segmented into two parts: tip for power and sleeve for ground. The camera supplies power to the microphone by way of this second connection, and because this is a somewhat unusual method of powering, the microphone as delivered is only suitable for use with these model cameras. Despite the relatively unique connector layout, powering an external microphone from the camera in this manner is simpler for the user because he or she does not have to remember a separate battery.

A third method of powering called *phantom powering* is used with a wider range of professional microphones. *Phantom powered* microphones are

often named with a “P” suffix in the microphone model number to indicate their method of powering. In this method, power is supplied on the audio signal wires directly in a particular way. The signal runs from the microphone to the camera or external mixer, while the power runs superimposed on it, from the camera or mixer to the microphone. As with the previous method, this is fairly simple for the user in practice, because it means there is no additional battery to remember. However, the input to which the microphone is connected must supply power, and it is often switchable on and off, because applying phantom power to other microphone types might result in damage. Also, the voltage of phantom supplies is not standardized; while 48 V phantom is the most common, 12 V phantom supplies are also found. So to use a microphone with phantom power, the user must switch on the power and set the voltage of the camera or mixer to match the microphone requirements. Forty-eight-volt phantom power is the method used by the Sony PD-150 and 170 camera microphone accessories, for instance.

A similar but not interchangeable method is called “T” or “A-B” powering; it is considerably rarer than phantom powering, but its widest use has been in production sound for film, so you may encounter it. Usually designated by a T suffix in the microphone type, it is an older method used particularly on some models of Schoeps and Sennheiser microphones, but is rarely if ever encountered with newer digital video equipment. Still, external mixers such as the Shure FP-33 and Sound Devices 422 are internally switchable between T and P powering, and the user must match the switch setting in the mixer to the microphone type in use for it to work. Just to confuse things further, T-powered mics were traditionally supplied in two polarities. Normally they had pin 2 of the XLR connector for positive (+) voltage and pin 3 for negative (–) voltage. On the other hand, those with a red dot and some others, perhaps unmarked, were designed for connection to Nagra tape recorders, the film production standard for many years, which had pin 2 (–) and pin 3 (+). If you encounter a microphone with a red dot, you may use a polarity reversing (also called polarity inverting) cable, or build one, to interchange the leads for pins 2 and 3 to connect such microphones to conventional T-powered inputs. The Shure FP-33 and Sound Devices 422 mixers both supply phantom power at 12 V and 48 V and T power at 12 V in the standard polarity, but not the reversed “red dot” polarity, so if a microphone formerly used for film sound is connected to a portable field mixer such as the FP-33 or 422, it may not work, and polarity is usually the reason.

Dynamic Microphones

Another method of transduction is the *electrodynamic* principle, with microphones built in this way called simply dynamic microphones. These behave like a conventional loudspeaker, only in reverse. Sound moves a diaphragm attached to a voice coil, which resides in the field of a magnet, causing an electrical voltage to form directly, using the same principle as an electric generator. These microphones require no power, and models can be built that are quite rugged. Their output voltage for a given level of sound is usually lower than that of electrostatic mics, and their uniformity of responding to all frequencies from bass to treble is often not as good as for electrostatic types. However, there are several places where they excel and are virtually a necessity: news operations where they can better stand the effects of weather and abuse, and as a backup for electrostatic microphones, especially in harsh shooting conditions. When all goes wrong with the insulation on an electrostatic microphone because of high humidity, battery supplies, and so forth, a rugged dynamic microphone will virtually always still work. Thus documentary filmmakers will carry a dynamic microphone with them to be sure that they can get sound under practically any conditions encountered.

Generally, connecting most dynamic mics to phantom power will not damage them, but T power may do so because in T powering the voltage is applied across the voice coil rather than between signal and ground. Moreover, connecting a ribbon microphone—another type of microphone transducer using the same principle as the dynamic mic for transduction but a different internal construction—to either phantom or T power is likely to damage it. Thus the need for microphone power switches on cameras with powered microphone inputs is apparent. As mentioned previously, portable field mixers intended for more general purposes may have switching for both phantom and T power, potentially of different voltages. Of course, if these are left off, electrostatic microphones of the correct type, P or T, will not work; if they are switched on and fed to a microphone that doesn't match their type, the result could be microphone damage.

Polar Patterns

The term *polar pattern* describes how a microphone responds to sound from various directions. The simplest polar pattern is omnidirectional.

The smallest microphones, such as lavalieres or microphones built into telephones, and larger microphones that can be identified by having only one entrance for sound, are typically omnidirectional. They accept sound equally from all directions, except perhaps at the highest frequencies, where the size of the microphone makes it discriminate somewhat for sources in front rather than behind the mic. Thus they generally need not be aimed precisely, but, on the other hand, they discriminate the least against undesired sound.

Where microphones have to be small, an omnidirectional polar pattern is usually the only option. This is true of almost all lavalier microphones, which gain their utility by being located close to the source rather than by discriminating against sound from certain directions. We'll examine how these are related below.

All other types of microphone polar patterns show a directional preference. While none of them can be said to magically reach out and record something at a distance, they do attenuate sound off their main pickup axis to various degrees, depending on their polar pattern, while maintaining their sensitivity to sound along the principal axis of pickup. In real rooms and locations, off-axis attenuation has the effect of reducing environmental noise and reverberation. This typically leads to a preference for directional microphones over omnidirectional types for microphones that are on cameras, booms, or planted on set because omni mics are at a disadvantage when it comes to discriminating against noise compared to more directional types.

Any directional microphone can be aimed such that a desired sound source is put on-mic—that is, the microphone is pointed toward the source—and undesired sound is placed in the null or cold side of the polar pattern. While almost everybody understands intuitively that a microphone is better aimed at a source than not, almost everyone also overlooks the second fact: that aiming the side of the mic that discriminates against undesired sound is perhaps equally valuable. Shooting a scene outside a nightclub for a documentary about music on noisy Sunset Boulevard in Los Angeles on Saturday night means getting the most directional mic the closest possible to the subjects, aiming it at them, and keeping the back side of the mic to the street, thus reducing the noise of the street cruisers as much as is practical.

Another type of noise insofar as production sound recordings are usually concerned is reverberation. Reverberation is sound that has bounced

around a space several times until individual reflections have merged into a foggy cloud of sound; reverberation arrives at the microphone distinctly later than the direct sound and comes from virtually all directions. Reverberation can be helpful in small amounts to set the feeling of distance, because recordings with no reverberation at all (overly “dry” recordings) can sound as if they come from outside the space of the movie, like narration. However, in larger amounts, reverberation can be damaging to speech intelligibility because each sound emitted produces reverberation that tends to cover up later sounds.

In contrast with reverberation are direct sound and early reflections, the other two sound fields present on set. Direct sound is the sound that originates at the source and propagates in a straight line to the microphone, assuming nothing is in the way. Early reflections are the reflections of direct sound off walls, the floor and ceiling, and so on that work like balls on a pool table, bumping off the edges in a defined way and eventually making it to the microphone. Each of these sound fields has distinct directions associated with it, and thus directional microphones can be used to emphasize or deemphasize one over the other, depending on where the microphone is placed and in which direction it is pointed.

A demonstration of appropriate reverberation on a voice in an interior compared to voice-over narration occurs in the first scene of *Apocalypse Now*. The first voice is that of Capt. Willard, in voice-over, as we hear his interior thoughts. This voice recording is on-mic and dry. Then, when the officers arrive to pick him up, they are heard in the hall of the Saigon hotel. Their voices, doubtlessly from production sound, are more reverberant: they are real.

While used appropriately in *Apocalypse Now*, excessive noise and reverberation on voice is one of the biggest problems exhibited by many movies, whether shot on video or film. This problem arises more and more often today because of the routine use of location sets, particularly in sparsely furnished modern housing. Kitchen scenes can be particularly bad, because all of the surfaces in a kitchen are hard, and thus the room is quite reverberant. This problem is also demonstrated by the large, hard interiors of Savannah, Georgia mansions in *Midnight in the Garden of Good and Evil*. Many lower-budget films suffer even more from the effects of excessive reverberation, to the point where the dialogue becomes difficult to understand. Basically, it is impossible to remove reverberation in postproduction, and relatively easy to add it, so production sound concentrates on

picking up the voice without excessive reverberation, because it can always be added but never subtracted after the fact.

There are basically two ways around picking up excessive noise and reverberation: get the mic close while aiming it properly and make it directional to reject off-axis reverberation and noise sources. The case of the lavalier is covered later; for boom and planted mic operation, we typically want the most directional mic we can use without side effects dominating.

The types of polar patterns that can be found on film and video microphones are as follows (Fig. 5-1):

- *Omnidirectional*. This type accepts sound equally from all directions. As such, it must typically be closer to sources in order to pick up an adequate amount of direct sound compared to reverberation. It is the simplest construction, and thus is found in the smallest microphones, such as lavalieres that are placed close to the source of sound.
- *Bidirectional*. This type prefers sound from the front and back but nulls out sound from the side. Its main use in film and video sound is in making the side-facing component of single-point stereo microphones. Random sound energy like reverberation originating from all angles is attenuated to a factor of one-third of the level of on-axis sound.
- *Cardioid*. This type has a broad frontal acceptance and a null at 180 degrees from its principal direction. While this polar pattern is probably the most often found among directional mics in the world, film and video boom mics are usually more directional, such as the following types. Like a bidirectional mic, random sound is picked up at about one-third the level of on-axis sound. There is also a subcardioid available in some types, with its polar pattern between that of an omnidirectional mic and a cardioid.
- *Hypercardioid/supercardioid*. These types have a little narrower frontal acceptance than a cardioid, being somewhat more attenuated at 90 degrees than a cardioid, and they have a null in the shape of the surface of a cone, at an angle from the front of 110 degrees or 126 degrees. There is a small rear lobe of sensitivity. These are the most directional types of simple microphones, with random sound being picked up at a level about one-quarter that of on-axis sound.

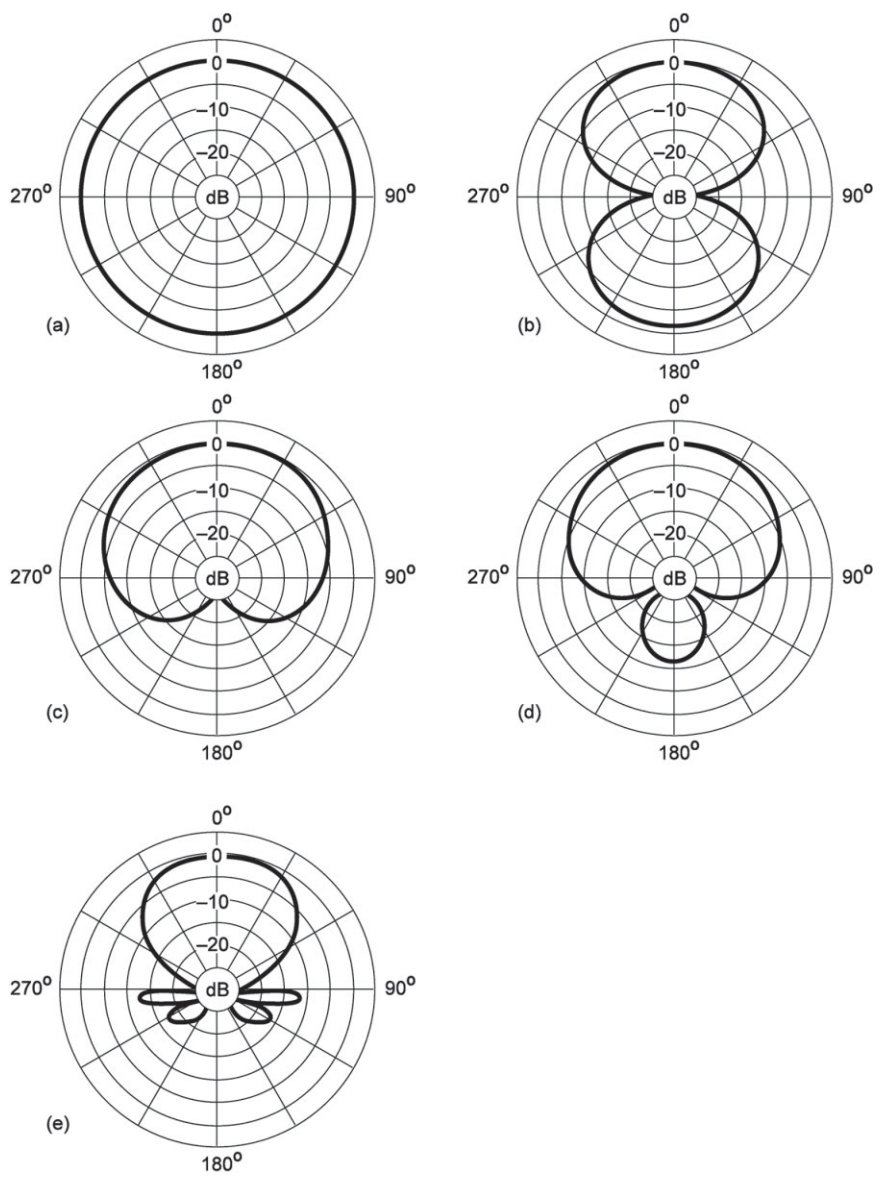


Figure 5-1 Microphone polar patterns. (a) Omnidirectional. (b) Bidirectional. (c) Cardioid. (d) Hypercardioid. (e) Interference tube. Although these are shown in two dimensions, the actual polar pattern is the two-dimensional pattern rotated through three dimensions along the main pickup axis of the microphone.

Because the null does not coincide with the rear axis, these types can be put on a boom and aimed at the actor while the null is aimed at the camera without being in the field of view of the camera.

- *Shotgun.* These types vary in their polar pattern depending on their length, from a fairly narrow club shape (about ± 30 degrees about the axis) for short shotgun models (meaning about a foot in length) to very narrow (about ± 10 degrees) for those around three feet in length. Because their angle of acceptance is so narrow, it is critical that they be aimed correctly. Sounds from off the axis are colored because of the principle of operation. The short shotguns only develop their directional property at higher frequencies, so they do not discriminate any better than a hypercardioid/supercardioid at bass to midrange frequencies. In practice, they *must* be aimed if the subject is allowed to move at all, so they require a boom operator in most instances. With these potential drawbacks in mind, they are probably the most used film and video production microphones, followed by hyper- or supercardioid types.

Long shotguns are really only suitable for outdoor use unless they are used in a large, dead studio like a Hollywood shooting stage, because through their principle of operation, their off-axis sound—which constitutes a large fraction of the energy received at the microphone in a normal room—is so colored that the problem becomes obvious. On the other hand, they can be useful in outdoor tracking dolly shots, to give just one example. For such a shot, lavalieres have a peculiarly intimate quality that does not match the picture well, whereas a long shotgun mic on a boom may get perfectly usable sound.

Differences Between Microphones Having Various Types of Polar Patterns

All other things being equal, an omnidirectional mic will be less susceptible to wind and possibly to handling noise than nondirectional types. In fact, most omnidirectional mics are less wind susceptible than most directional mics. For building a microphone system with maximum wind noise rejection, see pg. 164.

A corresponding fact is that omnidirectional mics have less sensitivity to pop effects. Try this experiment: hold your bare arm up vertically with

your hand about six inches in front of your face. Pronounce a “P.” You should feel a puff of wind in your hand. Now pronounce a “T.” You may now feel a puff of wind lower than you did the “P.” Ps and Ts are called *plosives* in speech. They create these puffs of wind when pronounced, and how microphones respond to them distinguishes how useful they are on close-miked speech. Omnidirectional microphones have an advantage in close miking because they are less susceptible to pops.

Directional microphones have another potential problem in that they exhibit boosted bass when the source is close to the microphone, an effect called the *proximity effect*. This can exaggerate a woofy quality in voice when used close up, for instance.

Here is advice for selecting a directional pattern in short form:

Use omnis:

- Where the source-to-microphone distance is small, so that the fact that the microphone is nondirectional does not cause a severe disadvantage.
- Where the source is a talker and the microphone can be placed very close—right in front of the talker—for two reasons: this type demonstrates less susceptibility to pops and exhibits no bass boost at close range.
- Where wind is a significant problem.
- Where you must because of the requirement for small size, such as lavalier use and for planting hidden microphones where the size must be kept small.
- Using multiple spaced omnis is one method of stereophonic recording that may be particularly useful with certain large-scale effects (e.g., a train going past the microphones).

Use cardioids:

- Where you can differentiate the sound field into desired sound coming from one direction and undesired from 180 degrees away

from the desired sound, such as placing a microphone near a talker on a street and pointing the mic at the talker and its back side at the street.

- However, note that cardioids, which in the world outside film are the most common directional microphones, are not as widely used in filmmaking as in music recording or public address systems, because hypercardioid and supercardioid mics are more directional and thus discriminate against noise and reverberation better than cardioids.

Use hypercardioids or supercardioids:

- Where you want to discriminate the most against reverberation in a mic having a relatively small body.
- Where the source of noise can be placed in a null in the range of 110 degrees to 126 degrees off the front axis, and the desired source on axis, such as when the mic is a boom capturing an actor on-mic and placing the camera in the null of the mic (typically more important with film cameras than video ones).
- These considerations lead to the selection of these types for boom microphone use.

Use short shotguns:

- For greater discrimination against reverberation and noise at high frequencies than hypercardioids or supercardioids, although in the bass and midrange they are equivalent to those types.
- Only when, if the actor or event is moving, a boom operator can accurately aim the mic, because off-axis sound is poor.
- This is probably the most often used boom microphone type.

Use long shotguns:

- For the greatest discrimination against off-axis sound over a wider frequency range than the short shotgun.

- Out of doors, for wide shots, dolly shots, and so forth.
- Indoor use is not generally recommended, because the interaction of the complex polar pattern of this mic type with room acoustics leads to coloration.

The following recommendations also apply to microphone selection, but not to selecting specific polar-pattern types. They apply instead to particular microphone techniques.

Use boom mics:

- Whenever practical, because they produce better timbre and perspective than lavalieres.

Use dead-hung mics:

- This is a type of inflexible boom mic that might be used in a fixed setting, such as the set of a sitcom; it is useful where it makes sense for coverage, such as upstage away from the reach of a conventional boom but in a well-defined position.

Use on-camera mics:

- For ambience or establishing shots with sync sound.
- For close-ups if there is no boom operator.
- As a last resort for a scratch track.

Use planted mics:

- In conventional car interiors, on the headliner, positioned so that the sound makes sense relative to the picture.
- When set pieces clog up the working space for a boom and the position of the actor is well defined, such as for Norm seated at the end of the bar in *Cheers*.

Use radio mics:

- When there is no practical way to use a boom mic and the problems with them have been considered.
- When a wired lavalier won't work because of the needs of actors/subjects to move about.

Problems with radio mics include:

- The sound perspective never matches the camera perspective. If an actor turns and faces away from the camera, the sound doesn't change, when it should. This isn't always important, because a news-reader for instance is not shot at a variety of angles, but for most fiction and documentary work it is important.
- Noise from clothing rustle and the direct interaction of clothes with the microphone body is troublesome.
- The microphone is in a poor position for picking up voice—it can't see the lips. Many lavalier microphones exaggerate high frequencies, effectively turning up the treble to try to overcome this effect, but the effects of putting a microphone on a chest instead of in front of a talker are to color the sound in a more complex way than can be compensated for with a simple treble rise. The sound is literally chesty and lacking in certain midrange to high frequencies, while emphasizing others. Methods of dealing with lavalier sound are described in Chapter 10 on mixing, but, all in all, lavalieres have to be said to be a compromise, varying in intensity from small to severe.
- For fiction filmmaking where the microphone cannot appear in frame, burying the mic beneath layers of costume tends to muffle the sound.
- For documentary filmmaking, putting lavalieres on subjects invades their personal space. For very small crews, the director is usually the one who mikes documentary subjects. Running the mic cable outside of clothing so that it shows on camera is bad form, even for news coverage, and so the cable must normally be run under clothing. Actors are very used to being handled by makeup, hair, and costume people, so they

usually don't find the addition of dealing with a radio mic to be much trouble, but for documentary subjects it may be a strange experience.

However, sometimes the only solution for a given shot is to use one or more radio mics. In a large space with a wide shot from a low angle, there is literally no place to put a microphone that won't be seen in the shot. In an exterior street scene in a busy city, there may be no way to conceal the actions of the crew, such as a boom operator, from bystanders without use of concealed radio mics, and thus other techniques draw more attention to the filmmaking than might be desirable.

The Radio Part of Radio Mics

The term *radio mics* used in film and video sound usually means lavalier microphones attached to belt-pack transmitters, often used with similar sized receivers placed directly on the camera or in a bag with the audio mixer (Fig. 5-2). However, there are other instances, such as handheld microphones with an internal transmitter or an external transmitter on a plug-in called a butt plug, and receivers that are rack mounted for fixed installations. You can pay between \$130 and \$7000 for a radio mic, so what distinguishes one mic from another? You might think that it would be the transmission power, because more power will reach farther, but that is unlikely to be the main difference because of strong Federal Communications Commission (FCC) regulatory limits on transmitter power.

The distinguishing features across the wide range of radio mics have to do principally with the reliability and sound quality of the transmission system. You can beat any of these systems with a hardwired microphone cable, but you would be giving up the freedom of movement that is so cherished among actors and subjects. Here are some qualities that give radio microphones such a wide price range:

- *You typically pay a premium for operating on radio transmission frequencies that are less likely to be occupied by others, such as television transmission, paging systems, industrial radios, government channels, and so on. Higher-frequency radio mic systems that work at UHF frequencies may be more costly than VHF because the technology is a little more complex, but you might be willing to pay the difference because the frequency spectrum is less busy at higher frequencies and thus*



a



b

Figure 5-2 A typical radio microphone (a) transmitter and (b) receiver.

less likely to have unanticipated sources of trouble. Even with sparser radio traffic at higher frequencies, interference remains a possibility at different frequencies in different locations. A competent seller in a particular market will know the frequency coordination issues of that market and sell or rent you a radio mic that isn't on the same frequency as a powerful UHF digital television transmitter, for instance. This is why local knowledge is important in the use of radio mics: to know what obstacles might be encountered in a particular market.

- *Fixed frequency versus agile frequency tuning of transmitter and receiver.* A fixed-frequency system is the simplest and may well work in your case, but in case interference develops on that channel, a frequency-agile system can be set to a new frequency. Some are even smart enough to scan across the frequency range and find an available channel—one without interference. Others offer a spectrum analyzer function, showing a graph of all the radio frequency transmissions occurring across a band of frequencies so that you can choose a frequency where no one else is broadcasting.
- *More expensive systems will typically work under more difficult conditions.* This is because the filters that separate the desired signals from undesired ones will be more elaborate in more expensive systems and do a better job of discrimination. If you were to use a single radio mic in a rural environment away from other sources of radio frequencies, then an inexpensive one might do, whereas if you needed 10 radio mics in a city near a large transmitter, you'd have a much bigger problem—perhaps an insurmountable one for a given budget.
- *The transmission method commonly employed for radio mics is frequency modulation (FM), and it suffers from the potential for multipath reception.* That is, the receiver may pick up the principal signal from the transmitter, but also the same signal reflected off a building or other large surface. That reflected signal could come in at just the right delay relative to the direct transmission so that the sum of the two signals adds up to nothing—we say they are out of phase with respect to each other. This is the effect you hear when listening to FM radio in a big city while driving around in a car. The momentary noisy dropouts are caused by multipath interference from reflections off the buildings. You can even find that if you drive up to a red light there may be a distorted signal or no signal at all, but when you move forward a foot or two, the signal is restored. This is a result of multipath interference.

The best way to minimize multipath problems is called *diversity reception*. Here two antennas and receivers physically spaced apart are scanned and the output is selectively chosen as the one having the best signal, moment-by-moment. The idea is that although the signal may drop out at one point in space because of multipath, it is much less likely to drop out at two points in space at the same time. Full diversity reception operates as described, with two full receiver systems used, in one or several chassis. Antenna diversity systems use a single receiver system and simply attempt to add signals from two antennas together to achieve a similar goal, but such efforts are rather limited in quality.

Another way to minimize multipath is to use directional receiving antennas tuned to the frequency range of the transmitter and aimed at the source transmitter, thus reducing the multipath signal strength. Such directional receiving antennas work rather like directional microphones, picking up signal along one axis and attenuating signals from off the axis. In some such cases, if a boom mic is not in use, the boom operator will instead aim the antenna for a radio mic system to follow the action. This method is sometimes used on feature film production, although it is comparatively uncommon.

- Because of the nature of noise on FM systems, analog transmitters and receivers use companding noise reduction, with the transmitter variably boosting the level before sending and the receiver cutting the level back in an equal and opposite way. The goal of such a system is to hide the noise of the transmission channel behind the program content. The companders used are of various levels of quality; this is a tricky process, varying in time, and can add artifacts to the voice, so a more expensive system is likely to be audibly better, although that might not always be true.
- *Audio performance varies from model to model.* Because the dynamic range of the RF channel is limited, it is commonplace to employ an audio limiter, capping the maximum loudness of sound. These limiters can range from benign to obvious in use, depending on their design and how hard they are pushed.
- *Recently, more expensive systems have gone to digital transmission instead of FM transmission,* with claimed benefits in reliability of the channel,

which is what radio mics are all about. In order to do this, the audio must be strongly coded to fit within the channel, and this coding may or may not be audibly transparent, so it is a good idea to audition these systems before you need them.

- *The most elaborate digital systems employ encryption to prevent someone from tapping the transmission.* With high-profile movies this may be a consideration, but it probably doesn't affect most producers.
- *So far, all known radio mic systems use FM or digital transmission assigned to one particular frequency* (that is, for the carrier; the audio-modulated radio frequency signals extend out around the carrier). In the future, spread spectrum systems may be expected to come on the market. Spread spectrum means putting out radio frequency energy over a very wide band of frequencies with a particular method of coding such that the transmitter and receiver are synchronized to one another—there is no single carrier frequency, and this is what keeps the system more immune to multipath and less likely to be detected by an interloper. In a bizarre turn of history, the co-inventor of this technique was the film star Hedy Lamarr. Kept as a government secret for many years, Lamarr's method of communications was to use something very like player piano rolls synchronized in both transmitter and receiver to determine their carrier frequency moment by moment and to thus frequency-hop the carrier so the transmission could be hidden. The largest use of a related technology dating back to Lamarr's World War II invention in the world today is the global positioning system (GPS).

It is important to start level optimizing for radio systems by setting the level at the transmitter rather than at the receiver. This is because the transmitter audio gain control best matches the range of the actor's voice to the available dynamic range of the channel. Overmodulation will run into limiting and possibly cause obvious distortion, while undermodulation will make for noisy recordings. Often two LEDs are provided, perhaps marked SIG for signal, indicating some noticeable activity in the channel, and OVLD for overload, indicating so much signal that distortion is risked. The idea is to set the manual transmitter level control so that the SIG LED flashes often but the OVLD flashes little. An actor or subject's entire performance must be taken into account, because overload of radio mic channels upon actors shouting is all too common. In fact, in many films

excessive limiting of the transmitters is often the most audibly identifiable part of there having been a lavalier in use, because of the artifacts surrounding the gain changes made by the limiter.

There are a great many radio mics on the market, and perhaps a simple system will do for your purposes, but it should probably be tested in the conditions of use to prove that. One way to tell what systems seem to work best is by consulting film/video sound-specific rental houses for large urban areas, such as www.coffeysound.com and www.locationsound.com in Los Angeles, or similar companies in your area.

Director's Cut

- Microphones have two main characteristics: how they convert sound to electrical energy and their directional properties. Electrostatic (condenser, capacitor) microphones require a power supply, so the need for special batteries is a consideration for some of them. They are typically more delicate than the other main type but generally have the best sound quality. Dynamic (electrodynamic) microphones do not require power and are generally more rugged.
- The only directional types usually used for boom mics are super- or hypercardioid, or shotgun, because these are the most directional types, and the idea behind using a boom mic is to reduce noise and reverberation and pick up the direct sound of a voice, separating it from the background as much as possible.
- In high wind, omnidirectional microphones perform better. Since most lavalieres are omni and are covered with costume cloth while in use, they can have an advantage in windy locations.
- The best use of radio mics requires professional advice pertaining to the location where they will be used.

CHAPTER 6

Production Sound III: Dealing with the Output of Microphones

Microphones are called, as a general class, *transducers*, changing acoustical energy, sound, into electrical energy, voltage. They typically put out much lower voltage levels than most audio gear, so they must be amplified by a microphone preamplifier to be brought up to usable levels. A typical amount of voltage delivered by a professional microphone is 13 mV (0.013 V), a rather small voltage, for 94 dB SPL,¹ a rather loud level. Thus microphones must be connected to microphone inputs as opposed to line inputs, and the converse is true, too: line-level devices such as CD players must be connected to line-level inputs and not mic inputs. If a line-level device is connected to a microphone input, the result will typically be overload or bad distortion. If a microphone is connected to a line-level input, and the gain is raised high enough to hear the signal at all, the result will typically be quite noisy or hissy. The mic/line situation is complicated by the microphone attenuate (MIC ATT) switch on some cameras that allows high-level microphone outputs (when encountering loud sound) to be accommodated by low-level (more sensitive) microphone inputs.

This chapter will give you the details needed to scale the voltage output of the microphone as needed to the input capability of the microphone preamplifier, whether in a camera or an external mixer. Before putting these steps into action, it is important to understand that over-recording digital audio is a very bad thing to do, as the onset of audible distortion is very quick when the maximum level is exceeded. On digital peak meters, the maximum recordable level without distortion is defined as 0 dBFS. Figure 6-1 shows a good recorded level for dialogue. The bar on this camera display indicates the current peak level, and the separated section indicates the recent peak level. Here, the signal is somewhat dominant in

¹ Sound Pressure Level, dB relative to the threshold of hearing, which is a pressure of 20 $\mu\text{N}/\text{m}^2$.



(a)



(b)

Figure 6-1 (a) The audio meters on most digital video cameras show left and right channel peak levels both currently (the bar) and in the recent past (the separated line sections giving “peak hold” functionality). This peak level is good for dialogue recording. (b) Levels for each channel are usually adjustable using audio gain controls on the side or back of the camera.

the left channel, which could indicate one of three things: (1) the sound source was off center to the left; (2) the signal was greater from the microphone plugged into the left channel input; or (3) the left channel gain was set higher than the right channel gain in using the audio level knobs on the camera. Offsetting the audio level as shown in Figure 6-1b is a common tactic when a single mic is being fed to both channels in order to provide a backup recording at a lower level in case the primary recording exceeds the available headroom on channel 1.² If your audio meter includes level markings, peak levels for normal dialogue should not exceed -12 dBFS, leaving 12 dB of headroom for unexpectedly loud sounds.

While in cinema at large the name *Cries and Whispers* normally refers to the Bergman movie of that name, in sound we can use the phrase to describe being able to record very loud sounds without distortion and very soft ones without excessive noise, at least of the electronic variety. Accomplishing this is a matter of matching the dynamic range of real life to that of the recorder.

Real life has the potential for wide dynamic range. While the span of the quietest sound you can hear to the loudest sound you can normally tolerate is about 120 dB or a ratio of one million to one, even perfect 16-bit digital recordings can produce only 93 dB range,³ and actual practical cameras set to 16-bit recording have much less range. Measuring a Panasonic AG-DVX100Ap through its microphone input to its analog audio output showed it to have 58 dB dynamic range, and from its microphone input to its FireWire (digital) output to have 69 dB range.⁴ The analog Nagra 4.2 recorder used on many older feature films, measured a similar way, has a dynamic range of 72 dB. So in practice, this DV camera

² In order for the second channel recording to remain clean, the levels must be offset at the proper place in the recording chain as described later in this chapter. In the case of a microphone plugged directly into the camera, the recorded level can be offset using the camera's audio gain controls, but if the microphone is plugged into a mixer or recorder as described below, it is better to offset the level at the first possible point in the chain: once a signal distorts, it cannot be saved by turning the level back down at a later point in the audio processing chain.

³ For aficionados, the reason this is not 96 dB is because of the necessity for dither to linearize the medium. Dither is some deliberately added low-level noise used to smooth the steps of the medium.

⁴ For experts, this was from the 1-kHz clip point to the weighted noise floor, weighting consisting of a 400-Hz high-pass and a 22-kHz low-pass filter. This seemed appropriate because the camera had a lot of low-frequency noise, which is usually acceptable because it is less audible than mid- to high-frequency noise.

is a little less good than the mono Nagra that it replaces when the move is made from film to DV as an original capture medium.

Recording Level Parallels to Early Cinematography

In the early days of the cinema industry, light meters did not exist. Cinematographers learned that with a certain amount of light on a subject, a certain f-stop, frame rate, and shutter angle would produce an exposure that contained a full tonal range from black to white: the image was said to be exposed correctly. Light meters were not introduced until the late 1930s and have been in use ever since.

Certainly one of the most famous still photographs of all time, *Moonrise over Hernandez* by Ansel Adams, was photographed without an exposure meter: there simply was no time to take a meter reading because of the fleeting weather conditions. What Adams knew from experience was the film speed, the f-stop, and the shutter speed for the full moon, and he exposed the negative for that.

Sound people today are like cinematographers before 1935, or like Ansel Adams in 1941, without meters and in such a hurry that we have to know what we're doing. We learn what sound levels are expected, and once we have a good feel for it, we can apply the methods of this section to help fit the real world into the recordable space—the dynamic range—on the medium. However, it does take time to learn what to expect. Table 6-1 gives some day-to-day sound levels so that you can calibrate your ear to know when you may be in trouble and have to resort to special measures. The table lists two different measurements: a type of average level called L_{Aeq} and the peak level of various sources.

The first thing to notice about the table is that the first and second columns of data are very different. Just watching television shows a difference of some 27 dB from the average level to the peak level, a huge variation.

The L_{Aeq} method of measuring sound pressure level involves first frequency weighting—that is, modifying the frequency response rather like tone controls do, but in a specific way that better matches human perception (rolling off the bass and extreme treble because we don't respond to them as strongly as to mid-frequency sound). This process loses some

Table 6-1 Sound levels encountered in various shooting situations.

Item	L_{Aeq}	dB SPL*
		Flat response, Peak detector, Impulse time response
Quiet room suitable for recording without extra precautions	20	68
Fairly quiet room but would require careful manipulation of background both in editing and mixing	28	62
Apple G5 measured at user's head, with computer on the floor below	36	60
Fairly quiet Hollywood street, peak represents car and truck passbys; suitable for shooting at least close-ups, especially if traffic could be controlled	48	65 (car by); 93 (truck by)
Quiet model dishwasher at 1 meter (m)	51	63
Television news listening at 3 m	55	82
Normal speech measured at typical close boom position at 0.5 m	65	86
Noisy Hollywood street: 7900 Sunset Blvd. at Director's Guild of America on a Saturday afternoon, peak level caused by city bus driveby across the street	65	105
Interior Los Angeles delicatessen at lunchtime	70	104
Normal speech measured at lavalier position	75	95
Urban street including bus passing observed from bench at side of street	82	107
Actor shouting at 0.5 m	na	128

*The reference level of 0 dB for SPL is a pressure of $20 \mu\text{N}/\text{m}^2$.

level relative to the peak because content is taken away. Second, a type of ac-to-dc converter called rms is used, progressively under-reading shorter and shorter events. This process also loses level relative to the short-term peak, because short time events slip through this relatively sluggish device. Finally, the *eq* part of the term means equivalent, that

is, a long-term running average of the thing being measured, with excursions in level both above and below the level measured. All three of these items cause the reading to be reduced compared to the instantaneous peak, and used together, they accumulate to a very large difference.

The “flat, peak, impulse” method of measurement means no frequency weighting is employed (flat frequency response, i.e., tone controls set to neutral), the detector used captures the peak (highest) value, and the impulse time constant of 35 milliseconds (ms) is used. While we can hear distortion in as little as 2 ms and thus must leave some headroom for instances even above the level stated, 35 ms is the fastest setting on many professional sound level meters.

Note that average levels are often more widely published, but the peak level is what causes distortion problems, and so must be observed for our purposes here. Also note that the difference between the two values is not fixed: it all depends on the situation. For instance, look at the first two items; both are rooms that we could record in, but we would certainly prefer the former. Although the first is some 6 dB noisier on the flat, fast measurement, its weighted average level is 8 dB quieter! The higher peak level wasn’t even audible because the increase in noise in the quieter room was apparently at very low frequencies, where hearing is not nearly as good as in the midrange.

This is the stuff that makes people throw up their hands and say that measurements are meaningless and you can’t quantify something like hearing, but that would be wrong. We can make great use of L_{Aeq} to check the background noise level of a space, say while scouting locations, whereas we must use the faster measurement (and even leave some room above it for shorter-term events) to determine how much input level the microphone and recording system will see on peaks. The peaks require special care in handling, because they can cause distortion if over-recorded. Having a professional sound level meter available is rare, but coming to know what background levels and peak levels are like, even experientially, is useful.

Fortunately, the wide potential dynamic range of real life is not encountered often while recording many kinds of programs. Take an interview situation: it probably won’t get too loud, although having a low background noise level is important because speech in a quiet setting leaves holes in the program material: silence on the part of the talker between the

syllables—or more exactly, the phonemes, which are the fundamental molecules of speech—that can reveal noise. The noise in the silences may be either the acoustical noise in the space or the hiss caused by the electronics involved in recording, and it is the latter which we can control using the methods discussed in this chapter. Acoustical noise has to be reduced by paying attention to background noise levels during selection of the site, and by utilizing the directionality, location, and aiming of the microphone.

Just the sound of an actor or documentary subject talking normally measures an average of 65 dB SPL at a conventional boom location for close-ups, but we must leave headroom for occasionally louder brief sounds, measured to be 86 dB⁵ peak.⁶ And getting to the instantaneous peak value so as to prevent any audible clipping probably takes this up to something over 90 dB. With a lavalier on the chest, the levels will be even higher. So will the camera remain undistorted at these levels? And how much room do we need to leave for an occasional over-the-top performance?

To know, we have to make a measurement. We set the Panasonic camera to its normal input sensitivity of -50 dB using the menu system. Then we connect a sine wave generator to the input and put in a single mid-range frequency tone, such as 500 Hz. Starting at a low level, we increase the level of the generator until the input abruptly distorts while we are listening to the monitor. (We have to be sure that nothing else in the chain is distorting first, by setting the camera's main recording level control correctly so that the signal stays in bounds, in this case meaning stays out of the red on the meter, and setting the monitor level low enough so that the monitor output is not distorting.)

When we do this, and measure the voltage where the input is just below the point of obviously distorting, we find a value of 48 mV (0.048 V). So we have to keep the microphone's maximum output to 48 mV or we will distort the camera's microphone input. It will make no difference if we

⁵ dB is used colloquially here and at many places in the text. Strictly speaking, the term dB only specifies a ratio, and without further delineation is unclear. The term SPL is implied anytime we are speaking of acoustical levels.

⁶ These numbers were measurements of the author Holman talking at 0.5 m, measured L_{Aeq} (for the 65 dB SPL measurement) and impulse peak (35 ms) on a B&K 2230 sound level meter.

turn the camera's level control down: *If the sound is distorted at the input, no amount of level setting of subsequent controls will fix the problem.* This fact eludes many people, especially ENG crews, as stand-up reporters' audio is often grossly distorted, and we can presume that the operators have done their job and set the recording level. What's happening is that the radio mics they are using have relatively high output, and they are overloading the microphone-level input of the cameras, so no level setting of the main level controls of the camera can fix the problem. In fact, one popular ENG radio mic receiver puts out a nominal -20 dBu, whereas mic inputs are intended for around -50 dBu!

So, for our purposes, let's choose a microphone, a good one. (This is meant to be illustrative, and by filling in the numbers from your particular example of microphone and camera combination, the method can be made universal.) Take the Schoeps MK41 hypercardioid capsule mounted on their Collette Series electronics, a CMC6U, and let's look at its specifications (www.schoeps.de). The microphone's sensitivity is 13 mV/Pa, the maximum undistorted level is 132 dB SPL, and the noise floor is 16 dBA. A scene recorded with this microphone that has exemplary sound for dialogue occurs in *Almost Famous*.⁷

While these numbers with their strange units may at first glance seem impenetrable to all but engineers, figuring them out will help us prevent massive distortion on peaks, among other things, so it's a good idea to understand them. The easiest item to get from the numbers is the dynamic range of the mic. It is 132 dB minus 16 dB, which equals 116 dB dynamic range from the input overload point to the weighted noise floor.⁸ We can note immediately that 116 dB is a whole lot more than our selected camera can accommodate, which we measured to be 69 dB. So anything we can do to adapt the "gallon" of the dynamic range of the world into the "half gallon" of the dynamic range of the mic, and thence into the "quart" of the dynamic range of the medium, is a good thing.

⁷ Boom operator Don Coufal, production sound mixer Jeff Wexler, dialogue editor Laura Harris, re-recording mixers Rick Kline, Paul Massey, and Doug Hemphill. Dreamworks DVD 87818 Chapter 23, 3:30–5:52. Russell: "So, this is where the enemy sleeps."

⁸ Weighting the noise floor means adapting the noise measurement with a frequency-by-frequency response curve, like a tone control, that more or less accounts for the low-level characteristics of human hearing.

The sensitivity specification is 13 mV for 1 pascal (Pa). One pascal is a unit of pressure equivalent to a sound pressure level of 94 dB. With the microphone overload point at 132 dB SPL, we have

$$132 \text{ dB} - 94 \text{ dB} = 38 \text{ dB headroom}$$

above the reference level, so we have to figure out what is 38 dB above 13 mV, a mixed bag of units. The term dB specifies only a ratio, so in this sense it is a multiplier times the voltage. There's an exact way to do this and a way to crib the result. Let's start with the crib, because that'll get us there faster: 38 dB is 40 dB less 2 dB. And 40 dB is 100 times, and 2 dB is 20 percent less (small numbers of dB are fairly linear). So 100 times 13 mV less 20 percent is 1.04 V. (If we'd been a little more observant, we would have found that the spec sheet tells us the overload is 1 V, and we wouldn't have had to calculate it!)

In order to generalize this, we need the real math solution. In order to reverse the logarithmic effect of decibels, what we have to do is the following:

1. Take the number of decibels and divide by 20.
2. Use the result of the division as a power of ten: 10^x .
3. The result is the ratio that corresponds to that number of decibels.

So to get the final answer, multiply the ratio by the starting point for the measurement. For our example:

1. $38 \text{ dB} \div 20 = 1.9$
2. $10^{1.9} = 79.4$
3. $79.4 \times 0.013 \text{ V} = 1.0326 \text{ V}$

The microphone can put out a little over 1 V, but the camera's input overloads at 48 mV (0.048 V). We've got a significant problem, at least potentially. First let's see what the maximum SPL is that we can record without mic preamp distortion. To do this, compare the sensitivity of the mic to the input overload:

$$48 \text{ mV} \div 13 \text{ mV} = 3.692.$$

So the output of the mic can be 3.692 times the level it would put out at the standard reference SPL of 94 dB. What is that in decibels?

$$\text{dB of the given voltage ratio} = 20 \log_{10} 3.692 = 11.3 \text{ dB}$$

And 94 dB SPL, which is the level for 13 mV, plus 11.3 dB, which is the factor above 94 dB where the mic input distorts, is 105 dB SPL.

Many scenes fall under 105 dB SPL, and much of the time you'll get away with a setup that plugs this mic directly into this camera. For instance, the ordinary dialogue measured in the table is easily accommodated, with about 19 dB remaining as headroom, plenty to accommodate even short-term peaks. Conveniently, the Panasonic camera supplies phantom power to the microphone, so all you need is the mic and an XLR cable and you're in business. You'll also need accessories such as a boom pole, shock mount, and windscreen, but these are incidental to our purpose here. However, if the actor suddenly shouts, the signal output by the mic will increase greatly in level and it is almost certain that gross distortion will result, even if the main level control is jerked down; to reiterate, the distortion is occurring when the signal output by the microphone overloads the camera input *before* the opportunity arises to reduce the signal with the main level control.

Cries

With this simple connection, then, you must beware of sounds above 105 dB SPL—they will distort. Because the mic can handle 132 dB cleanly, and the camera input only 105 dB when the two are directly connected, what we need is a pad, a kind of neutral density filter for sound. A neutral density filter is a gray glass disc put in the light path of the lens that is used to reduce the intensity of bright scenes to the capability of the camera's light sensor. Likewise, a *pad* is a device to knock down the level of the output of the microphone so that louder scenes can be recorded without distortion.

What we need is simple: a 27 dB pad, if we want to maximize the capability of the system, matching the maximum undistorted output of the microphone to the maximum undistorted input of the camera. Unfortunately, however, it's not a simple thing to just go to the store and buy such a device. Perhaps you could get a 20 dB pad premade, but not a 27 or even 30 dB one (most audio people have never heard of such large amounts of padding!). So we'll have to build one (Fig. 6-2).

The first thing to understand is that by putting a pad between the output of the microphone and the input of the camera, we are giving up on the



Figure 6-2 The exterior of a custom-built pad.

phantom power capability of the camera supplying the microphone, and we'll have to supply power externally. A battery-powered box in line will do the job; one such is the Professional Sound Corporation (PSC) 48PH phantom power supply. Follow that with a 27 dB pad on the way to the camera and we will have achieved our purpose.

A pad for a balanced line such as a microphone line consists of three resistors (Fig. 6-3).⁹ Pin 1 of an XLR connector is both ground and the shield connection and is wired through (pin 1 female to pin 1 male) and connected to the body of the connector as well. Pin 2 is connected with a series resistor between input and output, as is pin 3. A parallel resistor goes across pins 2 and 3 of the male end of the connector. Together, the three resistors form a voltage divider. Here is how it is calculated:

We need a -27 dB pad. We select a 1 k Ω resistor for the branch that is in parallel with the camera input,¹⁰ and now we have to calculate the two

⁹ Today at least. Very much older designs of pads had to take into account the source and load impedances in a matching type of impedance environment, such as 600 ohms. Nowadays we don't worry about such things because source impedances are so low and input impedances so high, although if we were to calculate the pad with the source and load impedances in place, we'd get a *very* slight difference in attenuation.

¹⁰ Because that will provide a low enough source impedance to drive a reasonable amount of cable.

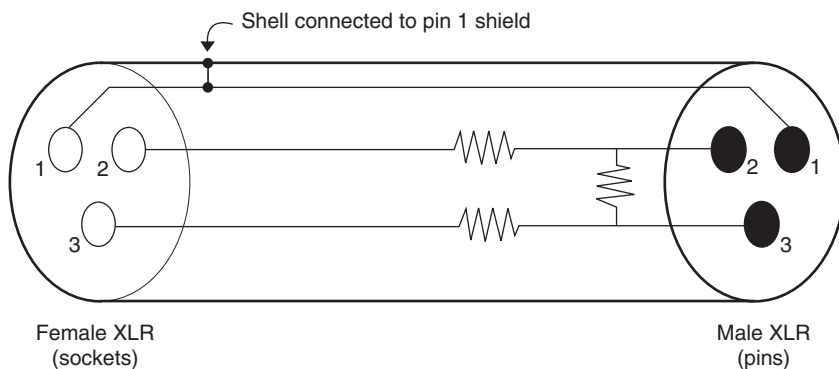


Figure 6-3 The schematic of a pad, showing the XLR pin numbers.

series resistors (Fig. 6-4). First think of them as one. The parallel branch represents the fraction of the voltage that we want delivered by this voltage divider. The equation for such a voltage divider is:

$$-27 \text{ dB} = 1 \text{ k}\Omega / (x + 1 \text{ k}\Omega)$$

Since -27 dB is another decibels ratio, we can undo it as before (divide by 20, the result raised to the power 10^x):

$$-27 \text{ dB} \div 20 = -1.35$$

$$10^{-1.35} = 0.044668$$

Thus, for our desired resistance,

$$0.044668 = 1 \text{ k}\Omega / (x + 1 \text{ k}\Omega)$$

Doing some high-school algebra means rearranging terms to solve for x :

$$0.044668(x + 1000 \Omega) = 1000 \Omega$$

$$0.044668x + 44.668\Omega = 1000 \Omega$$

$$x = (1000 \Omega - 44.668\Omega) / 0.044668 = 21.4 \text{ k}\Omega$$

Dividing this in two to build our balanced pad makes it $10.7 \text{ k}\Omega$. Looking up the closest 1 percent resistor from an E96 chart of values, we find $10.7 \text{ k}\Omega$ falls exactly on a 1 percent part value. (For the case of any arbitrary pad value, the closest 1 percent values have a negligible error for this purpose.)



Figure 6-4 (a) In our example, a 1 k Ω resistor is first mounted between XLR pins 2 and 3 of the male connector side of the barrel adapter.



Figure 6-4 (b) Then, two 10.7 k Ω series resistors are added as shown. The free ends are then clipped and extended with hookup wire to the corresponding pins 2 and 3 of the female connector, as is the ground, pin 1.

So, with this pad in our system, we can now record up to 132 dB SPL cleanly. Because this is a higher level than most people have heard of (after all, isn't 120 dB the threshold of pain?), aren't we being a little crazy? Well no, not really. An actor screaming measured 0.5 m away produces peaks of 128 dB SPL, and Lewis Fielder has recorded maximum peaks of 129 dB in the audience at 36 live music concerts.¹¹ This is close enough to 132 dB that having a method of recording to higher levels without distortion, for unusual situations, would be a good idea. For instance, I (Holman) measured the peak level inside the bass drum head of REO Speedwagon in the early 1970s and got 138 dB!

There is a way to record with low distortion to higher levels than 132 dB with this and similar microphones: add an additional pre-electronics pad. A pre-electronics pad is inserted between the microphone pickup capsule and the microphone's own electronics. It reduces the level before distortion can occur within the microphone electronics, and raises the headroom correspondingly. For some models of microphones, these are built-in and switchable. For the Schoeps mic, they come as screw-in devices made in two values: 10 and 20 dB, Schoeps models DZC 10 and DZC 20, respectively. For instance, I had to use the DZC 10 attenuators to record the University of Southern California marching band at homecoming because of the high peak levels that a marching band can produce. With these in place, we can record levels up to 142 or about 150 dB SPL for the two models, respectively. Note that levels above 140 dB cause instantaneous hearing damage and so should be approached with caution. For the loudest screams of an actor, the 27 dB pad described previously is perhaps just enough, but without much headroom for instantaneous peaks, and perhaps a 10 dB pad would be warranted if the boom mic is near the actor, as well as earplugs for the boom operator! Also, with these pads you could record some kinds of gunfire cleanly at reasonable range, but you'd better be wearing hearing protection. In some extreme instances of close gunfire, it is possible to stretch the microphone diaphragm beyond its elastic limits and wreck it.

One reason that sound levels in the 135 to 140 dB range sound unusual to many knowledgeable people is that in many publications, tables of sound pressure levels for many everyday activities are given as average levels over a period of time. Louder sounds than speaking, when evaluated in

¹¹ Fielder, L.D., "Pre- and Post-emphasis Techniques as Applied to Audio Recording Systems," *J. Audio Eng. Soc.*, Vol. 33, pp. 649–658 (Sept. 1985).

the short term instead of as a long-term average, can produce correspondingly higher levels. For instance, a bus passing by when viewed from a seat at the bus stop in an urban setting measured 107 dB peak. The combination of the Schoeps microphone and the Panasonic camera would distort on this combination without a pad for this rather ordinary scene.

Whispers

At the other end of the dynamic range is the very quiet. Some very quiet things to record include quiet room ambiances, Foley sound effects such as clothing rustle, and the ticking of a clock as used in Bergman's *Cries and Whispers* to indicate that the scene is quiet. This is not a sound design point to be overlooked on set: the *sound* of silence is not silence. Think of it this way: the picture equivalent of silence is going to black, not a view of an empty room. A quiet space might be represented by the sound of a fly, a single cricket, or some such, but it would rarely be represented as pure silence.

Recording such quiet sounds requires quiet microphones and quiet electronics. The Panasonic camera has 69 dB dynamic range below its input overload of 48 mV. If we plug the Schoeps mic into the microphone input directly (no pads) because we're recording quiet things, how quiet can they be? The simple way to calculate this is that it is 69 dB below the 105 dB SPL that we calculated previously. Why? Because that is the input overload point for no pad and the dynamic range with the microphone directly connected. The difference is the equivalent sound pressure level of the camera preamplifier noise, 36 dB SPL, subjectively about 16 times noisier than our hearing threshold. And it is 20 dB noisier than the microphone, from its specifications, so the camera preamp noise *masks* or conceals the real noise floor of the microphone. The combination may do well enough in shooting ordinary scenes, but in order to reach down to the quietest acoustic levels, a quieter external preamplifier is going to be needed.

The Sound Devices 442 is a battery-powered field mixer with 115 dB dynamic range¹² on its microphone inputs—a number derived from

¹² This 115 dB of dynamic range may not be the actual performance because it may be measured at the two extremes: noise with the gain all the way up and referred back to the input, and headroom with the gain controls minimized for maximum headroom, two settings that can't be achieved simultaneously.

its specifications. Comparing its dynamic range to that of the camera shows an improvement of 115 dB minus 69 dB, which equals 46 dB. With each 10 dB representing subjectively twice the dynamic range, these dedicated mic mixers are at least 24 times quieter than the camera's electronics. Furthermore, doing some more calculations shows that the 442's equivalent input noise is about 12.5 dB quieter than the microphone, thus making its noise floor negligible in this comparison.

To use a dedicated microphone mixer, its output, set to line level, should be sent to the camera's inputs, also set to line level. Be absolutely certain that both the output of the mixer and the input of the camera are switched to line level, because feeding a line-level signal into a microphone-level input, even if line up on tone can be done correctly, will cause serious distortion. Set up properly, the combination of the Schoeps mic and the mixer will be able to record sounds with far less noise than could be achieved by plugging the mic directly into the camera's microphone input.

If we are really going for the quietest sounds, we might want to pick a microphone with an even lower noise level like the Neumann TLM-103. By virtue of its large diaphragm, it can be quieter than the Schoeps, at the expense of having greater variation in the off-axis frequency response. Thus we probably would not want to use the Neumann as a boom mic because, among other things, it is much larger and heavier, but it may excel as a specialized Foley mic, for instance. Its noise floor is equivalent to 7 dB SPL (A weighted), about the lowest among recording microphones, and its sensitivity is such that the output noise is still above the noise floor of typical field mixers.

To come full circle, what about the other end of the dynamic range, the loud end, for these combinations: the Schoeps and the Neumann microphones into the mixer? The Schoeps can produce about 1 V, but a Shure FP-33 mixer, for instance, distorts at an input level 10 dB less, thus requiring a 10 dB pad in order to reach 132 dB SPL. Unfortunately, putting in such a pad will kill the phantom power available from this mixer, and we'll have to use an external power supply, just as described for the Panasonic camera. With the simple connection without an external power supply and pad, and the Shure providing the phantom power, the combination clips at 122 dB SPL, useful for many day-to-day purposes, but not reaching the highest levels of a screaming actor without distortion.

The Neumann has a maximum output level of 3.5 V, so with the Shure mixer a 21 dB pad will take the undistorted equivalent level up to 138 dB SPL, or, looked at another way, with no pad a level of 117 dB SPL can be handled by this combination.

Cries and Whispers

While we've discussed handling loud sounds and soft ones, what if they happen simultaneously within one scene? This might occur if an actor in a quiet space, murmuring to himself, suddenly shouts.

Digital overload, like microphone preamplifier clipping, comes on abruptly and generally sounds pretty horrible. It is most noticeable on voice content in the vowels. Consonants are sharper, harder, faster sounds, which means they take up a wider frequency range, thus tending to hide distortion. Vowels, on the other hand, have a more orderly waveform, repetitive from cycle to cycle, and a well-defined fundamental and harmonics at twice, three times, and so forth times the fundamental frequency. This spectrum means that distortion, which tends to spread the content across a wider frequency range, will not be masked or hidden. So the first thing to give an indication of distortion is typically distorted vowels.

The most common way to handle a scene with wide dynamic range is to ride the gain, that is, adjust the microphone channel level control to adapt to the sound being recorded so that the loud passages don't distort and the soft ones are well above the noise level of the electronics. Naturally, this is only effective if the potential distortion is caused by over-recording on the medium and not clipping at the microphone input, as described earlier. Riding, or adjusting the gain constantly, does no good if the signal is already distorted or noisy because of the conditions at the microphone preamplifier *input*. The main recording level control on cameras or mixers is *after* the microphone preamplifier, so its use is to adapt the dynamic range presented at the output of the microphone preamplifier to the dynamic range of the medium.

There are two contravening factors at work here. Constantly adjusting the gain means that within a given take the recorded level goes up and down so that the foreground sound, usually speech, is well recorded.

A problem with this method is that then the background sound is constantly varying. This causes problems editorially, where a cut between two recorded sections could result in a large change in the background sound. So there are two conflicting goals for the production sound recordist in this situation:

1. Keep the foreground sound undistorted in loud passages by turning the level down, and not electronically noisy in quiet passages by turning the level up to keep the signal above the noise level of the medium.
2. Maintain the one best level within a whole scene, including shot-to-shot setup variations, which will produce the most cuttable sound by *not* riding the gain.

How does this dilemma get resolved? First, the degree of the dilemma depends on the dynamic range of the recorder. When early film sound optical tracks had little range, about all that could be done was that in order to have something sound loud, something quiet had to come before it—it was only through contrast with the quiet preceding passage that something could be made to sound loud. This was called “the cinema mixer’s trick.” There really wasn’t any dynamic range to spare.

With modern high-capacity recording, less gain riding needs to be done, because the range is larger. However, that doesn’t mean that you cannot protect the loudest sounds from distortion—you should. What it does mean is that through training and experience one comes to know what is acceptable in terms of gain riding. It is especially helpful if you edit your own sound recordings, because the feedback that this produces makes you a better mixer. No one prescription can be given, except to consider the following:

- Human mixers, as opposed to limiters and automatic gain or volume controls called compressors, can anticipate the action, particularly if there are rehearsals for a fiction piece. One can sneak the gain riding in *before* it is needed and in an otherwise quiet place.
- Boom operators can be especially helpful in leveling out performances from various actors/subjects. This is particularly true if a hypercardioid rather than a shotgun microphone is in use because the hypercardioid typically has much smoother response off-axis: it

sounds mostly just like attenuated on-axis sound, so a stronger actor can be picked up at a more off-axis angle, while a weaker one is favored by being recorded on-axis. If this is tried with a shotgun, the off-axis sound coloration is too great to accept.

- Certain sounds can afford to be recorded with distortion, because they are distorted anyway. Gunshots come to mind: keeping them within the undistorted range in a scene containing other sounds to be recorded would probably require just too much change of the background sound to accommodate good recording. It is easier to simply let the system distort briefly, and then, if needed, edit in cleaner gunshots in postproduction.
- Once a wide dynamic range is available—say by use of an external mixer and staggered levels between the two channels as in the following discussion—then riding the gain gives way to setting the level, that is, setting the level just once for the loudest part of the scene to be certain it's undistorted, then leaving it alone. This produces the most editable sound. Any gain riding that will be necessary for intelligibility and performance can then be done in postproduction, when there is time to go over and over it as many times as necessary.

Another specialized way to get a gallon of dynamic range into the half-gallon of the camera's dynamic range is to divide the gallon into two different half-gallons. You could, for example, following the microphone battery power supply, put in a Y-adaptor and send one signal to both inputs of the camera, with one of them having a pad and the other having no pad. In this way, the padded version would capture the loudest part of the performance cleanly, but with noisy lower-level passages, while the second channel would have distorted high-level passages but be quieter in low-level ones. This requires a lot of coordination between production and postproduction sound—after all, one track *will* be distorted some of the time, but it is a way of capturing a wider range and using the second channel, when what is desired is monaural (single microphone) boom operation.

Using this tactic with the Schoeps mic and Panasonic camera, for instance, could theoretically extend the dynamic range of the recording from 69 dB to 69 plus 27 dB which equals a 96 dB range, a great improvement.

That would cover sounds from an equivalent 36 dB to 132 dB SPL. We'll see how to set the levels on this combination in the next section.

Multiple Level Controls in the Chain

A chain consisting of a microphone and two potential places to pad its output has been described. Real-world situations with a camcorder using an external mixer have many more points in the system where the level may be affected. For instance, in the order that the signal encounters them, these could be:

- Microphone pre-electronics pad, previously described.
- Microphone level in-line pad, previously described, which may be a part of an external power supply or a separate device placed in the microphone line. If one is used, it does not pass phantom or T power, which must be supplied before the pad (often in the same box as the pad).
- Mixer's input sensitivity switch, Mic/Line, or various Mic levels and Line. For instance, changing from 48 V phantom power (implying an electrostatic microphone) to a DYN input (for an electrodynamic microphone) may adjust the input sensitivity of the mixer, because electrostatic mics are normally "hotter" than electrodynamic mics (more output for the same SPL) by something on the order of 20 dB.
- Mixer's input trim control (pad). Replaces the need for external microphone-level pads with a variable input gain or pad on the mixer. This is only available on typically larger and more expensive mixers.
- Mixer's channel level control.
- Mixer's master level control.
- Mixer's output level switch: Mic/Line. This switch is provided so that if the only input provided by the camera is at mic level, it will not be overloaded. If mic/line switching is available at both the mixer output and the camera input, line level is preferred because the higher level is less susceptible to electrical interference.

- Camera's input level switch: Mic/Line. Set to correspond to mixer's output.
- Camera's input level control.
- Camera's monitor level control. Also, there is often a switch that provides monitoring for channel 1, channel 2, or channel 1 plus channel 2.
- Mixer's monitor return level control.
- Mixer's headphone level control.
- Standalone headphone level control, such as in a distribution box for production sound operator and boom operator (potentially separate).

At first this can seem to be almost hopelessly complicated, because there are so many places for the level to be set. One of the primary reasons for this is that each component part in the chain has to be prepared in the marketplace to face all kinds of conditions; hence, for instance, the mic-level output on a microphone mixer. That only exists because sometimes the only thing available on the next piece of gear is a microphone input, but there is a way to simplify the settings for this chain. First is to understand the signal flow through the chain so that troubleshooting can be done if the signal goes awry. Second is to know the nominal setting of all parts of the chain, so you know when you are deviating from normal. Third is to calibrate the level for those parts of the chain that are metered, so that the mixer and camera are adjusted to the same sensitivities, and what is read on the mixer will read on the camera; then the mixer becomes an extension of the camera. Fourth is to understand the consequences of getting the levels wrong. Let's take each of these up in turn:

1. The signal flow is given in the aforementioned order. However, it is a fairly comprehensive list, and any single set of equipment may have fewer parts. It is worthwhile to draw a diagram of your particular system with all of the places that affect level, so that you can follow the logic. A typical example with an external mixer and using single-system recording is given in Figure 6-5.
2. We've already had a discussion about the dynamic range at the mic and on the medium. Setting levels correctly helps the gallon into half-gallon into quart problem, and some of this has already

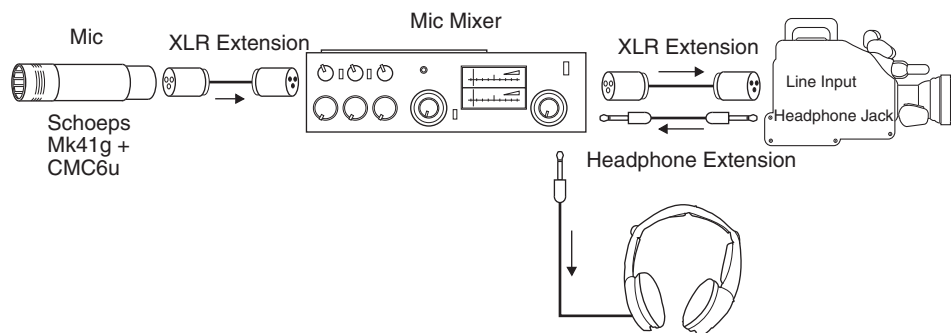


Figure 6-5 Block diagram of single-system shooting setup with external microphone mixer. The return of the headphone signal from the camera to the mixer, and monitoring by way of that path, ensures that the signal has made it to the camera.

been discussed in the context of the microphone input. For the whole system, nominal settings need to be applied. For instance, it is useful to know that level controls normally are run at 7 on a scale of 10 or near 0 dB (unity gain) on a scale from $-\infty$ to +15, that headphone monitor controls on camera are often operated fully up, and so forth.

3. Calibration is necessary in a system with a separate mixer and recorder. Start with a tone generator from the mixer, set to a standard level such as 0 VU, and then set the mixer output switch to line level, the camera input sensitivity switch to line level, and adjust the camera input level control for a standard meter reading, such as -20 dBFS. Because two different types of meter are in use, VU and digital peak, the readings for the nominal setting are *not* the same.
4. The consequence of getting the level wrong at one point in the chain and making up for it at a later point can be added distortion or noise. This may be a small problem or an overwhelming one, depending on the degree of the mismatch. Say in one stage the level is down by 10 dB and then subsequently restored to full level. Any noise added to the signal in stages between the two level controls will be exaggerated by 10 dB. Worse is clipping distortion early in the chain, which we attempt (unsuccessfully) to fix later in the chain.

Once clipped, the signal remains distorted no matter what you do with subsequent level controls.

Another Kind of Overload Distortion and How to Avoid It

So far we have described the overload distortion called clipping and explained how to use pads before and/or after the microphone's own electronics to maximize headroom. There is an invisible type of this overload possible too, and it plagues all directional microphones especially: infrasonic overload. Directional microphones of all types are much more susceptible to wind noise, boom panning noises, and other very low-frequency noises, compared to omnidirectional types. But directional mics are necessary in most film/video recording applications, just to emphasize the direct sound and suppress reverberation and noise occurring in most settings. The only exception to this rule is for lavalieres, which are mostly omnidirectional and therefore less wind susceptible than other types.

Infrasonic (formerly called subsonic) overload from wind or boom panning leads to bobbles in the audio—little moments of time where the sound seems to be compressed or even go away. Once you've heard it, you'll be able to recognize it. Electronic infrasonic overload may occur at two potential places in the chain: right in the microphone's internal preamplifier or in the camera or mixer's microphone preamplifier. Either of these may be prevented by putting a pre-electronics low-frequency filter (called alternately low-cut or high-pass) between the pickup capsule and the microphone body. In the case of the Schoeps, this is a screw-in device called a CUT-1, with adjustable low-frequency response above a steep cutoff frequency. In some other microphone models, the low-cut switch may be built in; however, in that case it is not necessarily clear whether the low-cut switch occurs before or after the electronics. Placing it before the microphone's own electronics most improves the headroom, whereas placing it after protects the rest of the chain—in particular, the microphone preamplifier in the camera or mixer—from infrasonic overload.

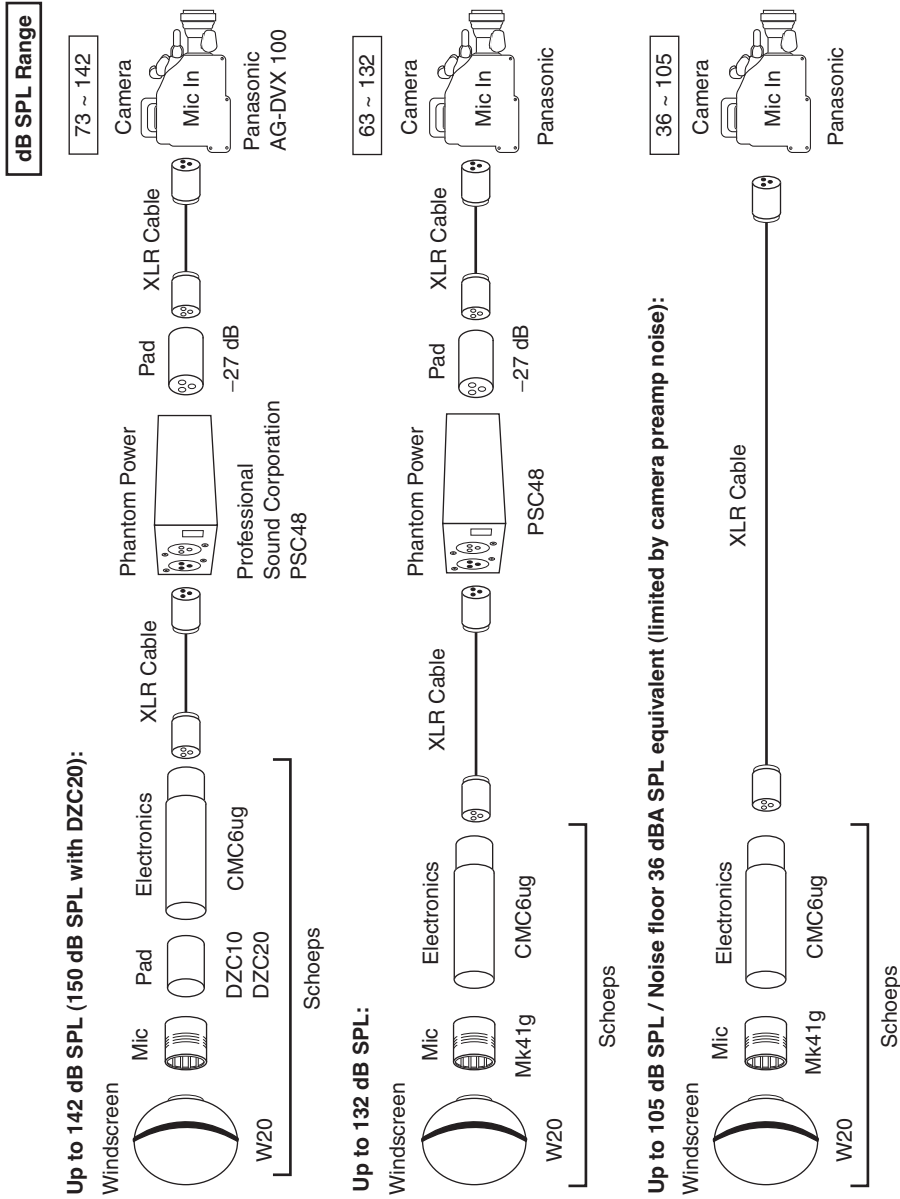
Generally speaking, for recording dialogue one can use a 60- or 80-Hz high-pass (low-cut) filter and do little or no damage to the voice while

decreasing the intensity of wind and panning noises dramatically. This is especially true of steep filters, like the Schoeps CUT-1, a 24-dB/octave filter below 60 Hz.¹³

Combining Features for Best Wind Performance

The combination of ingredients that produces the lowest wind susceptibility involves using an omni pickup capsule, protecting the capsule from wind, protecting the electronics from low-frequency and broad-frequency range overload, and protecting the camera or mixer microphone input from excessive amounts of signal (Fig. 6-6). A system consisting of a Schoeps MK2 omni capsule plus DZC20 pad plus CUT-1 filter plus CMC electronics plus BBG windscreen with Windjammer should have the greatest attenuation of wind effects possible. The output level of the mic should be scaled to the input capability of the mixer, possibly with an additional pad, and an additional mixer high-pass filter might not be out of place (Tables 6-2 and 6-3). An omni capsule implies that one must use the closest possible working distance, but the advantage of low wind susceptibility may overcome the working distance problem.

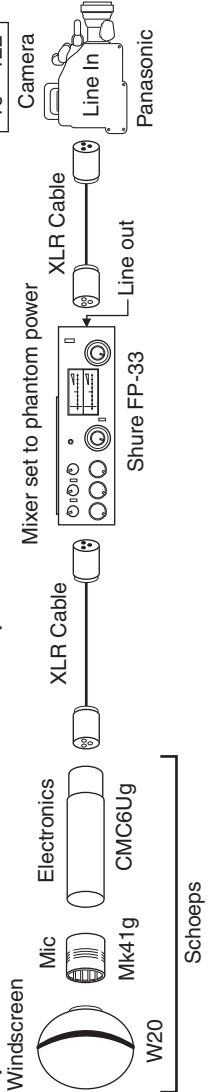
¹³ If this particular Schoeps CUT-1 is used, there are other considerations: its use increases sensitivity by 5 dB, but decreases headroom by 5 dB. If an attenuator is also used, it is placed on the pickup capsule first, followed by the CUT-1 filter.



dB SPL Range
**

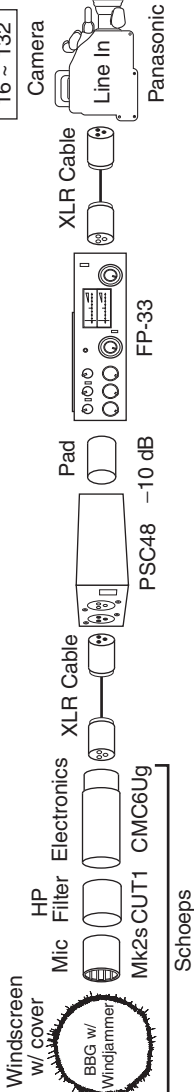
16 ~ 122

Up to 122 dB SPL / Noise floor 16 dBA SPL equivalent:



Best wind performance (mic pads may also be used to reach higher undistorted levels):

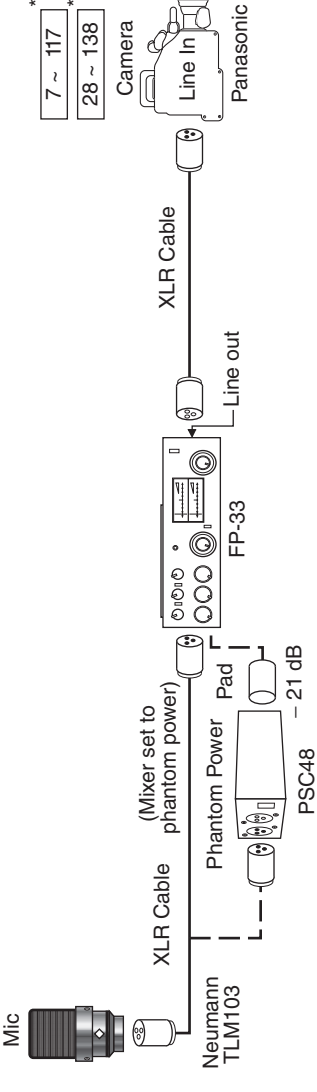
16 ~ 132



Best Noise performance / Noise floor 7 dBA SPL equivalent (High levels require external power supply and pad):

7 ~ 117

28 ~ 138



** These extremes are not available simultaneously, but rather represent the range over which the mixer may be adjusted to work.

Figure 6-6 Cont'd

Table 6-2 Some common microphone specifications.

Model	Description/Where Used	Sensitivity at 94 dB SPL, mV	Max SPL	Noise (SPL* Equivalent)	Dynamic Range, dB
Audio Technica AT835b	Shotgun/inexpensive boom mic	12.5	130	24	106
Audio Technica 4053a	Hypercardioid/inexpensive boom mic	22.4	146	19	121
Neumann TLM 103	Cardioid/very low noise studio mic, Foley	21	138	7	131
Neumann KMR 81 i	Short shotgun/good-quality highest SPL shotgun, boom mic	18	128 (138)	12	116
Neumann KMR 82 i	Long shotgun/good-quality outdoor boom mic	21	128	12	116
Schoeps CMC641Ug	Hypercardioid/good-quality, highest levels, an industry standard	13	132 (142)	16	116
Schoeps CMC68Ug	Bidirectional used with above for MS stereo	10	134 (144)	18	116
Sennheiser MKH 60-1	Short shotgun boom mic	40 (12.5)	125 (134)	18 (25)	107
Sennheiser MKH 416-P48U3	Short shotgun/industry standard boom mic	25	130	13	117
Sennheiser MKH 70-1	Long shotgun boom mic	50	124 (132)	5 (13 with pad)	119
Tram TR50	Standard lavalier	16	134	26	108

* A weighted
Values in parentheses are with built-in pad enabled.

Table 6-3 Input overload and pad values needed for selected microphones with selected audio inputs to match the microphone maximum output to the electronics microphone input maximum undistorted level.

	Beachtek DXA-6	Sony DSR-PD170	Panasonic AG-DVX100A	Shure FP-33
Mic input overload, mV	880	141	48	316
Mic input overload, dBV	-1	-17	-26	-10
Microphone Make, Model	Required pad value, dB			
Audio Technica AT835b	None	-15	-24	-8
Audio Technica 4053a	-20	-36	-45	-29
Neumann TLM 103	-12	-28	-37	-21
Neumann KMR 81i	None	-16	-25	-9
Neumann KMR 82i	-2	-18	-27	-11
Sanken CSS-5	None	-13	-22	-6
Schoeps CMC641Ug	-2	-18	-27	-11
Sennheiser ME66	-4	-22	-31	-15
Sennheiser MKH 60-1	-4	-20	-29	-13
Sennheiser MKH 416-P48U3	-5	-21	-30	-14
Sennheiser MKH 70-1	-5	-21	-30	-14
Tram TR50	-5	-21	-30	-14

Director's Cut

- Microphone electrical outputs have a potentially very wide dynamic range, from very small to rather large voltages depending on the scene they face. The range can be tamed through the use of pads, which work like neutral density filters.
- If the microphone output voltage exceeds the input capability of the microphone preamp, the sound will be distorted *no matter where the level control is set*. Pads are necessary to keep the maximum output of the mic from exceeding the input capability of the preamp.
- Setting the record level is the most important act of a recordist. Because digital recording distorts abruptly when overloaded, it is better to under-record somewhat than risk ever over-recording.
- To record very soft sounds, like Foley and some ambiances, a camera mic preamplifier may be too noisy, and use of a better outboard mixer or preamp is necessary to avoid hearing hiss.
- Multiple level controls in the chain can be confusing, and the best way to work through the confusion is to know what the nominal settings of all the controls and switches affecting level in the chain are and to set them there to start.
- Wind noise can be avoided through use of omnidirectional microphones, high-pass filters, and efficient windscreens.

CHAPTER 7

Media Management

This chapter covers both file management and the management of physical media (memory cards, tapes, hard disks, and optical discs) on which recordings and backups are stored. Media management is an increasingly complex process as new formats for recording and backing up information enter the marketplace. Indeed, on some independent films (especially multi-camera productions) it is common to have a crew member whose sole responsibility is to transfer recordings from camera original cards onto an editing system and backup each set of recordings onto one or more hard drives for redundancy. The importance of backups cannot be overstated in an environment where memory cards are constantly used, transferred, erased, and reused over the course of a production. We'll look at how media management and file transfer operations can most effectively be accomplished on card-based, tape-based, and disk-based systems, in both single-system (camera recording audio) and double-system (audio recorded on a separate device) setups.

What Counts as a Backup?

The shift from a tape-based recording paradigm, in which it was common to use tapes only once and keep them as the primary backups of recorded footage, to a card-based recording paradigm, in which flash memory cards are used to record and then reused once the video files are transferred to an editing/ingest station, has removed the security of keeping camera originals as backups and put a premium on making sure you have backup copies safely made before reformatting a card. In a workflow recording on tape or write-once optical discs, the original recording medium is like the camera negative in a film process: it is generally kept as a record of the full-resolution, unprocessed source material, used to retrieve material for online sessions, and saved as the archival copy of the original video even after a

project has been edited.¹ Thus camcorders shooting on Mini DV, DVCAM, HDCAM, and other formats give productions a primary backup medium “for free” as long as a new tape is used for each reel of each production. The downside is that enough tape must be purchased to capture all the footage shot for a project, which could add up quickly, particularly for documentary projects. Depending on the tape format in use, buying tape could be quite expensive, and the archival copy this workflow provides “for free” might end up costing a considerable portion of the production budget.

One advantage of non-linear file-based recording, then, is the reusability of memory cards once files are transferred to a hard drive for backup and ingested into an editing system to verify their integrity. Unlike copying a tape into an editing system, in which the tape is typically saved as a backup of the original footage, copying the contents of a memory card into an editing system does not constitute backing up the footage if the memory card is meant to be reformatted and reused later in the production. Generally, an additional off-line backup is made by copying the files onto a second detachable hard drive or a second system before the card is wiped and reused in the camera, and productions may even make a second backup of all material shot during a day of production onto another hard drive overnight so that only one backup drive is connected to a system and a backup copy exists on an off-line drive at any given time. The needs for backing up footage vary from production to production, based on the type of material being shot and the projected expense of recreating footage if it is lost in a system failure; but usually having at least two copies of the footage shot for a given project is a good idea, particularly when no tape original exists to fall back on as an archival backup, since unexpected hard drive failure is always a possibility and always seems to happen at the worst possible moment.

Although reusable memory cards introduce additional requirements for backing up footage during production, non-linear file-based recording workflows offer additional hidden benefits such as the ability to review footage on set without the necessity of rewinding tape and the corresponding danger of writing over the end of a shot with the next recording. With memory card systems designed to record each clip as a new file until the card fills up, at which point the card is swapped out for an empty or

¹ Magnetic tape is more generally considered an appropriate archival format than an optical disc, which may have an uncertain shelf life, and professional formats with greater footprint on the tape offer better archival properties than consumer DV tapes.

reformatted card, there is no longer the danger of rewinding and recording over a shot accidentally as there was with tape-based systems. In order to record over a shot, the camera operator would have to enter the menu on most cameras and intentionally erase a file or reformat the card being used. Moreover, transferring video and audio files off a memory card is as simple as copying any other type of file between hard drives or portable devices. Whereas recordings on tape must be read linearly, typically in a real-time streaming transfer, files on a memory card or hard drive can be transferred and backed up much faster than real time. The peculiarities of both streaming and file transfers will be covered in more detail later in this chapter.

The principal danger in reusing media to record digital video lies in the requirement that the camera operator must be sure that the media have been backed up properly before reformatting a card to continue shooting, or an entire reel of footage may be lost. Unless there is someone on set to track which cards have been backed up and to verify the integrity of transferred footage, the danger of losing footage will be greatly increased. This person may be the director or camera operator in small crews as long as they have time to oversee transfer and backup operations effectively. Note that is very common to have multiple memory cards in use on a digital video production. Even the smallest productions may have two cards in use at any given time: one being transferred and backed up while the other is in the camera so that production can continue without waiting for the first card to download. It can be extremely helpful to have a system of marking which cards are live (currently being recorded), which are being backed up or awaiting transfer, and which have been transferred and are ready for reformatting in camera. Without a physical system in place such as marking cards with colored tape or storing cards ready for reuse separately from those that have just come out of the camera, there is often no way to know whether a card has been backed up aside from going back to the production hard drives and comparing the footage on the card and on the hard drive shot for shot. A live card and a transferred card, after all, look identical in all respects.² This fact

² It may be possible to mark cards that have been backed up by physically deleting the files from the card once they have been transferred to disk, but generally speaking the cards are reformatted in camera so that recording begins on a newly initialized file system and so that footage is not accidentally deleted from live backup drives while an operator is attempting to delete the files off a transferred memory card. Disconnecting a memory card but keeping the files on the card until the integrity of the transferred files is verified is also good practice to ensure that an accurate off-line copy exists before the card is reformatted.

underlines the importance of being sure backups exist before a card is given back to the camera operator to be reused. Having a system in place to mark live cards, transferred and backed-up cards, and empty/reformatted cards can save a lot of stress, headache, and potentially heartache in the heat of the moment on set. That's why having a dedicated crew member to transfer and backup footage on large enough productions can be a boon to operational efficiency, akin to the loader on a film set whose job it is to load and unload film magazines in a changing bag without exposing the film, so that no footage is lost but the production never has to wait for media to become available to continue shooting.

On smaller or documentary productions, it is common for the camera operator or director of photography (who may also be operating camera) to offload cards onto a hard drive or computer editing system since generally he or she will be the crew member most familiar with the file systems and file formats in use, the protocol for transferring footage so that it can be imported correctly into an editing system, and the protocol for reformatting the card using the camera menus. When transferring footage into an editing system and verifying that video and interleaved audio can be read properly, one peculiarity is that there may be different qualities of footage recorded at once by some cameras—the full quality video and a thumbnail or preview Quicktime for low-resolution instant playback—and it may be important to set up an editing system correctly so that the full-resolution version is imported into a video editing application or can be verified and referenced correctly in an online/off-line setup.

Types of Media

The range of digital video recording media in use today includes memory cards, tapes, optical discs, and hard disks. In some cases, cameras and audio recorders may record to more than one of these types and may even record to multiple media simultaneously. In newer digital video cameras, including both prosumer camcorders and digital SLRs used for shooting video, solid-state media predominates, although in consumer models hard drive and optical disc recording remain popular. And all four media types, including digital tape formats, remain in use on earlier models of production equipment that are still available today.

For solid-state systems, the particular type of memory card called for by a camera varies by manufacturer and, in some cases, across the model line of a given manufacturer. Many Panasonic cameras such as the HVX200 and HPX170 use P2 cards, for instance, while other models like the AF100, HMC40, and AC7 and their siblings record onto standard Secure Digital (SD) cards, and some models record onto DV or DVCPRO HD tape instead of or in addition to solid-state media.³ Sony sells professional digital video cameras such as the F3 and EX-3 recording on SxS cards, while lower-end Sony camcorders may record onto hard disk, SD or Compact Flash (CF) cards, and/or DV tape stock in HDV⁴ or DVCAM modes.

Current offerings by Canon and JVC likewise record onto SD cards, CF cards, or digital tape stock, both in their consumer and professional lines. Digital SLRs with motion picture modes typically record onto industry-standard solid-state media such as Compact Flash cards (used by the Canon 5D and 7D cameras) or SD cards (used by the Nikon D3200, for instance) since they simply record video onto the same type of media used for taking still photos. Table 7-1 gives examples of the most commonly used media for recording digital video, along with examples of particular camera models using each type.

Streaming Transfers and File Transfers

In the early days of digital video editing, video was played back in real time and digitized into an editing system at low resolution in the online/off-line workflow described in Chapter 2. As storage capacity increased and digital recording via Mini DV, DVCAM, and DVCPRO became mainstream, the introduction of the FireWire⁵ bus allowed digital video to be transferred at full resolution into an editing system with no recompression, then edited, and recorded back to tape without ever leaving the digital domain. However, the video was still transferred over FireWire in a real time *streaming transfer* at the DV standard data rate of

³ The Panasonic HVX200 was one of the early prosumer cameras offering high-definition recording on P2 cards while still allowing for standard definition recording on DV tape.

⁴ HDV is a video recording format that stores high-definition MPEG-2 video onto standard DV or Mini DV cassettes.

⁵ Also known as IEEE Standard 1394 and iLink.

Table 7-1 Capacities and common video recording uses for various media.

Media Type	Capacity (Typical)	Typical Video Formats	Representative Cameras
Secure Digital (SD), SDHC, SDXC	2 GB (SD) – 2 TB (SDXC)	MPEG-2, MPEG-4, AVC	Canon VIXIA, Panasonic HDC series
Compact Flash (CF)	8 GB – 256 GB	MPEG-2, MPEG-4, AVC	Canon 5D and 7D (DSLRs), XF100
Sony SxS Cards	32 GB – 64 GB	MPEG-2, DVCAM	Sony F3, EX-3, EX-1 (XDCAM series)
Panasonic P2 Cards	4 GB – 64 GB	DV, DVCPRO, DVCPRO-HD, AVC	Panasonic HVX200, HPX170
DV/DVCPRO Tape	1 hour	DV, DVCPRO, DVCPRO-HD, DVCAM	Panasonic HVX200
Mini DV Tape	1 hour	DV, HDV*	Canon, Sony, Panasonic, JVC camcorders
Mini (8 cm) DVD-R/RW	1.4 GB	MPEG-2*, MPEG-4, AVC	Canon, Sony, Panasonic camcorders
Internal Drive	32 GB – 64 GB	MPEG-4, AVC	Canon XA10, Canon VIXIA, Panasonic HDC series

* HDV and Mini-DVD camcorders are often limited to recording MPEG coded or AC-3 (Dolby Digital) audio rather than uncompressed PCM audio (see the discussion on audio recording formats in Chapter 1).

25 Mbit/s (3.125 MB/s).⁶ This method of transferring video remains in use in some systems, particularly those based on DV tape formats, but has been largely supplanted by *file transfers* in newer tapeless workflows.

Streaming Transfers

A streaming transfer occurs in one direction (i.e. tape to computer, or vice versa, but not both simultaneously) and takes place in real time. The video or audio being transferred is actually played back off the tape or editing system as it is sent through an interface, and can be monitored on a video screen and/or speakers as the transfer takes place. Telecine from film to video tape is an example of a streaming transfer that was developed before the age of all-digital workflows: in telecine, footage shot on film is played back alongside synced audio, with both the film and recorded audio pulled down to NTSC speed for recording onto NTSC video tape in real time. Digital video transfer via a FireWire bus is one example of a streaming transfer in digital video, but other interfaces may be used, particularly to transfer audio; examples include AES3 and S/PDIF interfaces, in addition to any setup using analog connections to import and digitize audio in real time.

The first item to notice about an interface is whether it is carrying digital data or carrying audio in analog form. Unfortunately, the problem of distinguishing the two is made more difficult by the fact that the same connectors are often used for both analog and digital signals, even on the same piece of equipment! To get back and forth between analog and digital sound requires an analog-to-digital converter (ADC) in one direction or a digital-to-analog converter (DAC) in the other. Usually these converters are incorporated into other pieces of equipment, such as cameras, and remain unseen by the user, but sometimes in high-end applications they may be stand-alone boxes just for this purpose. A camcorder contains both (Fig. 7-1): the microphone or line-level analog inputs are converted to digital for storage on the medium, and converted back to analog for monitoring, while also being available digitally from

⁶ Higher data rates of 50 Mbps and 100 Mbps were used in some DVCAM and DVCPRO formats.

the medium or even during recording over FireWire or Serial Digital Interface (SDI).

While older methods used analog video connectors to import and digitize video and audio via an ADC connected to an editing system, digital camcorders use FireWire transfer as the primary method of achieving streaming audio–video transfers. FireWire connectors were ubiquitous on early DV camcorders recording on digital tape, and productions using tape-based camcorders will most likely continue to rely on streaming transfers via FireWire for ingress and egress of material due to the linear nature of playback and recording on tape. FireWire is an inherently digital transfer method of carrying data back and forth from camera to computer and vice versa. It can also be used for recording in real time onto an external hard drive, which may be connected to a camera for recording audio and video together or to a professional field audio recorder for recording digital audio in a double-system setup. A typical FireWire cable used for connecting a camcorder to a computer for streaming video transfer is shown in Figure 7-2.

Where audio only is concerned, one use of the FireWire bus is for offloading actual recording of the digital data onto an external hard drive, as described above. However, audio transfer between devices such as a digital audio tape (DAT) deck, CD player, or microphone preamplifier and an editing system would more commonly occur over dedicated audio interfaces. The most common digital interfaces for streaming audio transfer are AES3, a professional interface, and S/PDIF, a consumer interface that may nonetheless be found in the digital video production world.

This is where it becomes important to recognize whether a connector is carrying analog or digital data, because unlike FireWire video transfers, streaming audio transfers may take place on connectors also widely used for carrying analog audio signals. In one form of the digital professional audio interface, AES3, XLR connectors (Fig. 7-3) are used, whereas professional analog audio routinely uses XLR connectors too. So the only way to know whether a given XLR port on a piece of equipment is analog or digital is by its labeling. The consequence of connecting a digital output to an analog input could be blown tweeters in monitor loudspeakers and/or power amplifier failure. That is because the digital signal is composed of a wide range of frequencies extending very far above the audio range, and this is seen by subsequent audio equipment to which it is directly connected without first being converted to analog as very strong

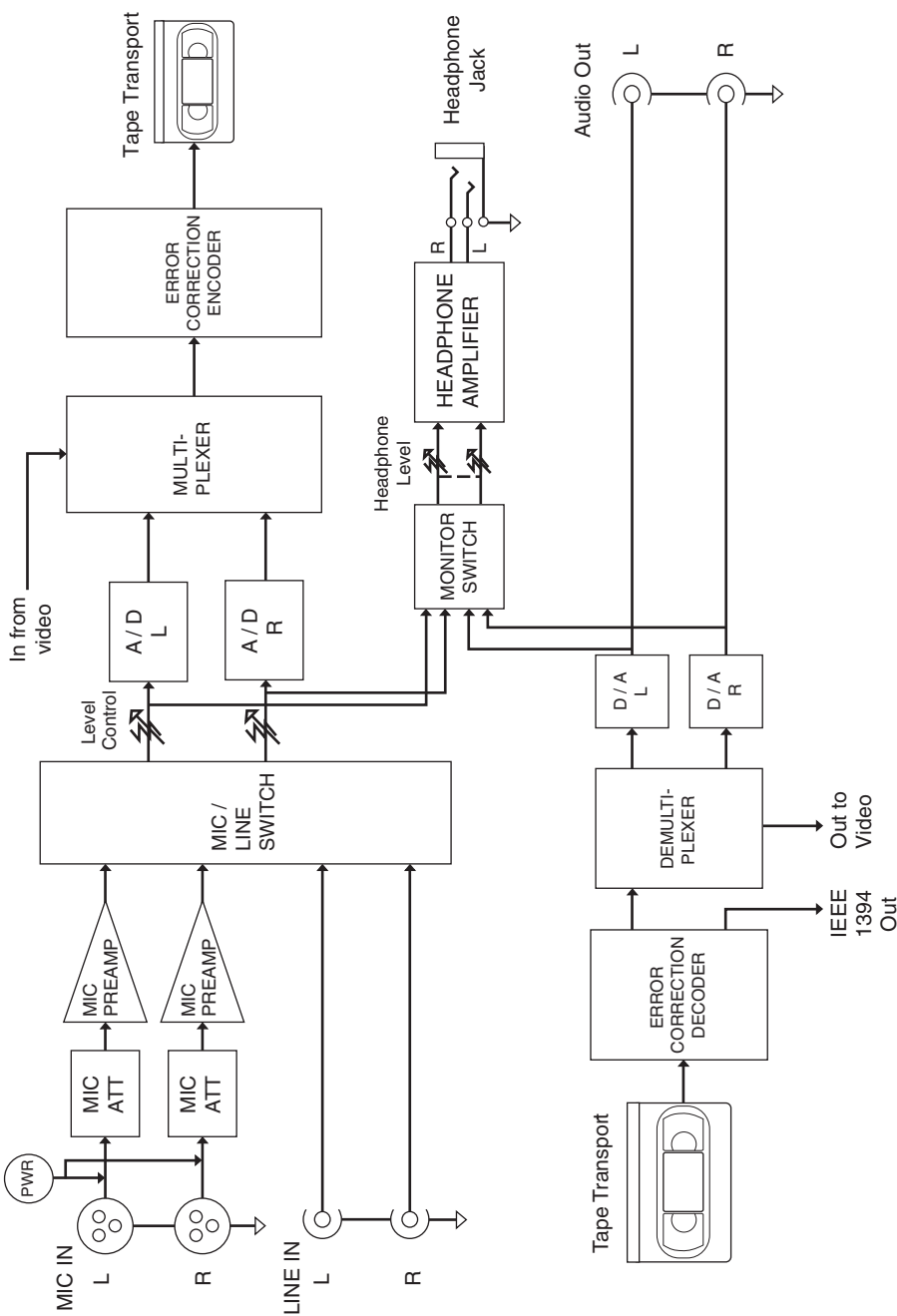


Figure 7-1 A block diagram of a camcorder.

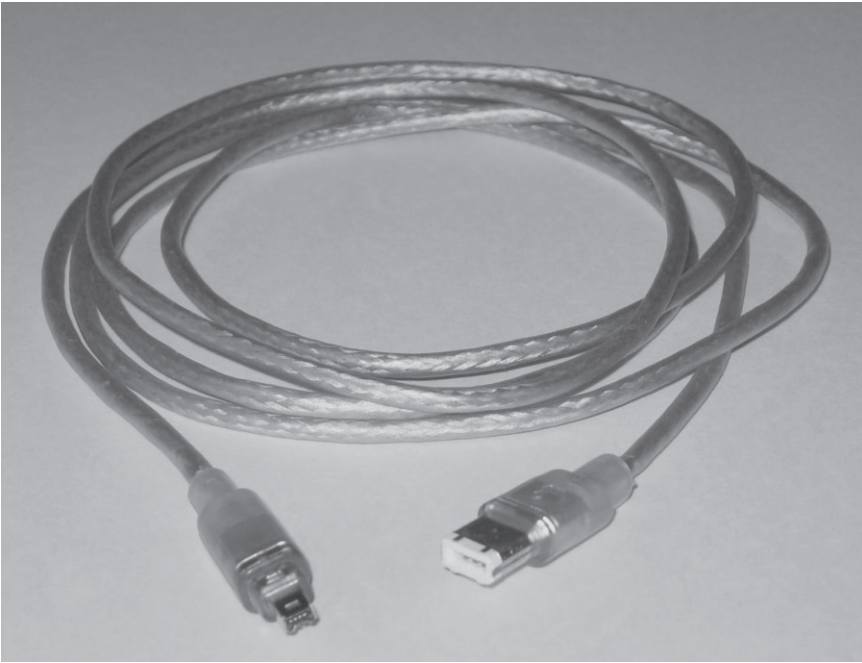


Figure 7-2 A typical FireWire cable used in connecting a camcorder (left) to a computer (right) for streaming video transfer.

ultrasonic levels, which may cause damage. Conversely, connecting an analog signal to an input intended for a digital signal will normally result in there being no sound, although damage is unlikely.

In professional audio studios, for digital audio connections the XLR connector carrying AES3 format signals is the most common.⁷ XLR connectors have a gender assignment: males, those with exposed pins, are today always outputs, whereas females having sockets are always inputs. Each three-pin connector carries two audio channels interleaved in time as one signal, and the wiring is balanced; that is, each of the two primary conductors carries a signal that is the same as the other, except that the two are upside-down mirror images of one another. The third conductor is a shield for the other two and sometimes may not be used. The two primary conductor signals are said to be out of phase or in opposite polarity from each other. The reason for using balanced wiring is that it is less

⁷ The AES3 standard is available at www.aes.org.

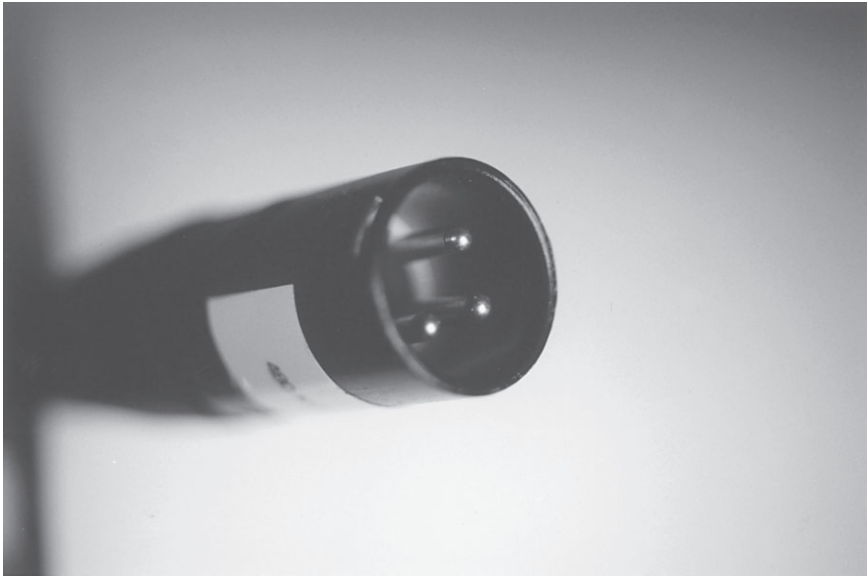


Figure 7-3 An XLR connector pair that may be used for both analog and professional digital connections.

susceptible to outside interference, particularly from magnetic fields, although these digital audio signals are already rather impervious to outside disturbances.

This system also has what is called *characteristic impedance*, a measure of the ratio of voltage and current found on the interface, and uses matched impedance, so the inputs are said to terminate the outputs. This means that it is bad practice to use a Y-adaptor to distribute a digital output signal to more than one input, because each output is expected to see just one input in a one-to-one relationship. The impedance is 110 ohms, and this may make the use of Cat 5 unshielded twisted-pair computer cable practical for studio-length runs, which should all be unshielded, including both installed and flexible leads (equipment cords). For such runs it is commonplace to use the accompanying telephone-style RJ-45 connectors instead of XLRs. Current practice favors the use of pins 4 and 5 of RJ-45 connectors for the first AES3 pair signal, and pins 3 and 6 for a second AES3 pair signal. For runs in harsher environments, such as near transmitters where many interfering signals are present, specific AES3 shielded cable with 110-ohm characteristic impedance is available.

In video studios with lots of video wiring and switchers to consider, another form of the professional interface is available, called AES3id, so that digital audio signals may be cabled and routed using standard video techniques. Television stations use this 75-ohm unbalanced version of AES3, for instance. In this case, the same cable (RG-59/U) and connectors (BNC) that are used for video are also used. There is no way to distinguish input from output connectors because both are of the same gender, so you have to make certain by observation that outputs are connected to inputs.

For both AES3 and AES3id wiring, each cable carries a pair of audio signals, such as a left-right pair. The two signals are interleaved in time and decoded by receiving equipment into the original pair with the correct timing.

On consumer equipment a different version of the digital audio interface is used, called S/PDIF. The signals are similar but not identical to AES3 ones, and this can lead to difficulties; the electrical interface is different as well. Because the connectors for both inputs and outputs are of the same gender, there is also potential labeling confusion over what constitutes an output and what an input. Does "Tape Out" on a unit mean connect this signal to

the output of the tape machine? It would be easy to think this, but that would be wrong. Better labeling is, for example, “Out to Tape,” meaning to connect this output to the input of the tape machine; unfortunately, such explicit wording is infrequently found.

There are several ways for S/PDIF signals to be transported. With a wired electrical connection, the interface uses RCA phono pin plugs⁸ (Fig. 7-4). Phono jacks are found on CD players, for example, on both left and right



Figure 7-4a An RCA connector pair typically used for consumer analog audio (left channel white, right channel red).

⁸ Called each of these four in turn somewhere, so all of the terms have been used. Just one of them is usually sufficient to specify the connector in most cases. Hereinafter they will be called phono plugs or jacks.

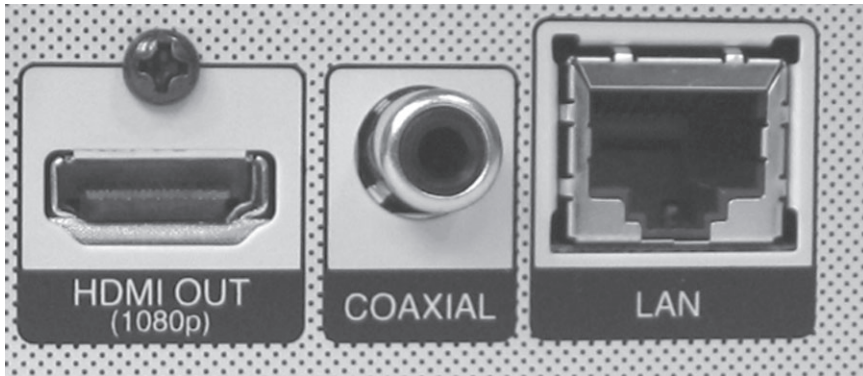


Figure 7-4b A single RCA connector may be used for S/PDIF digital audio transfer and is often labeled “digital coaxial” on consumer equipment.

analog outputs, and for the single digital output that carries two channels on one wire. There are two ways to distinguish between the analog and digital connectors. Usually the insulator of the left channel analog connector will be white and the right channel analog connector will be red, with the digital connector’s insulator some other color. Yellow is typically reserved for composite video signals. The other means of distinguishing analog from digital is the labeling: “Left” used alone as a designator implies analog because the digital interface carries two signals on one conductor.

Another way that S/PDIF format data may be conveyed is by optical fiber connection, and these are commonplace today (Fig. 7-5). The most common of them is called TOSLINK, and the associated cables and connectors are widely available. Optical fiber connections have the advantage of using light as the medium over a strand of glass inside the cable, thus requiring no electrical interconnection and preventing certain problems arising on the interfaces because of grounding difficulties, namely ground loops that can lead to hum.

One neat thing about digital audio connections is that if you have the same connector type on both ends of the cable, you just need to plug them in, and in all likelihood the interface will work. The audio levels are represented within the digits transmitted, not by the electrical voltage on the interface. Thus you merely need to make the connection and the audio levels will be right, which is one advantage of digital interconnections over analog ones.

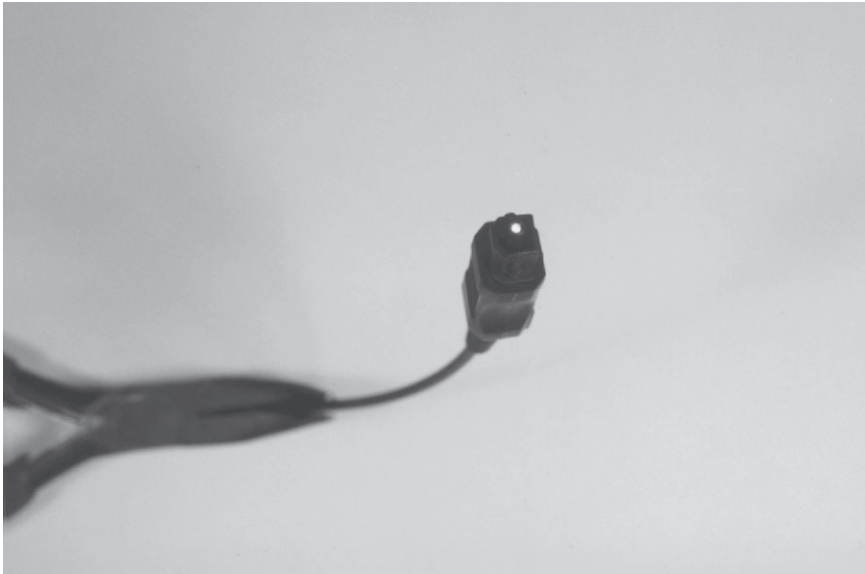


Figure 7-5 An optical fiber cable and connector.

If you face translating from one type to another, AES3 to S/PDIF or vice versa, it is possible to translate the electrical interface's level and impedance by means of transformers. However, even after correct electrical interconnection, AES3 signals are often not recognized as valid signals by S/PDIF inputs. If this connection does not work, you need a format converter from AES3 to S/PDIF.⁹ It is more likely that an S/PDIF signal will be recognized by an AES3 input, but this is also not guaranteed. Still, there is no harm in trying to connect the signals, once the level and impedance issues are out of the way, by means of an appropriate transformer.

Connection issues aside, there is another way you can get into trouble on a purely digital interface. Clicks or "snats" (loud, sharp, short clicks) can arise from clocking problems. Clocks for audio are like the quartz crystal oscillator in many watches, setting a pace for the internals of machines,

⁹ The following company makes transformers and format converters between AES and S/PDIF and vice versa: c4 Audio Systeme GmbH, Elmring 15, 38154 Koenigslutter, Germany. Telephone 0 53 53 94100 0 (From U.S. 011 49 53 53 94100 0), Fax 0 53 53 94100 88 (From U.S. 011 49 53 53 94100 88). http://www.c4audio.de/c4_Audio_Digitale_Interface_Pr/c4_audio_digitale_interface_produkte.html

and for the audio on their ports. But, just like different quartz watches, the clock in a DAT machine might be slightly different in speed than one in an editing system. When they are connected together, you can hardly notice the difference, and the small difference in rate that occurs can be mostly overcome by buffering—placing the incoming stream in a memory at the speed of the source machine, and removing it from the memory for use at the speed of the receiving machine. The problem comes when that memory becomes either under- or over-filled. At this point, the memory will have to jump to refill or remove data, and a snat results. This can occur as often as more than once a second (rarely) or as infrequently as once in, say, 20 minutes, and can thus be hard to track down.

There are two ways to cure this problem: (1) Operate the receiving equipment so that its clock is derived from the transmitted data, which means that it can't simultaneously be synchronized to the rest of the studio—for receiving data it must be operating as an island, divorced from the other studio equipment; or (2) The source equipment must be locked up to a studio master clock by connecting its word clock input to the same word clock that the editing system is using. A studio master clock is a piece of gear having a high-precision clock. Although it may have many types of outputs to cover both audio and video cases, the one most prominent for audio is called word clock, a reference signal that is clocked at the audio sampling rate, in most cases 48.000 kHz. Distributing this signal to all equipment that will accept a word clock input, and configuring that equipment through hardware and/or software switches to sync to their word clock input, overcomes the snat problem when sending digital audio between devices. By the way, it does not matter what type of digital audio connection—S/PDIF or AES3 on copper wire or optical—is used; snats may be a problem over any digital audio connection.

The remaining problem is with equipment that has no word clock input, a category that includes almost all CD players and inexpensive DAT machines. There are two possible solutions. The first is to make the receiving equipment follow the incoming signal as its clock source via a hardware and/or software switch. When set up this way, the receiving equipment—say an editing system—can't simultaneously produce a digital output that is in sync with the rest of the facility, but this may not be important. For instance, you would have to monitor the output sound by way of the editing system's analog out, rather than its digital out to other equipment, because the latter would then require the other equipment to operate at the original sending equipment's rate, and so forth. The second solution is

to use the converted analog output of the source machine, fed into an analog-to-digital converter on the receiving machine.

There is one more way that digital audio interfaces can be problematic, and it is even more fundamental: the source machine and the receiver must be operating at the same nominal sample frequency. If they are not, several consequences may occur. The first would be no sound at all as the receiver fails to lock up to the source. The second is that the receiver might be tricked into operating so far off speed that it accepts 44.1-kHz samples from a CD player, for instance, on a 48-kHz input, in which case, when the sound is subsequently played back at the normal rate for the editing system, the pitch will be raised by 8.8 percent, a large amount, and the time shortened by a corresponding amount. In some cases this may not matter, but it would usually represent an obvious distortion of the original audio element.

The third possibility is for the system to transfer the audio at 44.1 kHz, but tag it with that information, and then perform a sample rate conversion upon playback to get it to play at the correct pitch and speed in 48-kHz files (the sample rate always used for accompanying a picture). The problem is that very good sample rate conversion is taxing on computer hardware, limiting at the very least how many tracks this can be done on simultaneously. This is the method that Final Cut and Soundtrack Pro use in order to accommodate mixed sample rate files, such as 48-kHz DV files and 44.1-kHz CD files intermixed within one editing session, but at the very least it uses up considerable computer horsepower to do well.

Recommendations for streaming transfers:

Reviewing the requirements for digital audio interconnections to work well leads to the following recommendations:

- For FireWire interconnects, the first issue is that you need the correct connector and cable matched to the correct standard as shown for the task in Table 7.2.

Note that FireWire 400 ports may also be called iLink.

For FireWire interconnects, the main software issue is that the source tapes or files for a given program are all at the same sample rate

Table 7-2 FireWire interconnects.

1st Cable End	2nd Cable End	Used To	
FireWire 400 IEEE 1394a standard 4 pin	FireWire 400 IEEE 1394a standard 4 pin	Connect two DV camcorders together for copying or for continuous recording using certain models	
FireWire 400 IEEE 1394a standard 6 pin	FireWire 400 IEEE 1394a standard 4 pin	Connect DV camcorders to computers or to DV recorders, either having IEEE-1394 FireWire ports	
FireWire 400 IEEE 1394a standard 6 pin	FireWire 400 IEEE 1394a standard 6 pin	Connect FireWire hard drives to computer IEEE-1394a FireWire 400 ports	
FireWire 800 IEEE 1394b standard 9 pin	FireWire 400 IEEE 1394a standard 6 pin	Also known as a bilingual cable, it connects either a 6-pin FireWire 400 device to a 9-pin Firewire 800 port on a computer, or it connects the 6-pin FireWire 400 port on a computer to a 9-pin FireWire 800 device	
FireWire 800 IEEE 1394b standard 9 pin	FireWire 400 IEEE 1394a standard 4 pin	Connects a DV camcorder to a FireWire 800 port on a computer	

and word length, with a strong preference for two-channel, 48-kHz, 16-bit audio over lower bit rates or sample rates for DV footage. Violating this recommendation leads to additional work either by you (in transferring files to a common format) or by the workstation (which may slow it down and cause audio quality problems),

at least. Otherwise this is the most straightforward of the interfaces for digital video transfer, containing as it does both picture and sound. However, there may be synchronization issues to consider because of the use of unlocked audio in certain DV formats (explained in Chapter 2).

- For AES3 interconnects with XLR connectors on both ends, the same sample rate and word length issues apply to make inputs consistent with other files used within the editing system (48 kHz strongly preferred); but if only a 44.1-kHz source is available, such as from a CD, acknowledge this and provide for it within the editing environment or by other means. Snats occurring on this interface usually mean that there are clocking issues. If these are occurring, see if the source machine has a word clock input, and provide one from the same source used by the editing machine. If not, then set the editing import software so that it clocks from the input source, such as by selecting “Clock derived from AES input” or some such option in a clock menu.
- For AES3 interconnects with BNC connectors on both ends, the same considerations apply as for those on XLR connectors.
- For connections between AES3 balanced 110-ohm connectors (XLRs) and unbalanced 75-ohm connectors (BNCs), a transformer adapter needs to be used.
- For S/PDIF interconnects with phono connectors on both ends, the same issues as for AES3 interfaces occur. For CD players, the sample rate is on the 44.1-kHz standard, and thus sample rate conversion considerations apply, as explained previously. However, CDs are more quickly imported by file copying rather than by playing them on a player, so the need for this interconnection is far less common than it once was.
- For mixed AES and S/PDIF systems, an appropriate transformer may be used in some circumstances. You must confirm that the software will accept the relatively few bit differences between the two systems.

File Transfers

By far the more efficient means of moving audio and video around between media and systems is by file transfers rather than streaming transfers. That’s because whole sound files and video clips can be imported into an

Table 7-3 Audio file formats for professional recording and transfer.

Name	Where used	Limitations	Endian	Handling of more than one channel	Metadata
Sound Designer II (SDII), on PC with .sd2 file extension	Editing systems; very popular but with some limitations due to its age	Sample rate up to 48 kHz, 2GB file limit produces length limitation	Big, due to Motorola roots	Channels interleaved	Mono/stereo, bit depth, sample rate
AIFF (Audio Interchange File Format) (.aif, .aiff)	Editing systems; very popular but with some limitations due to its age		Big, due to Motorola roots, although variants in little	Channels interleaved	Number of channels* (mono/stereo are only interesting ones), bit depth, sample rate, application-specific data area
AIFF-C (Audio Interchange File Format for Compression) (.aifc)			Big	Channels interleaved	Mono/stereo, bit depth, sample rate, application-specific data area; -C indicates potential for bit-rate-reduced (compressed) audio
WAV (.wav) for Waveform Data	Started on Window but widely used today, even on Apple, for interchange		Little, due to Intel roots	Tracks interleaved	Number of channels, bit depth, sample rate; may contain bit-rate-reduced (compressed) audio

Continues

Table 7-3 *Continued.*

Broadcast WAV (.wav); despite being another format, .wav is the correct file extension	Broadcasting; type of WAV with extensions developed by broadcasters	No specific hooks for accompanying a picture	Little, due to WAV roots	WAV format file with added metadata about production and timing
OMF 1				First attempt at producing interchangeability with audio and EDL, particularly between picture and sound editing stations
OMF 2				Second attempt
AAF				Third attempt
AES31				An extension of broadcast WAV; an attempt at workstation interchange

*But the number of channels jumps from four to six, with the six being other than the standard 5.1, so AIFF is not useful for 5.1 channel files.

editing system in one action when copied off a random-access file system, and it takes place much faster than real time. Among the considerations is that the imported file format matches that of the system into which it is being imported. Let's say that you have DV or MPEG format video files with sync audio and you wish to import music from an audio CD to include in your project. The problem is the mixed sample rates: 48 kHz for DV and 44.1 kHz for CD. Several things may occur: One is that the session plays back the samples of the CD, but at the new sample rate, thus raising the pitch and shortening the duration. This may be all right for some sound effects, but it will make voices sound thinner and music playback in the wrong key, so the other possibility is sample rate conversion upon import. Sample rate conversion from 44.1 kHz to 48 kHz involves recalculating the shape of the original waveform in hardware or software and sampling it at the higher rate, which can be done automatically by many editing systems if configured properly.

The other concern in making file transfers is whether the system doing the importing understands the syntax of what is being delivered. There are many different file formats in use. Some contain only audio with minimal information about how it was recorded, whereas other file formats contain a great deal of information about where a given sound is to be used in a time line, what fade files are to act on it, and so forth. One primary difference among the file types is whether the audio format represents the least sensitive information first (called little endian, such as WAV files) or the most significant information first (called big endian, such as AIFF files). Some file formats for exchange are listed in Table 7-3.

Audio File Formats

Today's audio file formats are the result of years of development, with many systems owing quite a lot to older ones. In many lineages, each new format is really an extension of an older format, accomplished by placing new wrappers around the file recorded in the older format to add new information. In this way some audio file formats today bear a strong resemblance to Russian nesting dolls—opening each one reveals a new, smaller doll within. Some newer formats that wrap around older ones restrict the range of choices of the older ones. For instance, WAV files may contain low-bit-rate coded audio (see Chapter 11), but when used as an AES31 file, the coding is restricted to linear pulse code modulation (PCM).

One of the first of these formats was called IFF, developed for the Amiga computer. It begat WAV, which begat BWF, which begat AES31, each one adding a layer of additional information to the basic audio. Basic digital audio files in widespread interchange use are the formats AIFF-C, BWF, WAV, and SDII. There are also numerous file formats used for given proprietary systems. Some of the proprietary system file formats may be interchanged with each other (e.g., Akai machines can read Fairlight files).

A second type of file contains both audio media and editing instructions as well as additional metadata, or data about the data. Such formats include OMF 1, OMF 2, AAF, and AES31. Audio file formats that may be embedded within OMF, for instance, include AIFF-C, SDII, and WAV. However, OMF files may also be just edit decision lists with pointers to the actual audio; such OMF files are called composition only files. Watch out: more than one assistant has delivered a composition only file across town, only to find it does NOT include the audio itself.

Digital audio with a sample rate of 48 kHz using 16-bit linear PCM coding requires 5.76 Mbytes per track minute (48,000 samples per second times 16 bits per sample times 60 seconds per minute divided by 8 bits per byte equals 5,760,000 bytes). So 1 GB of digital audio storage holds 173 minutes of monaural 48 kHz sampled 16-bit audio. Other sample rates, word lengths, or number of channels can be scaled from this number.

Media are filled to various degrees depending on their use. For instance, while media for exchange and recording can be filled rather fully, media for editorial purposes must leave room for editing changes to the audio, additional editing files, and operational overhead. For editing purposes, 50 percent full is probably normal operation.

Managing Double-System Recordings

In double-system setups, production audio is recorded separately from video files using a dedicated audio recorder on set. A production field recorder typically records audio files in WAV or BWF (Broadcast WAV Format, see Table 7-3) onto one of the solid-state card types listed in Table 7-4, although optical, hard disk, and tape-based formats are also used by some devices as primary or secondary recording formats, and the most common of these are listed in the table as well. Examples of audio

Table 7-4 List of commonly used media for recording audio in double-system setups.

Media Type	Capacity (Typical)	Representative Audio Recorders	Audio Channels
Compact Flash (CF)	2 GB – 32 GB or more	Sound Devices 702, 722, 744, 788 Marantz PMD660, PMD671 Tascam HD-P2	2, 4, 6, 8, or more
Secure Digital (SD), SDHC	2 GB – 32 GB	Zoom H3N, H4 Tascam DR40, DR680 (8-track)	2 on handheld recorders, up to 8 on professional models
Internal Hard Disk	80 GB – 256 GB	Sound Devices 722, 744, 788	2, 4, 6, 8, or more
External Hard Disk (FireWire interface)	variable	Sound Devices 702, 722, 744, 788	2, 4, 6, 8, or more
Digital Audio Tape (DAT)*	15 minutes – 3 hours	Older recorders	2

* DAT recorders are no longer used as commonly as they were before non-linear recording systems took over the market, but may still be used with older equipment packages.

recorders used for capturing sound on a double-system production include the Sound Devices 700 series recorders for high-end productions, the Fostex FR-2 or FR-2LE, various offerings by Tascam and Marantz (which today record onto solid-state flash memory cards, although older models used tape or optical discs as the recording medium), and smaller devices suited for low-budget shooting such as the Zoom line of recorders and Sony's comparable handheld recorders.

Any of these recorders may be used with or without an outboard mixer as described under double-system recording setups in Chapter 3. Without a mixer, one or more microphones or receivers would be plugged directly into the recorder, usually via XLR inputs. With a mixer, the output of the mixer would typically be fed at line level into the recorder's XLR inputs or, in some cases, purpose-built multi-pin inputs. In either case, the sensitivity of the recorder inputs would have to be set to mic or line level accordingly.

A more recent phenomenon is the advent of the mixer-recorder, a device built with mixing features in mind but with recording capabilities built in. Most of the Sound Devices 700 series recorders have enough mixing controls that they could be used as mixer-recorders, although the Sound Devices 552 would be more representative of the mixer-recorder family of devices: built as a newer version of the four-channel 442 mixer, it adds a fifth input channel but also incorporates the ability to record the two-channel mix onto an SD card, which can be used instead of or in addition to the line-level feed sent out through the mixer outputs.

Audio media management for double-system recording is made somewhat easier than video media management by the fact that audio takes up less space on the medium than high-quality video files, so that less swapping out of media is necessary for recording audio. Even a 2 GB memory card (Compact Flash or SD) is enough to record over two and a half hours of 2-channel, 48 kHz, 16-bit audio, and larger memory cards can easily be used to achieve similar recording time with 24 bits per sample or with more channels of recorded audio. Devices designed around Compact Flash card recording may also allow for recording onto an internally mounted hard drive or an external hard drive, most commonly attached via FireWire interface. Simultaneously recording to Compact Flash card and hard disk can provide a backup copy of the day's recordings. Larger-capacity hard drives can also be used to archive all the recordings on a production rather than wiping the recording medium after transfer each day.

A second use of the FireWire bus on portable field recorders is to transfer audio files to a computer or editing system via file transfer. Unlike the use of the FireWire bus in streaming video transfers from DV format camcorders, using the FireWire connector on a recorder like the Sound Devices 702 to connect to a computer puts the audio recorder into disk mode: it masquerades as a simple external hard drive so that audio files can be copied off its internal media in a high-speed data transfer.¹⁰ Either of the device's internal media—Compact Flash card or internal hard drive—may be read in this manner, and this may be the only way of copying data off an internal hard drive short of unmounting the drive from the field recorder.

¹⁰ Other devices such as the Zoom H4 and Marantz PMD661 offer similar functionality via a USB connection.

Unlike video files, which are copied off memory cards as they are filled throughout the day, double-system audio recordings are usually transferred to the production drive and backed up at the end of each shooting day or, if necessary, at lunch. Since audio media doesn't fill up as quickly as video media, an entire day's worth of recordings will usually fit onto a single memory card or hard disk. When transferring recordings off a memory card, the card is usually removed from the recorder and read through a flash memory card reader attached to a computer, although using the FireWire or USB bus for file transfers as described above is possible on properly equipped recorders.

External hard drives in principle require no transfer of audio files since they can be disconnected from the recorder and plugged into an editing computer for import directly into the editing system, but in practice they are best kept off-line as the primary copy of the original recordings, and it is normal to transfer the files via computer to one or more backup hard drives at the end of each shooting day.

While increasingly uncommon, digital audio tape (DAT) may still be used as a consumer format for recording double-system audio. If used, digital tapes must be played back either on the recorder or on a dedicated play-back deck and transferred to the editing system in a streaming transfer.

Syncing Footage

There are many benefits to double-system recording, such as the quality of audio components available in dedicated audio recorders, the ability of sound and camera crews to move freely without being tethered to each other by cable, and the ease of recording additional sound on set for use in postproduction—including wild lines, presence, and sound effects, to name a few elements—without rolling camera. On the other hand, shooting double-system adds the requirement of syncing sound to picture in postproduction, a process which should be undertaken before any footage is edited together so that edited sequences are properly linked with the primary production sound recordings and don't have to be re-cut after sound and picture elements are married.

The traditional method of syncing sound to picture involves the use of a clapperboard slate, ubiquitous in film iconography, to create a sharp

clapping sound that can be put in hard sync with the visual closing of the sticks. However, other methods for syncing footage exist in the world of digital video. It is important, using any of these methods, to have a plan and to verify in advance that the chosen method will work with the equipment in use on a particular production. At any rate, a clapperboard slate should be used whenever possible as a backup in case the intended method of syncing footage fails unexpectedly; having a failsafe backup for sync reference can save hours of time in the editing room when something goes wrong.

On multi-camera productions, it becomes particularly important to have a method of syncing footage, since there will be multiple video elements to sync with each production audio recording and, in principle, with each other. If each camera has the capability of recording an audio guidetrack or jam syncing to a common time code, it may be possible to sync the video and audio recordings automatically in software. The following methods of syncing audio may be helpful to consider when setting up a workflow on a digital video production.

Time code Autosync

If your audio recorder has a time code output and you have a video camera that will jam sync to an external time code source, the easiest way to sync footage is to set both the audio recorder and the camera to free run (or time of day) time code and to jam sync the camera's time code generator to the audio recorder's, or vice versa. Then, leaving both devices powered on, you can typically shoot for several hours without the clocks going out of sync. Due to the potential for slight differences in oscillation time of the crystals in the two devices to build up to a noticeable time difference over the course of a day, it is a good idea to re-jam the timecode every few hours, such as once in the morning, after lunch, and again later in the afternoon. It may also be necessary to jam sync again after either the camera or the recorder is powered down and back on again, since time code frequently jumps by some amount during a power cycle.

After importing both video and audio footage into an editing program with autosync capability, such as Final Cut, Avid, or Premiere, the video and audio clips can be synchronized automatically using the matching free run time code embedded in each clip. Most likely, you will need

to select which audio recording to match with which video clip, but the software will slide the audio and video into synchronization and create a new video clip with production audio in sync. It is a good idea to watch through several clips synced in this manner and verify “by eye” that the audio and video are in hard sync, or, if a clapperboard slate has also been used, to scrub through the first few clips sunk using autosync and verify that the slate clap seen in the audio waveform coincides with the matching visual element on screen.

PluralEyes and DualEyes to Autosync

If time code sync is impossible, it may still be possible to autosync audio and video from one or more cameras if an audio guidetrack can be recorded on the cameras, say from an internal mic, during production. Recording a guidetrack was one of the few recommended uses of an internal or on-camera mic mentioned in the production sound discussion in Chapter 4. In addition to providing an emergency backup for syncing audio by hand (see “If a slate wasn’t used” below), a guidetrack recorded on the video medium can in many cases be used as a sync reference for software designed to compare audio data and sync recordings by matching audio mathematically. This technique is relatively modern compared to time code and manual syncing methods that have been in use for decades. It represents an emerging use of technology that promises to help reduce the workload involved in this area of digital video production, but, like any other method of achieving sync between audio and video, it should be tested in conditions similar to what will be found on set to make sure that it is a viable method for a particular project.

PluralEyes and DualEyes are two examples of software solutions designed to analyze the audio data of one or more video guidetracks and sync video files to each other or to a master audio recording, without relying on a user to manually find sync points for each clip as with traditional slating. DualEyes is designed to sync a single video file recorded by a single digital video camera to a production audio recording made on a separate recorder, whereas PluralEyes can sync footages from several cameras operating in close proximity in a multicam environment by matching up the guidetracks recorded by each camera and, in the case of double-system audio recording, can sync up the master production audio as well. In order for the software to work properly, it is necessary for the guidetrack to exhibit some degree of correlation with the master production audio recording.

This is generally achieved by ensuring that the camera and production microphone are in the same general vicinity during shooting and that they both pick up the same voices, effects, and ambient sounds.

One situation where this would certainly be impractical is in the case of a distant subject being recorded via radio mic on the production audio recorder, but who is too far away from the camera to be distinguished by the camera's on-board microphone. In other situations, there may also be too much ambient noise to pick up a quiet subject as cleanly with an on-camera mic as with a directional boom mic or lavalier mic, and if the difference between what is picked up on the guidetrack and what is heard on the primary recording is too great, this kind of autosync will not work as expected. The workaround is to run a cable from the double-system recorder to the camera and record the same audio on both systems; the primary production sound recording will probably be cleaner, but the camera recording ensures sync.

Syncing by Hand with a Clapperboard Slate

This is the traditional method of syncing footage, and whenever possible a slate should be used even in conjunction with time code syncing in case sync is lost between devices or the time code fails to sync up unexpectedly. When a clapperboard slate is used, the sticks are clapped at the beginning of each take to provide a common picture and sound reference that can be matched up in postproduction: a spike in the audio waveform corresponding to the loud, sharp clap of the sticks, and the corresponding video frame in which the sticks come together. The full procedure for using a slate in a double-system situation is given in the section on double-system recording in Chapter 3, and it is recommended that the steps outlined there be followed to minimize confusion and make manual syncing of footage as straightforward as possible during postproduction. In particular, the sticks should be clapped each time the camera is rolled, *and* each time sound is rolled in a case where the sound cuts and starts recording a new file while the camera has continued recording. The sticks must be clapped in frame and should be located on-axis relative to a directional microphone or near a body microphone on one of the actors so that they can be clearly seen and heard, and they should only be clapped once camera and sound are *both* rolling; otherwise, either picture or sound will not have a sync reference for a particular shot.

If a Slate Wasn't Used (What to Do in Case of Emergency)

If a slate wasn't used for one or more shots and no time code correlation exists between audio and video footage, the best way to proceed depends on whether a guidetrack was recorded alongside picture using an on-camera microphone. If a guidetrack is available, it may be possible to slide the primary production recording into sync with the video by visually examining the waveforms of each track in an editor that will show audio waveform data. Most prosumer and professional video editing software will display audio waveforms on audio tracks in the timeline. In Final Cut, this option is called "show audio waveform," while in Avid the "show sample plot" option must be selected. Expanding an audio track in Adobe Premiere Pro will also enable audio waveform display in the timeline. Sound editing software such as Pro Tools or Soundtrack Pro will show audio waveform data as well, so if an original audio clip must be synced up to edited video during sound editing, a guidetrack from the picture editor can be used in the same way. Once a production recording appears to be in sync with the guidetrack in waveform view, hard sync can be confirmed by playing back the two audio tracks simultaneously and listening for an "echo" or comb filtering artifacts (described in Chapter 9), sliding the production track forward or backward as necessary. It may be necessary to slip the production track by increments of less than a frame to get it in hard sync with the guidetrack.

In the absence of a slate, time code reference, and guidetrack, your best bet for syncing up footage is to slide the audio into approximate sync with the video and look for a short, identifiable syllable or sound effect to use to put the file into hard sync by eye. Frontal plosives like "T", "P", "D", and "B" are the best audiovisual cues for syncing dialogue by eye since they are usually loud and sharp and have a more visible sync reference (the actor's lips) than vowels and other consonants. Other visible events, such as hands clapping or an actor hitting a hard surface, can also provide hard sync references provided the corresponding sound can be found efficiently in the audio track; that's why it's important to slide the audio into near-sync first before finding a hard sync point. While it may be possible to sync audio in this way if no other sync reference exists, this is a slow and tedious process for any more than a handful of clips and should not be thought of as a viable option for syncing double-system footage in any but the most extreme circumstances.

Common Problems in Digital Audio File Transfers for Sound Accompanying Picture

Here are some of the common problems found with digital audio file exchanges, including those that come from picture editing systems:

File operations:

- Different software version of the system that generates the export data than expected between source and destination, or among source, translator software, and destination.
- Changing software version partway through a project.
- File naming conventions unintelligible or illogical. A file named “tempdub” conveys hardly any information because there are a great many of these in a production.
- Leaving file extensions off the end of file names, such as .wav.
- Improper formatting of export media, including fragmentation caused by not starting from an empty file structure on such media.
- OMF files transferred between different editions of an editor before export to the audio system.

Editorial operations:

- Illogical track layout, like sounds jumping among tracks such as interchanging boom and lavalier tracks at random. Consistency in naming and logging of tracks is necessary.
- Audio not in hard sync before export; it is difficult to see sync on low-resolution video. To check sync, it is useful to have a production A track that has been carefully checked as a reference. Dialogue may then be split across other tracks and their sync checked by listening and sliding the split tracks, playing the edited track and the A track simultaneously, and listening for phasing, a swishy sound resulting from the comb filter that occurs when sound is nearly in perfect sync.

There is software on the market that helps to autoconform sound to sound, such as Vocalign from Synchro Arts.

- Insertion of wild sounds without a means of tracing sync back to the source. This means the sound department has to do the work all over again if the same source is to be used. For instance, laying in music from a CD in the picture editor provides no means to trace back to the source. Instead, copying the CD to a time code recorder before insertion into the picture edit, then importing from the recorder with time code, provides a means to get precisely the same sync in sound editing. Note that any sound the picture editor uses is exported, but the sound editor may need a longer length of it, and this is where it is valuable to be able to repeat the picture editor's work in the sound editing room.
- Over-reliance on exporting systems has led to less accurate production of human-readable EDLs, but these are the only backup if the export fails.
- Text information on a sound clip.
- Subclips/group clips in multicamera shooting.
- Start times for a session do not match sequence time.
- Noisy editing rooms with bad monitoring leads to the picture department saying "It sounded all right on the Avid" in explaining a bad transfer. Picture editing suites are notorious for having bad monitoring conditions, and thus they are usually no place to judge audio quality. The story is told of completely distorted audio being received by a sound department that, when they went to check how it sounded in the editing room, found that the tweeters in the monitor speakers were burned out.

Digital audio problems:

- Different sample rates between source and destination. This is particularly a problem when music interests are involved because they would prefer a 44.1-kHz sample rate of the CD, yet most film and television operates at 48 kHz. Just importing CDs as sound effects into projects necessitates a sample rate conversion if they are to be imported with the same pitch and duration.

- Sample rate conflicts caused by pulldown, when film or genuine 24P video is in use. Shot with 48-kHz standard sample rate at 29.97 fps, then pulled down on the telecine, results in a non-standard rate. If audio has been inserted without pulldown, it will drift in sync by 0.1 percent, with the sound being longer than the picture. This can be checked by measuring the length and calculating the difference. If it is 0.1 percent, which is one frame in a thousand with the sound being longer, the likely source is the lack of a required pulldown.

Ordinary audio production problems:

- Level too low or high.
- Channels mixed together or interchanged.

Improper setup for export:

- Incorrect consolidation: in the case of Avid, leaving Audio Suite plug-ins activated: group information, pitch change, time compression/expansion, fades, levels.
- The sound editor needs *handles*, that is, sound for each region of audio used in the picture edit before the beginning and after the ending, providing the ability to make smoother changes than are usually done by the picture department. Handle lengths range from a few frames to the full length of the take, depending on the material, the density of edits on the tracks, and the desires of the sound editor. For long-form work,¹¹ handles are provided that are the length of the take, so the maximum chance is available to find presence that intercuts.

Improper or no labeling of the media:

- The Multichannel Postproduction Media Label in Chapter 11 suggests content for a label for interchange.

¹¹ This is the industry term for feature films and television miniseries.

One frame or more sync errors in exported files (probably originates in software mathematics) are common:

- It is useful to have a sync check such as a clapperboard slate once in an export file so that hard sync can be checked after import.

It is highly useful not to rely on a crucial transfer, but rather to test the transfer path ahead of the time when a transfer will be essential.

Director's Cut

- Make multiple backups of audio and video media and ensure that there is a workflow in place to mark media ready for reuse, such as memory cards that have been downloaded and backed up and are ready for reformatting in camera.
- When recording audio over a digital connection, make sure the playback and recording devices share a common clock source or that the recorder syncs its sample clock to the incoming data to avoid clicks or snats in the digital recording.
- File transfers from hard disks and memory cards used for recording are much faster than streaming transfers involving playback from the device, but streaming transfers may be required in the case of tape-based or other linear recording workflows.
- When recording audio double-system, always use a slate reference for each clip even if another method of syncing footage is to be used, such as time code autosync. That way, a backup sync reference will be available in case the primary means of syncing audio to video fails to work as expected.
- Ensure that material recorded on set and in the studio matches the audio file format and recording parameters in use for a given project whenever possible; if not, converting audio to the project sample rate, bit depth, and file format on import can save time and processing power down the line. Usually, audio accompanying picture is recorded at 48 kHz sampling rate and a bit depth of 16 or 24 bits per sample.

CHAPTER 8

Sound Design

Sound design is the art of getting the right sound in the right place at the right time. This chapter concentrates on defining what the right sound is. The right place can be defined as placement of a sound into the correct track out of many possible tracks, an organizational skill exercised during editing that anticipates and strongly affects mixing. The right time defines the placement of sounds in a time line, often with a synchronous relationship to the picture or to other sound effects. The right place and time have to do with the details of editing and are covered in the next chapter.

One concept that filmmaking borrowed from theater is called *mise en scene*, defined as what the director puts into the scene. The scene in film is the frame that we see—what is in it and what is excluded, the lighting, costumes, and blocking of the actors; that is, their massing and movement. When first applied to cinema, the concept meant simply everything in front of the camera: the set, costumes, composition of the shots, patterns of light and dark, and so forth. More recently, the term has been broadened within its use in film studies to include cinematography and its effects such as depth of field, as well as editing and sound. Thus sound stands in modern critical thinking as one principal ingredient of the film art. Reading a film means understanding the role that each of these ingredients plays in its perception, including sound.

The first notion of film sound influenced by the *mise en scene* idea is that somebody somewhere has decided what you hear, moment by moment, within a movie; as with any creative work, there is an author, and what you hear in a film is not accidental but planned. What you hear could be just the sound of what the camera sees in documentary fashion, but it is often a more constructed sound, built out of bits and pieces by professionals to simulate a desired overall sound impression. Sometimes, just to portray reality, sound may have to be changed from that which was captured during production. Off-screen sound recorded during production may

include inappropriate noise, in which case constructed or substituted sound is more appropriate. At other times, various film sound styles may dominate during different parts of a single film, extending from the real to the surreal to montage, for instance, across a range representing increasing levels of abstraction.

In most mainstream pictures, certain film sound conventions prevail that are perhaps stereotypical, but that also have the advantage of transmitting emotional information quickly. At the end of the climactic scene after the Ark of the Covenant has been closed in *Raiders of the Lost Ark*, Indy and Karen are left exhausted, but happy. The ambient sound we hear is that of crickets, a quiet and peaceful sound that ends a remarkable sonic sequence. The fact that all is at peace is transmitted quickly to the audience by the use of such stereotypical sound.

We come to understand the conventions of film sound through exposure to a great many examples; we are all trained through listening to films, whether we know it consciously or not. Thus a scene with a particular piquancy for us is that of lovers gazing into one another's eyes accompanied by orchestral score. The camera dollies back to reveal—an orchestra playing in the foreground. What could be more likely? The literalness of this sound joke is funny thanks to the confusion between whether the music heard is source music or score. It is a play on film sound conventions.¹ Another example occurs at the beginning of *Diva*, when we hear score over the titles, but the mail carrier shutting off the radio on his motorcycle kills the score. For this film, the fact that music is going to play a major role is highlighted by the effect of this character's action in the opening seconds.

One of the basic film sound conventions is that keeping sound constant across a picture cut implies that while we may have changed our point of view, we are nonetheless still in the same space. This job of improving overall continuity is most often handled by presence (also called *room tone*, or in England *atmosphere*). Presence is defined as sound that matches the sound captured during production so well that it may be intercut with production dialogue tracks without discontinuity. It is the sound of

¹ Good jokes have a way of being recycled. Mel Brooks used this joke in *Blazing Saddles* when the sheriff rides to his new job across the Western landscape, only to reveal in a pan the Count Basie Orchestra playing the score.

relative quiet for the space being portrayed. Note that it is not silence, because cutting to silence is instantly noticeable and leaves us in an auditory void as opposed to a quiet location. It may well be very, very quiet, though, and hopefully this will often be the case because we want the best dialogue intelligibility possible most of the time, and the fact that the background is quiet rather than loud helps intelligibility. Sometimes it is impossible to get precisely matching presence, and the result is a bump at an edit. Dialogue editors smooth dialogue using such tactics as cross-fading between recordings or hiding a sound edit by delaying it until the beginning of the next utterance after the corresponding picture cut.

If dialogue tracks can't be perfectly smoothed, or if the director and sound designer wish to create a new sonic space for the scene to exist in other than the one captured on set, ambience tracks, also known as backgrounds, can play a similar role (Fig. 8-1). They provide continuity across picture cuts, as well as substitution of a new space for the original. If presence in the production track is low enough in level, backgrounds may be able to cover up or mask discontinuities at edits, in addition to providing a story-telling element of their own. Think of *Star Trek*: Each part of the ship has its own background sound, and part of cutting from Engineering to the Bridge, or to 10 Forward, is the change in ambience that accompanies the cut. We as the audience are taught by the filmmakers what each space sounds like as well as looks like, so we can orient ourselves quickly in a scene change, using both auditory and visual cues.

By using dialogue intercut as necessary with presence, and potentially one or more ambience tracks, a realistic-sounding dialogue scene may be constructed whether or not it was shot in a single day or at a single location, even if the background sounds that would be expected for creating

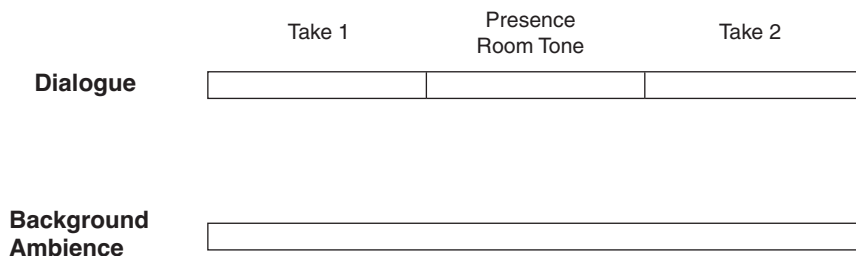


Figure 8-1 The difference between presence, which is intercut with dialogue tracks, and ambience, which exists on its own track(s).

verisimilitude were not actually present on set during shooting. The next element that is often used to help sell the reality of a scene is Foley. Named for the person who invented it at Universal,² Jack Foley, it is a series of sound effects made while watching the edited picture in a quiet, acoustically dead recording stage.³ The mic is typically operated close to the sound source. Recordings of footsteps, chair squeaks, clothing rustle, and small physical effects like pouring a drink, or large ones like a body hit or those of a dinosaur brushing up against foliage, may all be Foley effects. Given the close mic perspective and the quiet, dead room used to record it, Foley is most often a representation of exaggerated reality. The technique tends to hype the effects so that, heard alone, they seem bigger than life, but this is made necessary by the masking effect of having many other sounds playing simultaneously in the final sound mix. For Foley to read through the clutter of all those other sounds, it has to be recorded in this way. Even the lowest-budget films can afford Foley today, because the requirements for doing it well are fairly minimal: first, find a quiet absorbing space—for instance, a high-school stage with all the curtains let down to surround you. Then play the cut video over a picture monitor—possibly with headphones so you can listen to the production sound, although this is optional. With a microphone close to your feet, for example, proceed to record footsteps as necessary, placing them on another medium such as a production field recorder, computer audio interface, or even a separate camcorder. Repeat for other effects as needed.

One principal notion that applies to all types of sound effects is well illustrated by Foley, and that is the concept of layering. If you have a crowd walking on screen, you can get a crowd to walk in the Foley room and use multiple microphones to record them all at once, a technique I found in use at Beijing Film Studios; but if you are only one person, you can record sound that corresponds to one person in the shot, and do this repeatedly over several recordings, ultimately building up the sound of the whole scene. To accomplish this, it is useful to record each take to a separate track of an audio editing system such as Pro Tools or Soundtrack Pro so that each can be recorded in sync as you go. You may also want to have on hand a number of pairs of shoes! If you use a field recorder or other portable device to make your recordings, then every track has to be

² www.npr.org/programs/lnfsound/stories/000324.stories.html

³ The term *stage* comes from Hollywood practice: shooting stage, Foley stage, dubbing stage. In the outside world, this might be called a studio or recording studio.

transferred to an editing system and “sunk” up after the fact. Even for the multitrack recording where each track has been recorded in sync to the picture, getting hard sync often means moving Foley effects forward or backward in time by a small amount, but with this system, the process of getting each recording into near-sync is largely automated.

So far we have discussed dialogue, and with it intercut presence; ambience or backgrounds; and Foley effects. The next major category of sounds is cut or hard effects. These are sound effects, drawn from production sound, sound effects libraries, and original recordings made for the individual film, that typically bear a one-to-one relationship to picture. An anecdote illustrates this best. Some years ago, on the particularly tight postproduction schedule for *Tucker*, the re-recording mixers were spending long, long hours at work. The son of one of them had written his father a postcard that was promptly posted on the meter bridge at the top of the mixing console in the dubbing stage. It said, “Gee Dad, I get it—see a car, hear a car.” There could be no better definition of Hollywood postproduction sound cutting than that made by the young kid who missed his father. The Hollywood craft tradition says that anything we see that *can* make a sound, *will*.

Some directors accept the craft tradition that says that everything that can make sound will do so at face value, whereas others question it. A process often found in Hollywood is for sound editors to cut a large number of effects—basically everything they can think of—and then for the director to strip out whatever he or she finds unnecessary on the dub stage. This approach is exactly what sound design means to replace—the idea that decisions should be left for the end of the process and for the director to make. In *The Sum of All Fears*, for instance, an early scene in the film is a tension-ridden bomb shelter scene of the President of the United States commanding forces during a crisis. The phone rings and it’s his wife calling with trivialities, and we realize that this is not a real crisis: it’s only a drill. The spell is broken by the call. Sound editors provided all kinds of sounds, intended for use in both the screen channels and the surround ones, to embed the viewer in the scene and raise its tension level through engaging the listener. The director, on the other hand, found all of these communication chirps and whizzes distracting, particularly in the surrounds, so he stripped them out, and the scene played without most of them being heard.

For other types of filmmaking, especially where this process is carried out by just one person or by a very small crew, the selection of just what effects

to use can take place even before one gets to the mix. There is always some kind of limitation on the number of tracks available in an editing or mixing system. The number of tracks that is desirable seems always to be $n + 1$, where n is the number of tracks available. This begins with the simplest mixes with only a few tracks, and it continues through the largest mixes I have seen, with hundreds of tracks: “but if we could just have one more” is a familiar lament even in Hollywood studios. This is also a factor in sound design—keeping the cut tracks to the available range of equipment and software—and this problem is increasing today as editing processes take on what have traditionally been mixing processes. Plug-ins or filters that run on editing workstations take up some of the available computer horsepower, and it is often difficult to know when limits are being approached and what will happen once those limits are exceeded. Today the limitation is how many plug-ins per channel are allowed before the editing system can no longer run in real time.

Hard effects help illustrate and highlight certain truths about film sound. For instance, although supposedly realistic, those face punches in action films are anything but real for sound. I remember visiting a Hollywood recording stage some years ago on a studio tour. A large-fisted man was hitting a ham as hard as he could to get face punches. While this rather literal recording was better than the real thing, especially for the recipient, it’s really not enough for the larger-than-life punches of action-adventure movies. One thing to do is to slow down the original recording in an effort to make it appear bigger than life. Another is to layer together a whole parcel of sounds to make one enormous splat out of several merely exaggerated ones. Ingredients might include sounds of throwing an old leather jacket onto the metal hood of a fire truck,⁴ dropping overly ripe fruit onto concrete and slowing down the recording, or breaking a stalk of celery over one’s knee, to name a few examples.

One useful result of layering sounds is concealment of the original sound source used for a particular effect. Recognition is less likely when sounds are heard together simultaneously in time and panned to the same location. Additional concealment occurs with the pitch and time shifting that accompanies off-speed playback, which is often used with such effects to make them seem larger than life. As an aside, it is interesting to note that off-speed playback, particularly of an extreme nature, is much easier to

⁴ Ben Burtt at Skywalker Ranch for *Raiders of the Lost Ark*.

accomplish with analog recorders than digital ones—a reason in itself to keep one analog recorder around equipped with a wide-range varispeed in an otherwise all-digital studio.

What differentiates these kinds of sounds recorded for hard effects from Foley is that no attempt is made during recording to match a previously recorded picture in sync, but rather the object is to gather separately a whole set of ingredient parts that when blended together through editing and mixing take on a life of their own, becoming one auditory object, whether composed of a single ingredient or many layered together.

Walter Murch has stated an important concept in comparing sound design to a three-ring circus. Through evolution in time, big circuses settled on three rings, not two or five. How did they get there? It has to do with how much information can be processed by the brain at any one time. More than a certain amount of information leads to disconnection as the senses are overwhelmed. Yet the whole idea is to fully engage you, so growing to the point of diminishing returns is valuable too. Circuses evolved to the point where three things at once could be featured. Note that each of these three rings could contain many performers acting in concert, so there are many more molecular elements than three, but the molecular components are clustered into three groups, to which we can pay adequate, albeit potentially shifting, attention.

In our case the analogy has to do with being able to pay attention to what are called auditory streams. While we can only pay attention to three things at once, our attention may shift among more, because we can only hold so many in our heads at one time; but we can also shift gears among streams. Note that this means we definitely need more than three individual sound tracks, because a single auditory event may be built up from many different elements, but for any given point in time all of the elements clustered together into blocks form no more than three items we can typically follow or pay attention to.

Taking this idea a bit further, we might say that the three elements we can pay attention to are more easily delineated if they are of different types. Three dialogue tracks operating in overlapping fashion would be like three lion tamers in the three rings: confusing. But dialogue, Foley, sound effects, and music tracks, being rather thoroughly differentiated, are more likely to form separate auditory streams that we can follow.

Film Sound Styles

There is a range of styles used in film sound, and the style may shift within one film across the range through time. The most literal is the realistic style, which may wind up sounding as though it is just what occurred in front of the camera, but which may have to be built from multiple elements to sound that way. Surrealism breaks with realism by concentrating on unconscious, subjective sound often associated with a character's point of view. Montage goes even further in being the most abstract, with often a wide schism occurring between picture and sound.

Realism

For an ordinary dialogue scene, the sources are:

- *Dialogue from the original production sound.* Sound editors use both circled takes (the ones examined at dailies that have the potential of their picture winding up in the finished film) and outtakes (those that are likely never to have been seen by the picture editors) to find material from which to construct the dialogue tracks.
- *Wild lines.* Recorded typically the same day on the set, using the same microphone setup and so forth, but optimizing for noise in particular. For instance, if wind machines are necessary for a visual effect, they will be shut off and the dialogue repeated. Or if the boom mic cannot get close enough during shooting, the scene can be repeated for sound only.
- *Presence/room tone for above.* This is intercut as needed to remove unwanted utterances and otherwise smooth the track into sounding like a continuous whole. Today it is often captured from between utterances of the actors and looped to fill out the required length. However, it is also useful for the director to take a beat before saying "action," allowing the sound editor a moment where everything matches the shot and all is quiet.
- *Automated dialogue replacement (ADR).* Generally performed in quiet, dead spaces. Thus, it is typically used with an accompanying ambience track to produce a complete sounding picture. Used also for voice-over, which may have a tighter mic perspective (closer) to

sound as though the narrator is directly speaking to us from inside his head; occasionally this method of distinguishing a narrator from production sound (the narrator is less reverberant) is reversed, as in the soap opera when the camera sits on the unmoving face of a character with Vaseline on the lens and we hear them speak in a reverberant space: we know that we are inside their (big, empty) head. One benefit of ADR is being able to start with a clean slate for sound effects, ambiences, and music. Gary Rydstrom, who has done his share of them, says this is what makes animation so fun, and difficult, because no production sound exists.

- *Production effects (PFX)*. Sound effects recorded using sync sound on location, which may need separation editorially from dialogue because the processes employed during mixing may well be different between dialogue and effects.
- *Ambience/backgrounds*. These sounds form a space for the movie to live in. They may be used to set locale, period, threat level, and other exposition and emotional story-enhancing items. While presence serves to fill out production sound tracks so that they sound completely continuous, in the best of all possible worlds the presence will be rather quiet compared to the foreground dialogue and production effects, because production has taken place in a controlled environment. If this is so, then the sound designer has greater freedom, because dialogue intelligibility is already good. Then the sound designer can add ambience elements that enhance the story.

The specific track layout for these will be considered in the editing chapter. When cut and mixed, the attempt is to stick to the reality portrayed by the camera. As has been mentioned, in some instances use of sound from the location may be impossible, and ADR plus ambience might be substituted, but the point is the same—it is to sound as though it occurred in real time, in front of the camera.

A term used in video production (but not routine in film production) to enhance this real sound is *sweetening*. To add sweetness, other cut effects or music may be added to the smoothed track (smoothed by virtue of being filled correctly with presence dialogue tracks).

Music that is to appear as though it is present in the scene is nearly impossibly difficult to shoot in real time on location because, if it is heard at all,

it will proscribe many sound edits that might otherwise be made. Whenever possible, such source music, like playing from a radio within the scene, is faked and added in postproduction as a sweetener. About the only case where it cannot be faked is if a band shows up live in a shot, for instance, and even this can be controlled so that you collect all of the bits and pieces that make up the whole on location. You might wish to record:

- The band in a natural, moderately close perspective for the entire tune.
- The band in a distant perspective for the entire tune. This can be faked in postproduction as needed if sufficient postproduction tools such as reverberators are available, but if they aren't, the easy way to overcome their lack of availability is to make a separate recording, at the same tempo, energy level, and so forth, but from a distant perspective, to be available for intercutting or even cross-blending with the closer perspective sound.
- The foreground action, with the actors speaking their lines loudly enough to be heard over the band playing in the background.
- The foreground action with the actors speaking their lines at the same level as when the band is playing, but with the band miming their performance in the background.
- Of course, if the band doesn't show in a shot, then you don't need it; but be certain the actors are delivering their lines with sufficient energy that their timbre sounds as though it is being spoken over the loud background that will be built in postproduction.

Another common reality situation that needs careful consideration is shooting in a busy restaurant. In fiction filmmaking, the idea is to get the extras to *look* as though they are talking, but to keep them silent, or at least very, very quiet, so that the foreground dialogue (directed to have an appropriate energy level for the scene portrayed) is well recorded. In documentary situations, Table 6-1 shows that the general background noise level is just about the same as the dialogue level in a busy deli, so recording here is much more of a problem. Radio mics help a little by being so close to the actor, but there are still large requirements imposed on intelligibility and editing that are difficult or impossible to overcome. For instance, a quiet, secret communication in a busy deli is just about impossible to record in situ. It would have to be faked on some level, either

by controlling the original shooting or by constructing the sound from ADR and deli ambience tracks, possibly with some of the original sync tracks in instances where it can help sell the reality of the shot.

In order to construct a reality from parts, there are some principles used in recording and editing for sound effects. For moving objects such as cars, supervising sound editors ask sound effects recordists for several perspectives. Fundamentally these are:

- Start up from a stop.
- Stop from running.
- A steady. This is a constant perspective of a car running, for example. The microphone is “running” along with the car. It would be used when the camera follows along with the action, such as with a trailer shot. A steady could use either an interior or an exterior perspective.
- A pass-by. This is a drive past a fixed microphone position. It is an exterior perspective. Among the factors that constitute a pass-by is Doppler shift, the crowding together of sound waves that makes the apparent pitch go up as a sound is approaching you and the spreading apart of sound waves as the object recedes into the distance, lowering the pitch. The most obvious of all of these is on European police cars with their distinctive two-pitch sirens passing an observer’s location. If you hear such a sound on a sound track, you can associate the scene with European capitals immediately.

While this list describes four states of sound effects for one object, each of these could have many variations, with different speeds, squeals on various materials, and so on. Nonetheless, this is one way to think about what the microphone hears and how it will be used.

One factor in realism is the interaction between actors and sound effects. In the opening scene of *Raiders of the Lost Ark*, Indy collects the idol from the cave but has also set into motion the ancient mechanism to protect it. On his chase through the cave to get away from the machinery, at one point he looks up over his shoulder. At what? There is about to be a giant rolling boulder bearing down on him, but he doesn’t know this yet. What he looks at is a sound effect of the latch mechanism releasing the boulder, provided in postproduction. Steven Spielberg and Harrison Ford knew

that this was about to happen, and anticipation heightens the tension, so looking back over your shoulder makes sense, and then the question arises “what is he looking at?” and sound effects provide the answer.

Sometimes reality is just too strong for picture, and sound carries the day. This works in the sense that a radio play does, stimulating the imagination. *Fahrenheit 9/11* uses this technique when the airplane crashes into the twin towers. Because the pictures are very, very familiar, it is useful to go to black picture and tell the story only with sound for a period of time. This use of sound only can be compared to the picture technique of watching the reaction that people have to a horrible event rather than seeing the event itself, picturing the event in the mind’s eye. In fact this is what *Fahrenheit 9/11* follows the black sound-effects-only scene with, continuing and extending the thought.

It is always amazing how different a track sounds without picture; we pay a different kind of attention to it. *Fahrenheit 9/11* took advantage of this, but it is a fairly rare instance. On the other hand, an easy experiment is to listen to a section of a film with the picture off, which can reveal a universe of sound that most people don’t notice with the picture present.

The goal of using dialogue plus presence, ambience, and hard effects, including potentially source music,⁵ is usually to portray reality, but, as can be seen, this is actually a lot more complicated than it first looks. I (Holman) remember doing lighting for a play in which the director asked for just one spotlight on the single actor for the whole play. I thought about what the director was saying, which was that there was not to be a lot of self-conscious lighting, with light cues varying, and so forth, to detract from the intensity of the one-actor experience. I lit the scene with three basic sources, a spotlight on the actor, a strong backlight that set him off from the background, and several soft background lights. I got the effect of one light as you would perceive it if you weren’t consciously thinking about it, but it actually took several lights to do it. These were ganged together for dimming, so they went up at the beginning and down at the end together, and most of the audience was probably never the wiser that there was more than one light source, but the actor’s head was separated from the background better and the set illuminated dimly, all through the use of more lights than the director had called for: one light might have

⁵ Called *diegetic* in critical studies.

been literally what he asked for, but the effect of a single light source and simplicity was his goal, and that was better achieved with more than one light.

Sound tracks are a bit like this: sound editors may well produce a lot of tracks that, when finally perceived, add up to a whole that we take in as one. The idea that multiple sound tracks treated similarly tend to cluster together as one, such as all fading together, also proves to be useful to form auditory streams from ingredient parts. The island part of *Castaway* is a good example of creating the illusion of reality, how hard it can be, and how effective.

Stretched Reality

Sometimes it is necessary in an otherwise realistic presentation to hear things that we could never hear if we were present at the camera in real life. In particular, this applies to dialogue. An outrageous example is the opening of *American Pie 2*, which shows the exterior of the protagonist's frat house. As the camera dollies in, indicating we are being taken inside, we hear the conversation inside, and then we cut inside in continuous time. It is physically absurd to hear this in the same perspective outside that we will once we are inside, but it is an accepted convention to hear dialogue that we couldn't in ordinary circumstance. This is movie reality.

Another example comes from the history of telephone calls in the movies. In the earliest sound films, when an actor talked on a telephone we couldn't hear the other end of the call. So the dialogue had to be forced, with the recipient repeating all of the important content for the audience to hear. Soon the telephone filter was developed, and we could hear both ends of the call, with the end off-camera portrayed through the bandwidth-limiting filter that made it sound as though it was coming through a telephone. Already by the early 1930s such conventions were accepted as being realistic, when in fact you wouldn't hear the other end of the conversation in real life. The reason was to prevent the clumsiness of having to repeat what the other person said. This convention ruled for many years, and it grew a bit more complicated as editing got quicker; whenever a caller was off-screen they had to be filtered, but then you have to cut the filter in and out corresponding to what you are seeing, and, if you should just happen

to go to a split screen in the middle of the call to show both ends, no filter is appropriate.

In more recent times, the convention of filtering the opposite end of phone calls has been changing. Filtering takes away the presence of an actor, and if, for instance, a threat is being issued, then somehow it is lessened through filtering. While we need the filter to establish that it is a phone call in the beginning, gradually, over the call or multiple calls, the filter can be made less significant (by extending the frequency range). This applies to the threatening phone calls of the blackmailer played by Goran Visnjic as he speaks in multiple calls to the character played by Tilda Swinton in *The Deep End*, for instance.⁶

Another example is in *Vertical Limit*. Here mountain climbers have gone beyond the limit of their oxygen, and they know they are going to die because help cannot arrive in time. One has a conversation on a walkie-talkie with her brother down the mountain.⁷ At the beginning of the conversation, the sound is filtered, or possibly re-recorded through an actual walkie-talkie, but as time progresses this effect is gradually lifted. The camera also becomes more intimate with the protagonists. We come to a point where the brother is right there in the cave with the doomed climber despite being miles away, because the voice is heard full range. The full-frequency range voice is much more intimate sounding, especially than the low-fi walkie-talkie. We are hearing the reconstruction of the voice of the brother from the point of view of his sister, rather than the real thing. It is an interesting effect and the imposition of a kind of hyper-reality that works emotionally at this climax of the picture.

What is Seen versus What is Heard: On-Screen versus Off-Screen

Sometimes the effect of not seeing something is more horrible than expressly seeing it. Colonel Turner, the traitor who is exposed on the flight of the assault team away from the castle at the end of *Where Eagles Dare*, is given the chance by the lead of the assault team played by Richard Burton to commit suicide by jumping out of the plane, without benefit of

⁶ Supervising re-recording mixer Berger.

⁷ Columbia Pictures DVD 05066, Chapter 23.

parachute. The scene is played out such that Turner stands up to go to the door, and we cut to Richard Burton and Clint Eastwood. What we hear is the sound effect of the door latch opening. We don't need to see the body tumbling from the plane and landing splat on the ground; that's for rougher pictures. It is the sound effect that is so horrible in context. And one could say that this sound effect is the climax of the film, because the whole point of the raid was to expose leaks from the British side (which, by the way, didn't happen, but this wasn't known when the picture was written).

There are doubtlessly many such citations. Two that spring to mind are the sound of the dragoons attacking the marchers in *Dr Zhivago*, as played out in Omar Sharif's (as Zhivago) reaction to the event, rather than seeing the event itself. Another is that Jimmy Stewart alerts the city to the fact that a murder has been committed at the end of Hitchcock's *Rope* by firing a gun out the window, and we hear the off-stage sounds grow as people react to the shots, call the police, and then we hear the police arriving as the world of the killers collapses around them.

Hyper-reality

If stretched reality can be said to be pushing the limits of the real, Hollywood sound tracks for blockbuster action films go even further, beyond all possibility of the real to the hyper-real. There are a great many examples of this concept, available especially every summer in the local multiplex. Perhaps one way to describe these effects is with an inverse example. Gary Rydstrom recounts a conversation with Steven Spielberg about the sound for *Saving Private Ryan*. Spielberg said he didn't want a "Hollywood sound job," and Rydstrom knew exactly what that meant: he strove not for hyper-realism but for realism tempered with psychological sound. That is, although we hear the pitched battle realistically, including from an underwater as well as a more normal perspective, we also hear some of the sound from the point of view of the characters on the Normandy beaches, whose hearing has been affected by the extraordinarily loud sound around them. In speaking to veterans, Rydstrom learned that it was the particular sound that the German tanks made that they remember to this day, and because no working Panzer tanks could be found, he relied on the memory of the veterans to make a German tank sound that was realistic for them.

Surrealism

The next style to consider is the surreal, emphasizing the unconscious. Surreal sound might represent the point of view (hearing?) of a character as we slip from a realistic presentation into a more subjective one. There are several ways to do this: one is to strip out much of the sound and concentrate on just one or a few elements. We know the scene is noisier than this, but we also know we are making a transition to a more internal view of the world as sound disappears. It becomes more like our perception, following just a particular auditory stream. A good example is the opening scene of script reading in *All That Jazz*. The director is about to have a heart attack, and we cross over from a realistic presentation to a subjective one by stripping away most of the sync sound progressively and hearing just one or two Foley sounds, such as drumming his fingers on the desktop. To get back out of this state, the ambience corresponding to the reality sound is brought back in.

Dream states may be portrayed similarly, characterized generally by simplicity. We don't perceive ambience in dreams, and the sounds that we do hear may undergo transformations such as being in a different pitch range, made more or less reverberant, or changed in timbre. Any of these used singly or together is an indicator that we have left reality behind. When we subsequently cut to the face of our protagonist in bed, and we hear distant city ambience, we know through the reintroduction of the ambience that we've been jerked back into reality, and what preceded is just a dream.

Another way to lessen the reality and go toward a more outright filmic viewpoint is to add music. Here the big question is "what is the balance between music and realistic sound?" For instance, in the sequence of the biplane flying through flamingoes in *Out of Africa*, music is dominant. What makes the scene not all-out montage is the fact that we still hear the biplane, albeit nearly buried in music. The small sound of the distant biplane anchors us with the image and keeps it real.

In *the Bourne Supremacy* the title character suffers from amnesia. To get into his head—his subjective reality—a different type of surreal sound effects with more tonal sounds than normal is used in particular scenes, and one of its accompanying methods is to place more sound in the surrounds than would be there for a realistic presentation.

An example of manipulating realism versus surrealism or objective versus subjective sound is in the 2002 version of *Solaris*.⁸ When “Chris Kelvin,” played by George Clooney, makes preparations for going to sleep for the first time, we hear the Foley of him climbing into bed and music. As he goes to sleep, we see a shot of him in close-up, and the music comes to dominate. Interestingly, there is no ambience of the space ship at this point. Then the music is continuous over flashbacks of his life on Earth. This music, and the subjective handheld camera used as opposed to the more formal tripod-mounted one that has dominated in the previous scenes, tells us this is his dream. Also, the Earth scenes are warmly lit, while the space ship scenes are cold. Then, while remaining in subjective space, Chris meets the woman who will become his wife at a cocktail party. The reason we know this is still his dream is that the music dominates until the middle of a warm, handheld shot, and ambience and his clearing his throat breaks the music and causes it to fade out. The continuity that the music has provided over these preceding shots has let us know that it is all a dream, and the fact that the music remains over a shot and then is broken by an utterance of an actor helps us know we are still in his dream. He has a conversation with a colleague, and we hear it and cocktail crowd ambience (interestingly, only in center-channel mono). Along with some exposition, the colleague tells him to go ahead and deal with the woman he met earlier as he saw her on a train. He does, and during their conversation we cut back to him sleeping in the cool color of the space ship, in order to reinforce the fact that it is a dream. The conversation continues unchecked underneath this cutaway. Music begins under their continuous conversation, but an elevator shot of them together is shown for the first time, clearly just after the conversation has taken place. Thus at this moment there are three intermixed time frames at work, and yet they remain quite clear: the sleeping Chris in the space ship, the flashback to Chris and Rheya’s first meeting conversation as what is heard, and the flashback to their first touch as what is seen.

The music then comes to dominate once again, as they make love, intercut between the cool-color space ship (but wait a minute, where did *this* Rheya come from?) and the warm flashback, the past and the present. Finally, to come out of the dream state, we see a shot of Chris alone in close-up, and the music fades under the ambience, as though the music has been

⁸ Twentieth Century Fox widescreen edition DVD. Supervising sound editor and re-recording mixer Larry Blake. Sequence starts at 20:48 in chapter 8.

masking the ambience all along and we are just becoming aware of it, much like waking up from a dream. Interestingly, the music just disappears as her hand enters the shot: the abstraction is broken; we are back to reality. He is startled by her presence (after all, she's dead in real reality—it's just this particular reality that she's alive for), and this is a doppelganger. The sound is ambience, Foley of him moving rapidly, and his dialogue "God damn it" accompanied by heavy breathing: he's had a shock. Next he slaps himself awake. The surrealistic dream mood is thoroughly broken.

Interesting is that this new reality has more prominent ambience than the world left behind at the beginning of the sequence. The function of the ambience is to sell us that we are in the here and now. When he first goes into the bedroom and shuts—and on the advice of a crew member locks—the door, the music starts immediately and dominates the ambience right away.

Montage

If music completely overwhelms the other tracks, we move into the region of pure montage, typically represented by multiple picture cuts over continuous music. Here we are in the land of the most abstract. Montage can provide a little relief from the dialogue that delivers the ongoing story—it is the pause that refreshes. It often has a narrative role, moving the story forward by giving us information that otherwise would take yet more dialogue to convey. An example is that of Susan Alexander, Kane's love interest in *Citizen Kane*, singing as the camera cranes up to the top of the opera house and comes to rest on the skeptical stage crew. This is good storytelling because it shows us directly and through companion shots that she's on the road and always reviewed well by Kane newspapers, but she really can't sing, all in one brilliant sequence.

However, it is usually difficult to make a montage sequence as emotionally involving as those scenes where we have more direct connection to characters. This is because the nature of the abstraction of montage usually distances us from having empathy for the plight of the characters. There are always exceptions to the rule, such as one of the most famous of montage scenes, the Odessa steps sequence from Eisenstein's *Battleship Potemkin*, where the baby carriage gets loose and bounces down the steps.

Here we have a universal emotion built into our species of protecting babies that makes it work.

Sound montage, as opposed to filmic montage, although more frequently performed in other forms such as radio plays, is nonetheless useful for certain film sequences too. Sound montage is characterized by multiple sound tracks and threads coming and going, usually cross-fading among them. A technique to distinguish the multiple tracks of montage besides simply level is to change the relative balance between direct sound and reverberation, with a higher level of reverberation implying a greater distance, so that sound can seem to approach and then recede from the listener. Two examples of the use of sound montage occur in *THX 1138* and *The Conversation*.

A remarkable thing about montage is how often you find that picture and sound mutually reinforce each other, sometimes without actually doing anything to manipulate them together. Music has a beat, a downbeat, and other items regarding its phrasing. When a downbeat occurs on a picture cut, when a walker just happens to synchronize with the beat, when a key change accompanies a scene change, these are all items that resonate with viewers/listeners because they just seem to make sense of the abstraction in front of them.

In *The Wonderful Horrible Life of Leni Riefenstahl*, we see the title character at the editing table looking at the picture and listening to the sound of music during the Nuremberg Nazi rally. She glorifies in the sync that she produced between the waving of hundreds of Nazi flags and the music. She, as the filmmaker, knew very well that such things, which remain subconscious to most viewers, nonetheless help move the viewer emotionally.

Shifting Levels of Reality

A common opening for a movie is to start with scored music over titles, a montage. Such sequences function at a high level of abstraction, with a kind of psychic distance maintained from the audience—we are *reading* titles, not particularly following the background action, which can nonetheless serve to set the period, locale, and perhaps something about the characters, but on a rather abstract level. *Days of Heaven* starts this way, with old photos that set the era—Woodrow Wilson is President—and the

setting of an Eastern industrial city with which the West of the movie will be contrasted. After all, the filmmakers want us to be engaged, but at the same time, read the titles. When the main narrative starts, we are signaled this by the music coming to an end, and usually ambience of the first scene takes over on the sound track. A not infrequent alternative is for the music to fade under a voice-over, which also helps set the stage, such as is used in *Love Actually*. By either means we are eased into the story. Then we hear the real narrative begin, and the signal for this is production dialogue. In hearing this we have moved from a greater level of abstraction—montage and voice-over—to a greater level of reality.

Likewise, at the end of a movie rounding out the final denouement, typically the music score swells as the camera cranes up, a signal that the narrative is over. *Road to Perdition* has multiple endings, and the entrance of score on the first ending created the deliberate false impression that “okay, okay, thaaaaaat’s all folks,” and some audience members in the audience that I was in had been cued by the score that the movie was over and got up and left! Of course they weren’t remembering that the bad guy had not been called off the trail, and that this is a tragedy and shouldn’t have a happy ending. Those who left when the score told them to missed the real ending.

A movie opening sometimes used is to start with score and then modulate the score from sounding like score to sounding like source music originating from within the scene. This may be accomplished by worldizing the score and cutting it into different tracks in sync with itself. To make the transition, all the re-recording mixer has to do is to cross-fade from the score track(s) to the source track(s).

Worldizing is the process of bringing the world to the sound, by re-recording it through an acoustic space that deliberately modifies the sound. The idea is that clean sound is not always desirable; sometimes we want the sound dirtied up to make it sound more like the real thing. For *American Graffiti*, Walter Murch wanted the pop songs to sound as though they originated in a gym, not like needle-drop score. To do this, a loudspeaker and microphone were arranged to re-record the songs in a gym. Normally this would just add reverberation and move the perspective farther away. However, the difference in this case was that the loudspeaker and microphone were moved around continuously while the music was playing. This constantly changing path between loudspeaker and microphone made the sort of swirling sound that was right to accompany dances at a high school gym.

Interestingly, worldizing only became necessary when the channel from microphone to listener became so clear that sometimes it was too clear. The restored opening of Orson Welles's *Touch of Evil* as recut following Welles's instructions by Walter Murch employs worldizing. This was a major point of contention with the studio, which wanted a conventional score over opening titles. What Welles fights for in his memo is the following:

As the camera roves through the streets of the Mexican border town, the plan was to feature a succession of different and contrasting Latin American musical numbers—the effect, that is, of our passing one cabaret orchestra after another. In honky-tonk districts on the border, loudspeakers are over the entrance of every joint, large or small, each blasting out its own tune by way of a “come on” or “pitch” for the tourists. The fact that the streets are invariably loud with this music was planned as a basic device throughout the entire picture. The special use of contrasting “mambo-type” rhythm numbers with rock ‘n’ roll will be developed in some detail at the end of this memo, when I’ll take up details of the “beat” and also specifics of musical color and instrumentation on a scene-by-scene, and transition-by-transition basis.

In the version I was shown yesterday, it’s not clear where you have decided to place the credits. A brief report on this will determine whether or not my old ideas for sound and music patterns in this opening reel are still of some value.⁹

A process parallel to worldizing is futzing. To futz a sound is to deliberately screw up its frequency range, response, and distortion so that it sounds as though it is coming through a particularly bad channel. In early days of film sound, a telephone filter restricted the frequency range for telephonic voices, and this was enough. In today’s world, with much higher fidelity, somehow just restricting the frequency range to the bandwidth of a telephone is not enough; it’s got to be more screwed up to sound right. Here a box lined with absorption has a particularly bad small loudspeaker hung in it, and the output of it is picked up with a microphone. By overdriving the loudspeaker, or its amplifier, and by manipulating the signal path between speaker and microphone (direct, off-axis, around a barrier, muffled, and so forth), a variety of futz effects can be accomplished. An example of dialogue futzing is in the trench run at the end of *Star Wars*. In this case, side-band processing (i.e., transmitting and receiving the signal over a somewhat mistuned radio system) was also used.

⁹ Orson Welles’s memo to the studio, as seen on the bonus section of the widescreen DVD of *Touch of Evil*. There is further sound information in the memo, showing just how important Welles considered it. At this writing it is at http://wellesnet.com/touch_memo1.htm as well.

Sound Design as an Art

There can be no one prescription for what the sound elements should be and the interaction among them, because if there were, we could quantify art and then a computer could practice it. Instead, it is the funny juxtapositions and experiments that lead to a great sound track. There are some ideas that are fairly universal, though:

- *Emotional memory plays a role.* Here the most obvious is certain pieces of existing music. Music carries with it emotional weight, as when for instance nostalgia for an era plays a part. However, it is dangerous too, because it imposes its own order: beat, downbeat, key changes, and so forth. And the riskiest part of using existing music, especially of the pop and certain other genres, is that people have associations with it that are not under the control of the filmmaker: not everyone feels that “Leavin’ on a Jet Plane” means the same thing to them, and it may trigger unexpected emotions among a part of your audience, which means they are not paying attention to your story.
- *Sound effects with lots of low-frequency content tend to be threatening.* This can be said to be primordial for humans, as a thunderstorm or an earthquake are the natural events that contain a lot of low-frequency content. An early example of sound design like this is for *Butch Cassidy and the Sundance Kid*, where the off-screen and distant pounding hooves of the horses on their trail provide an intermittent threat to our rascally heroes.¹⁰ It is worth pointing out that this subjective equation has been enhanced in recent years as 5.1 channel sound with its dedicated low-frequency enhancement (0.1) channel permits greater levels of low frequencies.
- *Exaggerating reality may prove useful.* In the jungle scene in *Apocalypse Now*, the sound of insects is exaggerated and powerful, forming almost a musical score, as the threat level increases throughout the scene with the extreme, processed buzzing of mosquitoes before the characters come across the tiger. The particular sound raises tension simultaneously through its familiarity and its exaggeration—it’s like having a mosquito in your ear.

¹⁰ As edited by Don Hall.

If anything can be said about how sound design is practiced at the highest levels, it should be pointed out that it is the result of a lot of experimentation. This idea overarches recording, editing, and mixing. Hundreds of everyday objects are found by sound designers to make useful noises that, when synchronized to particular things on screen, sell the reality of a shot, even if that reality is far-fetched. The inventiveness of sound people in this search is remarkable, and good ones are always on the alert for interesting sounds. For instance, Ben Burtt went on a trip to an airbase to record jet takeoffs and landings. The most interesting sound encountered on the trip was that of the stalled air-conditioner motor in a cheap motel that, pitched down, got used for the Star Destroyer in *Star Wars*. You never know.

The motor sounds for the robots in *I, Robot* were the result of a lot of experimentation on the part of sound designer Erik Aadahl and of feedback from the director in diagnosing what he did and didn't like. During the months of recording, processing, and presenting sounds to the director, some of the rejected options included pitched-down hummingbird vocals, steel sheet wobbles, processed pneumatic air hisses, and cable "zuzz" sounds. "The final NS-5 robot sound was a combination of varispeed underwater bubbles, completely synthetic sounds designed on a 'Kyma' system, and telephone cord squeezes for closeups of the fibrous robot muscles moving," says Aadahl.¹¹

The dinosaurs in *Jurassic Park* came about by experimenting with many different animal sounds, layering them together and matching them to the screen action. One principal ingredient was that of a baby elephant trumpeting. The elephant did it only once despite multiple visits to record him, and one recording was extended into making this signature sound for multiple uses in the film by placing other sounds around it, editing it judiciously, and pitch shifting it through off-speed playback so that it could be reused without such strong recognition that it was in fact the same effect used over and over.

Spotting

Spotting refers to the director, the person in charge of the sound, and possibly the composer and/or others working their way through the

¹¹ Erik Aadahl, private communication, 2004.

edited piece and exchanging information and notes about what the meaning of each scene is in a film, and how to best supplement, or even directly tell, the story with sound. The term spotting may also refer to meetings in which the requirements are passed out to each subspeciality, such as at a meeting between the supervising sound editor and the Foley person.

A well-known sound designer tells the story of working with a famous director on several pictures. As they were spotting their first work together, the director would say “oh, there’s a car by,” and the sound designer would respond “okay.” After the literalness of this wore off, and they came to know each other better, the sound designer said to the director, “I can see the obvious—tell me what emotion you want the audience to feel?” On their next picture, the director asked, “Could you provide a morality tone here?” As Dennis Hopper’s photographer character says near the end of *Apocalypse Now* in explaining the shrunken human heads on posts, “Sometimes he goes too far.”

Director’s Cut

- Sound design means, among other things, that what you hear moment-by-moment is chosen deliberately by an agent, even for a nonfiction piece.
- Certain sound stereotypes transmit information quickly, such as the use of low frequencies equals a threat or a soft cricket indicates sound peace.
- Smoothing tracks, especially dialogue ones, helps overall continuity. Intercut presence and separate ambience or background tracks help provide continuity.
- Things that we see that we associate with making sound will normally make that sound. The source for this may be production sound, or it may be deliberately added sound effects, called hard effects.
- Film sound styles range from realism through stretched reality, hyper-reality, to surrealism. These are each portrayed using certain types of sound effects and/or music and even more rarely processed dialogue.
- Off-screen sound may have a strong storytelling utility.
- Sound effects of moving objects include start-ups, stops, steadies, and pass-bys.

CHAPTER 9

Editing

Computer-based workstations, either those based on picture editing with a sound component or those specifically designed for audio editing called digital audio workstations (DAWs), have revolutionized sound editing for movies and videos over the last two decades. There are several differences between analog and digital approaches that make the productivity of digital audio editing much higher compared to razor-blade editing of audio tape or film.

Non-Linear Editing

The first of these is the non-linear nature of computer-based editing. The term *non-linear* in this case refers to the way that the main storage mechanism, the hard disk drive in most cases, combined with the editing system software, basically works. It works rather like a phonograph record with a tone arm that you can pick up and place at the beginning of a groove anywhere on the record, quickly, as each piece of sound is needed. The editing system keeps track of where each portion of audio exists on the disk. It then plays them back in real time by skipping the “tone arm” around as necessary, filling up the immediate storage mechanism, the random access memory (RAM) of the computer, from the bulk storage mechanism, the hard drive, before the sound is needed. Then the edit system pulls the digital audio from the RAM, processes and routes it appropriately, and presents it to a digital-to-analog converter (DAC) for output conversion to analog, at the right moment. This is in contrast to linear-access media, like tape, where you have to fast forward, potentially past lots of intermediate tape, to reach a random position on the tape. Older conventional analog systems were linear access; computer-based ones where the audio is stored on disk drives are non-linear.

This description that involves hard disk drives and computer RAM shows the advantage of fast, random access and also gives a feel for some of the potential limitations of the process. The size of the hard disk drive(s) available on the machine will determine the amount of program material available, considering also its format (mono, stereo, or multichannel; sample rate; and word length). Roughly speaking, the available RAM will determine how many tracks can be played simultaneously. Another potential bottleneck is how fast the computer can move content from the hard drives to the RAM and from the RAM to the processor and out to the DAC. Moving a project from one computer system to another can reveal such problems: edit sessions that will play on one system will not play on another, for instance, because the resources available on one system may be so much greater than those on another that the processes bog down on the lesser system.

A typical digital video recording takes up around 25 Mbps (which is 3.125 MBps), the precise bit rate for standard definition DV format picture and sound. Newer, high-definition codecs such as AVCHD can record HD picture and sound using similar data rates. (Note the capitalization difference in the units: the abbreviation Mbps is millions of bits per second; MBps is megabytes per second, and the two are related by dividing the number of bits by eight to get the number of bytes.) That's in excess of 10 GB per hour. While storing this amount of data seemed impractical several years ago, today it's easy; at this writing, somewhere on the order of 100 hours can be stored for around \$100, and this number has improved dramatically over the years. Sound without picture requires much less space, with two channels of 48 kHz audio requiring less than 7 percent of the data space quoted above. Two-channel, 48-kHz sample rate, 16-bit audio uses 691.2 MB per hour, which at current rates can be stored for under \$0.10 an hour. Thus a large 2,000-hour stereo sound effects library at this resolution, having 1.4 TB¹ of data, can be stored for around \$100 to \$200, and the real problem becomes one of acquiring and cataloging the content, not the cost of storage. In fact, the cost has come down by a ratio of about 16,000 to 1 in dollars over the last 20 years, not even including the effects of inflation.

¹ Today a terabyte (TB) is considered to be a thousand gigabytes. An older method of counting that has largely been subsumed through time is to count from megabytes to gigabytes, and from gigabytes to terabytes, each by a factor of 1024.

Random-Access Editing

A particular set of editing features grows out of the random-access-based nature of editing on a non-linear editing system. It is another benefit, although sometimes it can get you into trouble. The sync relationship of a given sound to the picture is easily adjustable—you just grab and move a clip or region of sound along the time line, unless perhaps it has been locked, in which case you would have to unlock it first. Depending on the particular mode of the editing software, it may move only in whole frame increments, or down to single samples, at a time. Nudge edit controls let you move a sound back and forth in time a specified amount per nudge and sync up the sound to the picture as desired. A potential problem is that you may move more sound than you intend, given particular modes of the editing software, so you must be alert as to how various editing modes work. For instance, you may move one region, but affect all others coming after it, in some editing modes.

Non-Destructive Editing

With audio files stored on disk, separate from a session file organized to control the playing of the audio files in real time, comes another benefit: non-destructive editing. In cutting mag film, a sound editor would trim the film from the head of the individual transfer to the position where he or she just wanted the sound to start, then find the tail of the usable sound and cut off the excess. Thus there are three pieces of film: the head trim, the utilized sound, and the tail trim. In order to be able to extend the sound should that become necessary—say to accommodate a picture change—an assistant editor managed the trims and had to be able to find them if a scene had to be extended; that's the way to ensure that matching sound at the edit was available.

With computer-based editing, all of the sound exists for the full length of the sound files all of the time.² In an edit, nothing is really cut. The controlling software points to where it needs to go to play back the sound at the right time, and it is merely a matter of editorial adjustment to add

² This is true most of the time. There are circumstances where, to save space on disk, the portion of the audio file not used in a particular edit is discarded.

something to the heads or tails of a sound. This is much simpler than having to store and locate trims. On the other hand, for particular cases of exporting, for instance, if just a small sound clip is exported from a long sound file, it is overly burdening to export the big file just for the purposes of getting a small clip from it. In this case, export software may be used that specifies handles, an amount of sound occurring before the first fade of a utilized sound clip to after the last fade-out. Then, in subsequent work, if the file parts not used for the export are needed, they may be reimported.

Visual Waveform Editing

The earliest form of sound editing for film occurred on optical sound tracks. The modulation was visible, so most edit points were easily found, both by listening with a reader and ultimately, for the finest editing, by observing the waveform directly from the film. In the 1950s, magnetic recording became the standard, for several reasons: It could be immediately listened to and did not need laboratory processing, and it could be erased and the tape reused, which made it cheaper. However, the recording was invisible, and magnetic sound readers had to be used along with scrubbing to perform fine editing. Scrubbing consists of moving the film back and forth over a reading head to find the start of modulation, a typical place to cut because an incoming loud sound tends to cover up the discontinuity that occurs at the edit (even though the edit is just before the loud sound: it's called backwards masking, meaning that it works backwards in time). A system was developed to write a squiggly line down the back of the film with a marker pen to show the modulation, but this was abandoned as editors got used to editing an invisible medium.

Digital audio workstation editing brings us full circle, back to the early days of sound editing, with its visible waveform, where editing can be performed using both senses, sight and sound. However, it is a rather high burden on the computer to draw the waveform of each of a number of tracks in real time, and it is a testament to the increases in computing power and dropping costs over time that it works as well as it does. Editing systems typically provide a means to turn on and off the visual waveform display on a track-by-track basis, to reduce the computer overhead. The detailed waveform display is replaced with blocks that are much easier for the computer to draw. It might, for instance, mean that more processing

plug-ins could be accommodated operating in real time if the computer doesn't also have to be drawing the waveforms, although how these tasks are parceled out to the various parts of the computer—whether done by dedicated hardware or by the host computer—varies.

Edits and Fade Files

Most audio edits involve something more than just an instantaneous cut. To make such a cut while any sound is playing at all will typically lead to a click, as the waveform jumps at the edit to the new vertical position, cutting together the two instantaneous amplitudes of the waveforms before and after the edit. This applies to a new sound starting in a track, where before the edit the waveform will be flat and at zero amplitude, and to two sounds edited together, where the amplitude of the two waveforms before and after the edit are not identical. Only if the instantaneous level and the direction and slope of the waveforms before and after an edit match can an instantaneous cut be performed without producing a click. Clearly, such a condition does not occur very often. Such an instantaneous cut is called a butt splice, and it is useful in some specific cases, but most edits are a little more complicated. They involve a brief fade. In cutting magnetic film, this is usually restricted by the diagonal splicers used to one-quarter frame, and this is a good rule of thumb for digital editing—in the absence of any reason to do anything else, perform a 10-ms fade.

That 10 ms applies to both a new sound coming out of silence and to the end of that sound going back to silence. It also applies to cross-fades between two active sounds. However, it is just a rule of thumb. Digital audio editing systems give much more flexibility than was ever available from magnetic film splicers in length of a fade or cross-fade and shape of the fade (whether the incoming and outgoing sounds are abrupt or more slowly rising or falling during the length of the fade). In fact, to get longer fades with magnetic film, chemicals were applied to the oxide of the film to selectively remove the oxide coating containing the recording along the length. Some kinds of music editing of smooth string sections, for instance, use long cross-fade times. Two magnetic tracks were necessary, running in sync, with one fading out while the other faded in. This is another advantage of digital audio editing: you need not use up two tracks for two sounds that smoothly cross-fade because they can fit on one track. However, this can lead to some editing peculiarities, because the underlying

file must be long enough to complete a desired cross-fade, and this fact may not be obvious from the layout of files on screen.

In order to make a fade, or half of a cross-fade, the DAW has to multiply each sample during the fade time by a factor that increases over time for a fade-in or decreases for a fade-out. If such a fade-in is performed on several tracks at the same time, it has the potential to overwhelm the processing power available. Thus some DAWs use a separate file, called a fade file, for this purpose, wherein the fade has been precalculated³ and is stored. In terms used for pictures, it has been rendered. To play a complete region, the DAW first plays the fade-in file, then the audio file (which can be cut to instantaneously in this case because at the end of the fade-in all of the conditions are met for making an audibly seamless edit), then the fade-out file. The DAW jumps to fade files as necessary, an easier act for the computer than actually calculating the fades on-the-fly. This is why the structure of some editing systems contains both audio files and fade files, and explains why fade files can be reconstructed if they are lost—because all of the information to re-perform the fade exists in the audio files and the session file, which contains all the editing instructions for a particular project. This file is similar to an Edit Decision List of a picture editing system. It doesn't store the actual audio, but rather tells the DAW how to play back the audio files.

File Management

Three types of files have been discussed: project/session files, audio files, and fade files. It is crucially important to understand what these types represent, because many hours of postproduction time have been lost by exporting just a project/session file without the underlying required audio files, for instance. The DAW software must understand where to find the audio and fade files, and if their relationship is changed by moving them around on disks, for instance, the software can lose track of where it is supposed to get the audio. At the least, if files are moved, you may have to relink the broken connections. Of the three, the project/session files and audio files are essential, whereas the fade files are a little less important

³ That is, the audio has been acted on by the varying fade factor over the length of the fade, producing a new, specialized audio file called a fade file.

because systems like Pro Tools can reconstruct missing fade files from the underlying project files and audio files if necessary.

Exporting audio from Avid workstations and other picture editing systems utilizing OMF and AAF file formats may involve a process known as consolidation. In this process, just the portions of the audio files actually in use are exported (perhaps with handles, additional content attached to the heads and tails of a clip to promote the sound editor's ability to create smooth edits). Long segments that are not used are not exported, so the total export may be in some cases greatly reduced, thus speeding up operations. It is important to note in this case that if a portion of an audio file not used in picture editing is needed for sound editing—for example, to find presence or an alternate performance of a line of dialogue—it will have to be reimported and synced by hand or using metadata, according to the capabilities of the DAW in use.

How to Edit

The foregoing discussion describes many of the benefits digital audio technology has brought to the field of sound editing over the last 20 or more years, but all that is background for the real problem confronting us: how to edit. Of course there is no one way to do everything, as we're discussing an art here. But there are some time-tested methods that can help understand the process and thinking of sound editors. Not every program will fall into a neat category, but many will, and examples from the archetypal categories are important in illustrating what the task of editing sound involves.

An overall workflow diagram of the postproduction process starts with editing the picture and corresponding sound for content. Typically, at this stage, the sound edits follow the picture ones in a one-to-one relationship. Once the story has been blocked out this way, the picture editor continues to edit the sound a little more fully, such as adding in voice-overs, providing some simple sound effects, and adding some temp music. At this stage, they are not worried about smoothing the sound transitions from shot to shot and the like, but rather getting the story across most effectively. They may employ some of the techniques to make scene changes that are discussed as follows, for instance, but they will probably not use all of the dialogue-smoothing skills that are the province of the sound editor.

Where do you make a picture cut? Walter Murch explains in his book *In the Blink of an Eye* that picture editors watch the scene over and over and punch the edit key at just the right editorial moment, in real time—they feel it. Then they look at the cut they’ve made, again in real time, and perhaps adjust it a frame at a time until it works. Sometimes a clue lies in when an actor blinks. This process is analogous to switching cameras in live television work. So at this point the raw sound is cut synchronously with the picture by grouping the picture and sound together and making one cut across both. This discussion applies to dialogue-driven editing, but another possibility at an early stage for cutting montage is to bring in preexisting music and to cut the picture to the music rather than the other way around.

To do this work, editors usually use two to four production sound tracks. If monaural camera sound is all that is available, then one or two tracks will do, with two tracks permitting overlaps to be constructed without dealing with cross-fades. Cross-fades at this point wrap things together too much, which just may have to be taken apart later. Using four sound tracks permits two stereo elements to be edited and cross-faded. The next addition to this might be, if the style requires it, some added tracks for hard effects. Let us say that the sound effects have to drive the story, like providing motivation for action by the actors. In this case, deliberate effects tracks (which might get replaced) are needed in an early stage of sound editing, just to tell the story.

The next element to be added is score. Early in the process this may take the form of temp music, sources drawn from libraries of prerecorded music, for which the rights may not be attainable or the familiarity may distract from the story. Temp music is intended to be replaced with score composed for the show once one is written. However, music is needed to get the right pacing of the edit. Watch the main title crawl of *Star Wars* silently, and you will see that the presence of the music greatly affects the apparent speed of the picture—it speeds up with music present. Likewise, there is a large difference between editing on a small monitor and seeing a picture cut on the large screen. Presented on the big screen, edits seem faster, even frenetic, compared to a sequence of cuts on the small one.

This buildup of tracks, from production sound through hard effects to score, is enough to make early temp mixes to enable the producer to get the gist of what is going on. It is always best to lock the picture at a particular point before the export of sound to another system, or for detailed sound work to begin on the same system. The sound work is so meticulous that coming along and making picture changes while sound

editorial is occurring is often frustrating to sound editors, who have to work twice as hard in a constantly shifting environment. Some forms, like episodic television shows, do not have the luxury of permitting picture changes after lock. The workflow is pretty straightforward, from picture editing to sound editing to mixing. Feature films, on the other hand, often have many loops among these parts, as for example test screenings reveal that changes to the picture are necessary, which ripple through sound editorial and mixing. The parts of sound editing that are often left to after picture lock are dialogue smoothing, ambience/backgrounds development, hard effects in greater detail than required for picture lock, and cutting in the score.

So far, the layout of tracks is pretty obvious, because of the way they have been built up over the edit: Production A track(s), as the principal ones are called, possible Production B track(s) to produce overlapping tracks, effects, and music. In detailed sound cutting, however, expansion to a greater number of tracks is useful, to gain greater editorial control over the elements and because more elements will be added. Given all the tracks in the world, one might assign each sound to a track and let that be that. Adjustments could be made globally, from head to tail of each track, because an adjustment such as equalization would then apply only to that one clip. However, this is impractical on a hardware basis, and furthermore, automation has greatly improved what can be done to a single track, with many fine adjustments possible throughout a track. So multiple sounds occupy each track, and the question becomes how to disperse the sound clips among the tracks. The principles are:

- *Keep like sounds in the same track.* It is disconcerting to a mixer to confront tracks that are randomly laid out, with first a loud dog bark, followed by an exterior ambience, on the same track. These two sounds are at cross purposes and don't even match conceptually except to say they are both sound effects. One is a hard effect, adding tension to the scene, because the start of a dog bark raises the question "what is he barking at?" The ambience provides continuity, the connective tissue used over picture edits to assure the audience subliminally that although we've changed points of view, we are still in the same scene. Thus they should be on different tracks.
- *If there is scene coverage of production sound, and all the characters are well covered, then keep it in one track* (or, in the case of cameras with stereo original production sound, two tracks), as the equalization and so forth called for in mixing only needs to be applied to that track.

In the words of Randy Thom, “if it sounds good, it is good,” meaning, in this context, that given no obvious reason to split tracks, don’t split them.

The alternative is track splitting by character, but this is done only in the case where one character’s voice may need to be replaced, because he or she is off-mic or for performance reasons. It is usually difficult to match ADR with production sound exactly, so common practice would be not to even attempt replacing just one line of an actor within a scene, but the whole actor’s performance for the entire scene. This would then call for splitting out the track of that particular actor, for use as a guide track, which would then be replaced with ADR.

- *One split of the production track that may be useful is to split out production effects.* This could be any sound-effect-like sound that occurs in the production track. The reason to split it could be its replacement by other, hard effects, which would simplify the operation later. Another reason might be getting the right mix balance between dialogue and the production effect, which is often wrong in the original recording because the effect is in the wrong scale. Another reason could be that the production effect may be enhanced by cutting hard effects in sync with it, layering the sound to make it bigger, and it simplifies getting the balance among the parts of the layered sound correct if the production effect is split off from the production track.
- *The corollary to keeping like sound together is to separate disparate sounds.* If it is to be a shock cut, with a bang, you probably want to split the tracks before and after the edit, to provide the mixer with an easy way to adjust the degree of the bang. If you cut the clips before and after the cut into one track, the mixer won’t have the luxury of being able to anticipate the cut with a preadjusted fader, but instead will have to jerk the fader at just the right instant to get the effect desired. This can also be handled by microediting of the volume graph of an individual track, but the point is to leave the mix decisions to the mix stage of things, rather than painting the mixer into a corner.

Fine Editing of Production Sound

By the time it comes to fine sound editing, we expect that the actors or subjects are saying the desired thing and that the best performance has

been selected to use. On the other hand, sound editors for fiction films have a variety of sources on which to draw to make sound seem smooth and continuous, and to fix small slips of the lip and so forth, so the details may not be finished. The sources for constructing dialogue performances are as follows:

- *The speaking actor's close-up shot, which can be used for sound editing purposes whether or not the cut picture shows the actor's face.* Only on-camera performances require hard sync as long as the over-the-shoulder shots don't show the lips. So a continuous performance can be created in several ways. The simplest is to use the close-up sound of an actor over both the shot of him and the over-the-shoulder reaction shot of an opposing actor. Virtually no one ever complains that the sound perspective doesn't match, despite the character being alternately on- and off-camera. Here it can often be more important to provide ongoing continuity rather than verisimilitude of matching perspective.
- *Alternative takes.* This includes outtakes that no one thought were any good at the time, because perhaps just the syllable that's needed to fix a hesitation is present in the outtakes.
- *Wild lines.* If the production has looked forward to postproduction requirements, it will have planned for times when too much noise or reverberation exists to get clean sound and will have recorded wild lines for this eventuality. If the noise is under the control of the filmmaker, such as a wind machine, then it can be shut off and the actors can perform their lines with the same level of intensity as with the noise source on and as close as possible to the same performance as when the principal photography was shot.
- *ADR.* A necessity in some circumstances, such as shooting in a working Brooklyn used car dealer on a busy street. Editors work in preparation for ADR sessions by preparing cue sheets for each actor, who usually works alone, and attending the recording session to be sure that all bases are covered. A good ADR stage/recordingist will match the perspective of the shot so that the ADR sounds a great deal like the production sound, only not polluted with the background noise.

ADR requires a fairly quiet and dead space in which to record, a microphone and mic preamp connected to an input on an edit system, and the ability to play back the existing production sound track to actor/performer headphones while simultaneously recording to

new tracks. This is not too difficult to arrange with a conventional editing system, although professional ADR stages will also have such matters as arming tracks, communications via headphones, and a three-beep system for delivering beeps to the actor's headphones just before the start of the line. Digital picture greatly speeds things up and provides the ability to loop within the edit software.

Another issue with a straightforward two-actor scene is dialogue overlaps. If we are on an over-the-shoulder shot and the actors overlap their lines, then we will likely have captured one of them on-mic and one off-mic. This makes it difficult to cut the picture to the one on-mic, because the sound perspective will jump at the cut. One way to handle this problem is to make the actors clear each other's lines in production and to construct the overlaps in postproduction—although this depends greatly on the skill of the actors. Another approach is to be clever about the mic perspective while shooting, so that major bumps don't occur when cutting picture around to various views.

After potentially improving performances by selective use of dialogue from the sources mentioned previously, editors listen past the foreground dialogue to the presence, the sound behind the dialogue, because the evenness of edits that occur between pieces of presence before and after the cut determines whether the cut is successful in concealing the change. Hiding edits is what the work is about, because to hear them takes the listener out of the narrative. There are some alternatives if things don't match:

- A slow cross-fade will hide a presence change better than an abrupt cut.
- Delaying the sound cut compared to the picture cut to just one-quarter frame before the entrance of the incoming dialogue line can conceal the edit. Backwards masking (in time) works to conceal the cut. The incoming dialogue masks the presence for a moment, and when it emerges from behind the dialogue, the fact that it is a different level and timbre than before the cut is not very noticeable. Of course, this is all a matter of degree; a massive change of presence cannot be hidden in this way.
- If the leading sound before an edit has significantly less presence than the one after the cut, it may very well be worthwhile to dirty up the former, rather than trying to clean the latter. The example given is of

crickets with shooting during Golden Hour,⁴ wherein one close-up has a lot more cricket activity than one shot earlier in the hour. The principle is to make the scene smooth and continuous, so attention is not drawn to the edits, even at the expense of sound quality over some shots.

Stealing Presence

Places to find presence include the heads or tails of the takes in use, in between words, among outtakes of the scene, and from deliberately recorded room tone. Most users call presence “room tone,” or in England, atmosphere, or atmos. We use the term *presence* here, because it covers both interiors and exteriors, both of which need the treatment that having continuous presence provides.

If a short, neutral-sounding presence can be found, it can be looped and used to fill a longer interval. By the way, editors spend a significant amount of time listening to find neutral presence without small noises that would be revealed through repetition if the presence were looped. If the section found contains a tonal sound, then the loop can be done with tight editing, matching the waveform at the cut of the loop. If this is done carefully, no cross-fade may be needed. Otherwise, if a head–tail match cannot be made perfectly, or if the sound is more noise-like in character, cross-fades will be necessary between each clip. If the selection should have a tonal character with a rising pitch, then alternating heads-to-tails and reversed tails-to-heads clips permit there to be no abrupt change at the edits—the sound is heard undulating, rather than the sharper sawtooth pattern that would occur with straight looping. Presence loops can also use different lengths for each insertion, a method to conceal the loopy nature of the sound. The idea behind all of these techniques is to hide the fact that it is a loop.

Where Presence Is Used

When a single track or a stereo pair of tracks is involved, presence usage is fairly clear: you want it to be continuous within a scene. If the scene

⁴ The hour before sunset, when particular lighting effects can be achieved as day gives way to twilight.

involves moving from one location to another, then the presence will naturally change throughout the scene, and you can't use a bit from the first part of the scene to work over a later part where the scene has effectively changed. However, when characters are split between tracks, then things get a little complicated. If we add presence between lines of one character on their track, and between those of a second character on a second track, we've just increased the amount of noise: it goes up by about 3 dB if the two presence tracks are at the same level. However, if each character's track contains audible presence, and it is removed between the lines of one character while the other character's track is filled, we may hear the presence bumping up under the lines on the unfilled track. These two extremes represent the conundrum presented to the sound editor and demonstrate why you may not want to split tracks in the first place.

It may, however, become necessary to split tracks. Say one actor is well off-camera in a noisy location in a wide shot. We may find that the recording of foreground characters is adequate with respect to the noise, but using the off-mic performance is just too much of a stretch. What is done to cover this situation is to cut the off-camera lines out of the production track and fill it with presence. Now record and cut the line in from an ADR session on a second track. With good ADR, no additional presence will be necessary because the production sound presence will cover any small background noise in the ADR track. Good microphone perspective on the part of the ADR recordist will also improve the verisimilitude of the shot.

Documentary Considerations

Documentary films often draw on interviews to provide the narrative thread throughout them. However, seeing a person sitting down in an interview situation is often not very exciting, and at the least, supplementary visual material is shot to provide a means to cut away to some other activity as a way of compressing time. Really what's being done is to use the picture in support of the sound, rather than the other way around. Sound collected from many interviews, and at different times within one interview, will be cut together to form a coherent narrative thread for a character. However, there can be sound-editing difficulties involved in this process. One usual method in documentary production, given there is so much footage shot to collect enough to make the finished work, is to

type transcripts with accompanying time codes for locating parts and then to edit what the character says with scissors on paper. Almost no interviewee gives the straightforward, linear presentation needed by the filmmakers without help. So editors will intercut material from different interviews and times in order to make the story hold up and be told well. Although there can be an ethical debate over constructing performances in documentary, in fact it is done every day, and we hope that by cleaning up people's interviews that we serve them, and the video, well.

Editing by transcript causes sound difficulties. The matching from one shooting situation to another is particularly a troublesome case. Editors, while concentrating on story, seem to listen through the equalization, reverberation, level, and other imbalances to what is said, even though there may be strong audio mismatches, and edit for story first, then sound matching. This can leave sound editing a bumpy task to smooth the dialogue.

Fixing Bumps

Some of the differences between clips in either fiction or documentary videos may be ameliorated by the picture editor, such as level differences, but for the more subtle differences, that is left to sound editors. This may be the same person wearing a different hat, because fine sound editing usually proceeds only after picture editing has been done so that the story works. It is actually remarkable how much can be achieved by the application of appropriate, potentially small, changes to make the bumpy ride whole. These are applied on a clip-by-clip basis, and thus may use non-real-time plug-ins, such as Audio Suite in Pro Tools. Using such a plug-in, whole clips are processed at a time, out of real time. A graphics person would call such clips *rendered*. They have the advantage of being precomputed once overall playback is attempted, and thus do not add to the burden of signal processing in real time. The postprocessed clip is automatically substituted for the initial clip, which remains available if the process needs to be removed. The processes that are useful in dialogue editing are as follows:

- *Level matching.* May be accomplished with normalization or gain plug-ins, or by volume graphing the clip, all of which are equal in their result but may be more or less clumsy or faster to do one way rather than another in a given situation.

- *Timbre matching.* This involves the use of equalization and/or filtering, described in the next chapter. Overall equalization is usually left to mixing, but it may well prove useful to have a different equalizer on a single inserted sound clip to better match those around it.
- *Pitch shifting.* Sometimes pitch shifting, usually by small amounts, fixes the fact that a sound clip has been moved from its position in one sentence into another, because of inflection differences in the two uses.
- *Reverberation or worldizing.* Sometimes a sound insertion is noticeably less reverberant than the sound around it. In this case, a reverberation plug-in may do, or worldizing the sound by re-recording it through a loudspeaker and microphone in a room will do.

Figure 9-1 shows the sequence usually employed to edit dialogue.



Figure 9-1a A sequence of tracks as delivered from picture editing.

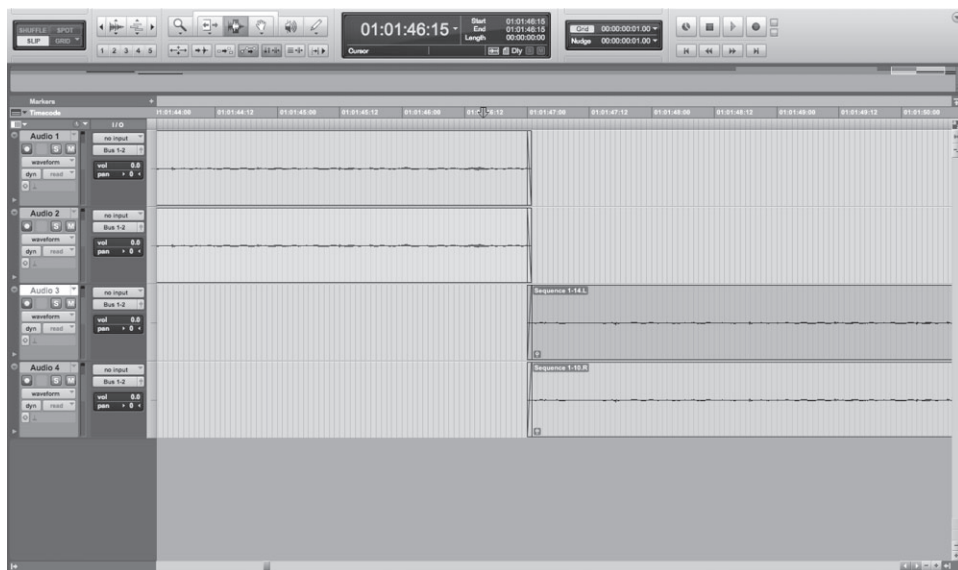


Figure 9-1b Tracks reorganized and split for dialogue smoothing. If a track with a cut in it matches across a picture cut, it is simply cross-faded; if it does not match, then it is split out into multiple tracks.

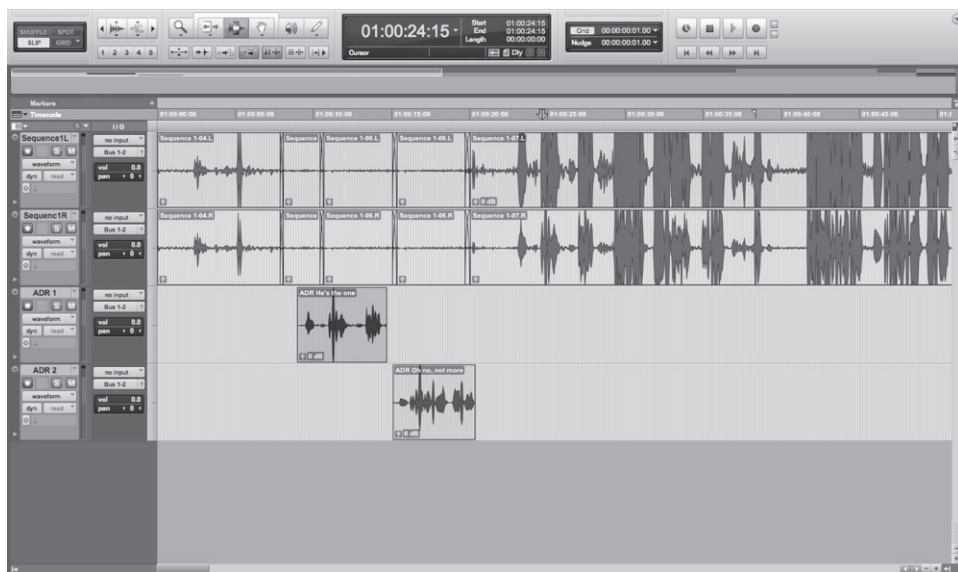


Figure 9-1c Tracks are interspersed with presence (room tone) and cross-faded as needed to sound smooth. ADR tracks typically do not need presence.

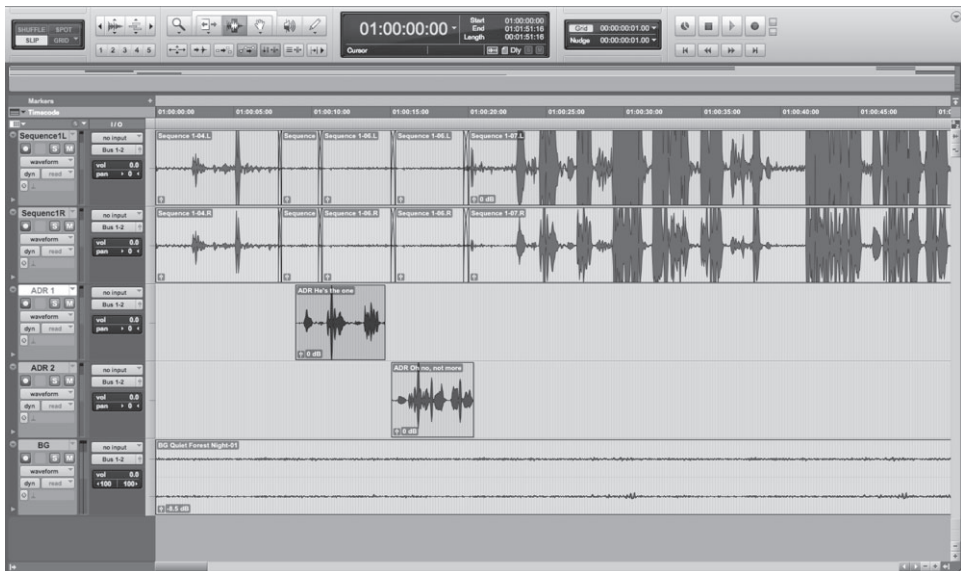


Figure 9-1d Dialogue tracks 1 and 2 have been filled with presence, ADR has been added in tracks 3 and 4, and background tracks (5 and 6) have also been added.

Sound Effects

Next, conceptually, once the dialogue tracks played together sound smooth and continuous, comes the editing of the hard effects. That's because such effects provide motivation for the actors to take certain actions, and they are necessary in storytelling, even at an early stage, to prevent viewers from wondering, for instance, "what is he looking at?" By providing a sound effect, we can tell what he is looking at.

The basic principle of laying out tracks for sound effects follows the thinking discussed before: keep like sounds in the same track and disparate ones in separate tracks. Remember that at this stage you are working in bits and pieces that are only part of a whole; trying to put all of your eggs in one basket, such as all sound effects in one track, only organizes them minimally. You want to separate types of effects and keep the various types on separate tracks.

In recording and selecting hard effects, you generally want them free of background sound. That's so you can cut in just the effect; the presence of

the production sound, or the ambience track to be provided, will cover up any small discontinuity that occurs when the sound is cut in—just the effect will be heard through masking of its inherent low background noise.

An example of what kind of tracks to keep separate is hard effects versus ambience or “BGs” (backgrounds) as they are often called. Ambience is like artificial presence or room tone: it is sound that is imposed by the editor as containing the storytelling threads of the environment. It is like what the set dresser does, providing an environment for the movie to live in.

Hard sound effects may be described as being of two basic types: those that come from sources that more or less emulate what you see and those that are drawn from a much wider range of recordings to fit the picture, but in a more oblique manner. An example of the first case is the sound for the propellers of the flying wing in *Raiders of the Lost Ark*. This aircraft was designed by the production designers and never flew. It had electric motors to turn the propellers, so it sounded wrong. Substituted for them was the sound of a helicopter’s blades rotating, recorded from below (with a strong windscreen), and layered together and cut into sync by making the propellers appear at the correct position for each picture shot. This meant cutting the tracks at the picture cuts into different sound tracks, so that the clips could be assigned by pan pot to different channels, because it is too difficult for a mixer to jerk the pan pots to new locations for each shot. This is an interesting example where the usual smoothness of keeping sound constant across picture cuts is deliberately broken, to make the sound effect seem more real. With a catalog of sound recorded at different speeds of the helicopter blades, and different recording angles and distances to the original blades, a full set of effects is useful in matching the perspectives of the various shots. This illustrates the more literal kind of effect, substituting a helicopter’s blades for the propellers of the flying wing.

Other hard effects, though, bear little resemblance to what is seen on screen. There are thousands of examples of these effects, but the basic method of finding them is to try out various effects and see how they work, one after another, laboriously, until something clicks. An example is the dinosaurs in *Jurassic Park*. Gary Rydstrom meticulously constructed these from existing animal sounds, because he couldn’t go to the zoo and record a brontosaurus. Even experts disagree about what sounds dinosaurs could make, because the fossil record is only of bones, not the soft tissue that creates the vocalization capabilities of an animal. By layering, pitch shifting, and surrounding the basic animal vocalizations with

ambience, footfalls composed of tree crashes, and sounds of the dinosaurs moving through the forest, a real-sounding world was constructed.

There are times too when the goal is not to construct a real world, but rather to leave it behind and enter the scene along with the characters, a more point-of-view driven or subjective reality. One method of doing this is to process sound effects, such as the insect sounds in the jungle scene in *Apolcalypse Now*, thus portraying a kind of hyper-reality. By far the most common method is the use of music, to take us out of the more linear thread of the narrative and into montage, as described in Chapter 8.

There are hard drive, CD, and DVD sound effects libraries such as The Hollywood Edge, Sound Ideas, and many others, containing many thousands of effects. Searching is made more rapid today with centralized sound servers containing sound effects libraries located in editing facilities, and even sources on the Web, such as www.sounddogs.com, www.findsounds.com, and many more. The biggest issue is perhaps that these libraries get so large that finding sounds in them by a written description is difficult, and one must spend time poking around just to see what's there. One company attempting to catalog elements by sound-alike means is www.comparisonics.com, who also use color enhancements to screen drawings to illustrate the frequency range of a sound in addition to the dynamic range represented by height of the waveform as usual. As of this writing, Google returns millions of hits when asked for "sound effects library." One such library shows more than 300 "body hits." Amazon.com offers hundreds of sound effects discs for sale, in addition to the multitude of independent websites selling individual effects and entire libraries. So there are a great many sources for effects, including the original effects used in the *Star Trek* television series and other sound effects culled from studio libraries, for instance.

Despite all of the sound effects libraries, which are quite effective, crews on many films record sound effects specific to that film. If you are doing *Top Gun* and you look in libraries, you probably don't find enough F-14 Tomcat takeoffs and landings to cover all of the possible variations you see in the movie; no effects library is that big. So a crew is fielded to record not only the planes but also everything else in the environment of the aircraft carrier. The attempt by this crew is to get sounds in isolation, because ambience interrupted by other sounds is not necessarily useful. If neutral-sounding moments can be found, however, then they can be looped to form a long ambience of, say, the bridge.

Recording sounds in isolation was the largest problem encountered by astronauts in recording space shuttle interiors for *The Dream Is Alive*. While the astronauts were a quick study, nothing could solve the fact that there are so many noisemaking items on the space shuttle interiors, so recording any one of them in isolation proved impossible.

Other event sounds, like the piston in the takeoff enhancement mechanism cocking, loading, and firing, may be recorded as a sequence. Then it can be taken apart into pieces and used separately, so long as the pervasive room tone underneath it is either low enough in level that the individual events take enough precedence that they overwhelm the bumps that might exist (by providing masking of the cuts from additional ambience tracks, for instance), or the ambience under it can be extended by looping to make a complete-sounding event.

Although sound effects are often kept constant across picture cuts to bridge them with effects, especially ambience, the function in doing this is to root the listener in the scene and prevent discontinuities when none are meant, excepting the point of view of the camera. There are other times when subtle changes are used in timbre across picture cuts to help the audience know where the scene is taking place. For example, in *Black Hawk Down*, which won the Academy Award for Sound,⁵ the sequence on the DVD in Chapter 1 from 10:07 to 10:56 is a conversation between two helicopter pilots closing in on and then passing each other. There are exterior and interior shots. The timbre of the two helicopters is subtly different, and this probably subliminally helps us out in recognizing where we are from shot to shot.

Ambience/Backgrounds

Backgrounds provide the connective tissue for sound tracks. They provide perspective on the scene, interior versus exterior, and more explicitly interior with traffic close by versus a quiet farmhouse interior, and so forth. Generally speaking, they are picked for neutrality, not drawing attention to themselves through any specific sound events. While removing them makes a scene unreal, most people don't explicitly notice them.

⁵ To Michael Minkler, Chris Munro, and Myron Nettinga.

However, they are a storytelling mechanism in and of themselves. In *Boys 'N the Hood*, John Singleton employed the interior perspective of a grade-school classroom with aircraft flying low overhead, as this is the fate of South Central Los Angeles, to be located directly in the flight path of Los Angeles International Airport. Actually shooting in such a classroom would cause problems, because the sound of the plane passing over has to be used in total. Cutting it in the middle of a pass-over, in order to get the foreground words right, leads to trouble. So the way to do this is to shoot the scene in an ordinary quiet classroom and then add the plane flyovers on a background track, separately recorded.

Weather events such as thunder or rain play a part in backgrounds, foreshadowing that things aren't quite what they necessarily seem, as seen by what's in the picture. Special consideration is given to the spatial nature of background effects today, although even the oldest mono sound movies had respect for how space was portrayed. By adding reverberation when desired, the space behind the screen was extended from the flat two-dimensional world represented to a space having perspective. In picture, converging lines produce an illusion called vanishing point perspective, where three dimensions are represented in two by suggesting the third. Likewise with sound, one primary way to represent space, at least interior space, and many exteriors as well, is through added reverberation. Backgrounds are more likely to use at least stereo original recording, if not multichannel, and to be spatialized, than are hard effects. One by-product of using matrix encoder/decoder technology, described in Chapter 11, is that two-channel stereo recordings, depending on how they are made, can appear spread out among screen and surround channels. Sometimes unintentional things will happen with stereo recordings and the matrix, and therefore it is useful to listen to the stereo originals through a surround decoder while laying them in, to understand the eventual effect. You can do this by monitoring through a simple consumer surround receiver.

Foley Effects

The small sounds of real life are often re-recorded for movies on a Foley stage, as described in Chapter 8, Sound Design. Operating much like an ADR recording session, picture is available to the Foley artist, along with many props collected on the basis of what is to be done. Foley is usually

hyper-real, recorded closely on-mic in a quiet, dead space. It has to be exaggerated, because with all the other sounds present, it would not read otherwise.

Cutting Music

Scored music is recorded to fit the picture, so it should be easy to edit in. However, this is often not the case, because *picture lock* is a term that might better be called picture latch. It takes time for the composer to write and record the music, and in this time frame the picture edit has often moved, so music editing becomes a serious task. It is something of a specialty, with a good knowledge of music preferred. To perform a seamless edit, a lot of things have to match: tempo, orchestration, key, and so forth.

Needle-drop music, that which is copied into the time line from existing music sources, is also a possibility, although rights are difficult to obtain. There are several sets of rights involved: the composer, the performer, the record company, and so forth. Existing contracts between broadcast organizations and those licensing rights to playing them on the air, such as ASCAP and BMI, do not include synchronization licenses in their basic agreements. Putting music to picture makes the product a whole new entity and invokes different licensing requirements. Some of the purveyors of sound effects libraries and others also have library music that can be used by buying a license to use it along with the recording.

Scene Changes

The foregoing discussion is largely concerned with what happens within a single scene. Within a given scene, continuity is often the rule, whereas a scene change involves changing the continuity, from making an abrupt break (“there, take that, the scene has changed”) to a gentle approach (music bridging a scene change, for instance). Also, the point at which the picture and the sound change may differ, with the principal sound for the new incoming scene entering before or after the picture change.

Scene change also permits a resetting of our hearing, so that if one recording method has worked in one scene, such as all on a boom mic, then when the

scene change occurs, it is relatively easier to change to a different technique, say lavalier microphones, so long as that stays constant within its own scene. When these methods, along with ADR, become mixed up within one scene, there's more trouble in maintaining continuity. A scene change is like erasing a blackboard—the lecturer gets to start over. This is not meant to say that a given character shouldn't be recorded in the same way and match across scene boundaries, it is just a tendency that there is less requirement to do so across scenes than within a scene.

There are several ways to get from scene to scene, and a great many of them are illustrated in the British comedy *Love Actually*. This film involves cross-cutting among several different characters' stories (which come together at the end). The scene change types are, with citations from the DVD of the film, as follows:

- *The straight cut*. While ambience may change, it is at a normal level before and after the cut. The cut illustrating this is from one dialogue scene to another. Chapter 3, 15:07.
- *The hard or bang cut*. This breaks the ongoing scene and slaps you in the face that a change has been made. Illustrated by the scene change to the blond and black guys in a car, first exterior with their radio playing, then quickly to the interior. Chapter 4, 24:49. Also used at the end of the scene.
- *A music cue bridging across a scene change*. This helps indicate that time has elapsed and softens the change compared to the hard cut. Chapter 4, 23:30; also 26:33.
- *Fade-out/fade-in*. Marking a more significant change than a simple scene cut, this implies a major marker, such as the passage of time or an act change. In *Love Actually*, it marks the change from the opening montage of people loving at Heathrow, with music and a voice-over, to the start of the more narrative part of the film, the recording studio sequence. This transition is actually made rather quickly, because there's no point in stopping the flow so early in the film, and might even be called a cross-fade, but fade-out/fade-in better characterizes it. Chapter 1, 2:02.
- *The prelap sound edit*, also called a J-cut (because of the shape of the letter "J" indicating something happens in the audio track before the video) among video editors. Sound is heard before the picture

change of the incoming scene. This helps propel the story and gives it a little more energy than a straight cut. It causes anticipation for the picture to change. Chapter 2, 9:45. The start of the marriage vows are heard before they are seen. Also the word “so” in the scene change at Chapter 3, 20:28.

- *Post-lap or L-cut.* Sound established in one scene overlaps into the next. Music of the wedding carries over a character arriving home unexpectedly, both over the exterior establishing shot and into the interior, where it comes out. Chapter 2, 11:22.
- *Source music becomes score.* By changes in orchestration or worldizing, the two can be transformed. In *Love Actually*, the recording studio live-sound source music changes into score in several ways: the scene changes to an exterior, the vocals drop out, and the strings swell. Chapter 1, 4:00.
- *Same source music across a cut, but with different perspectives.* In the funeral scene, Liam Neeson fulfills his deceased wife’s request to play the Bay City Rollers’ tune *Bye, Bye, Baby* at her funeral, and in the middle of it we cut to the wedding with the same tune playing in continuous time. The change is marked by a difference in how the sound was worldized to sound different. Also, listening to a radio station in an office, the tail line in the scene is “what is that?” and the picture cuts to the radio studio originating the broadcast, with continuous sound but in the new perspective of the station, and the announcer says “That was....” at Chapter 3, 20:07.

Table 9-1 summarizes the opening sequence of *Love Actually* in terms of the above techniques.

Plug-Ins/Processes

With the creative side of sound editing in mind, it is informative to consider the technical capabilities of sound editing programs such as Pro Tools and Soundtrack Pro that allow sound to be edited efficiently and tailored to a particular production. Traditionally, editing personnel have performed edits and mixing personnel have done virtually all of the signal processing necessary to ensure a good result. The only exception to this rule was one where the transfer room that made the mag film transfer

Table 9-1 *Love Actually* scene transitions.

Scene Change or Event Chapter/Time	Sound Transition/ Technique Used	What Happens	Notes
1/0:50	Music, then voice-over	Montage: Heathrow airport greetings, then explanation “Love Actually Is All Around”	Montage reinforced by hearing music and voice-over only, no dialogue or ambience
1/2:03	Fade-out/fade-in, music (score) to music (source)	Go to music studio interior where singer and backup vocalists perform	Scene becomes more real, less montage-like because of sync sound. When manager speaks over talkback circuit, first we hear feedback, further establishing where we are.
1/4:00	Music swells, vocal drops—source music transitions to a score-like cue	Cut to exterior establishing shot, then title “5 Weeks to Christmas”	Continuity provided across multiple montage shots establishing period
1/4:18	Music continues but we start to hear sync sound	Cut to interior	Music fades under dialogue but keeps playing
1/4:53	Same music continues, but now we hear sync sound too	Cut to exterior establishing shot, new residence	Music with sync sound has a way of sounding more immediate and less like a montage than music only
1/4:57	Music continues over cut and phone conversation	Cut to interior, Liam Neeson at his computer	
1/5:15	Music continues under conversation	Cut to recipient of phone call	

Continues

Table 9-1 Continued

1/5:33	Music continues under dialogue	After cutting back and forth, we wind up on a scene with Emma Thompson in her kitchen	Scene transition wasn't all at once, but rather back and forth, winding up on the far end of the initial phone call
1/5:50	Music continues	Exterior office establishing shot, blond guy enters door	
1/5:54	Music continues	Interior office, continuous time	Music continuity over these edits emphasizes continuous time
1/6:20	Music continues	Interior porno set day	
1/6:49	Music continues	Interior church day	
1/7:24	Music continues	Exterior 10 Downing St. day, the Prime Minister arrives, camera follows him inside	
1/8:06	Music fades out under scene, not a scene change		This is the first time since the beginning that we are without music. It is a relief in a sense and shows how far we've come in getting into the story.
1/9:10	Music cue fades in under scene		Music cue underscores the importance of what has just happened
1/9:42	Prelap sync sound of next scene	Interior church day; back to marriage ceremony	"In the presence of God" is heard over the outgoing scene, yet belongs to the incoming scene

Continues

Table 9-1 Continued

Scene Change or Event Chapter/Time	Sound Transition/ Technique Used	What Happens	Notes
2 /10:00	Source music wedding march pipe organ, transitions to Beatles "All You Need Is Love"	Singers, soloist, band members come out of nowhere	Note each part of building players is panned to the correct position
2 /11:22	Source music continues unchanged across cut	Exterior establishing shot, wedding guest returns home to check on sick wife	Before the cut it is source music; after, it is score
2 /11:25	Music continues across cut then fades out	Interior, day, continuous	She's not sick after all
2 /12:06	Hard cut	Wedding reception interior, day	
2 /13:15	Hard cut	Wedding reception food prep area day, moments later	Compresses time as Colin jumps from embarrassment to explanation with his buddy
2 /13:55	Hard cut	Establishing wide shot of set with lights, etc.	
3 /15:07	Hard cut	Interior, day, funeral, quiet presence, then Liam speaks	
3 /16:07	Bay City Rollers: <i>Bye Bye Baby</i> as source music		He cues it in over his tears
3 /16:48	Music continues across cut as source music in a new location	Interior, wedding reception, continuous time	Timbre change at cut indicates scene change in source music—both scenes are playing the same song!

Continues

Table 9-1 Continued

3/18:02	Source music continues across cut, becoming score	Exterior establishing shot footbridge over river	
3/18:05	Music continues across cut, then fades out	Interior office day	
3/18:48–19:22 19:24–19:36	Music cue: score		Emphasizes the point: she’s in love with Carl, and everybody knows it. Broken into two pieces, it is probably more effective than one.
3/19:48	Tune is on the radio		We’ve heard this one before
3/19:51		Her obnoxious phone rings	
3/20:04		“What is that?”	
3/20:07	Music continuous across cut, but with different timbre. It is source music in both cases, but now we’re in the originating studio.	“That was...” a very funny way to answer the question posed before the cut.	
3/20:22	Hard cut: we hear the voice of the announcer continue, but worldized	Interior, day, continuous time; faded pop star and manager in radio studio waiting room	
3/20:25	Hard cut	Back to studio announcer on the air	
3/20:28	Prelap edit “so”	Interior, day, new studio, interview	Time is compressed: it passed in the cut

Continues

Table 9-1 *Continued*

Scene Change or Event Chapter/Time	Sound Transition/ Technique Used	What Happens	Notes
4/22:40	Source music plays across picture cut, but fades out rather quickly	Interior, day, the cabinet room	Music not desirable under tense cabinet meeting, set up by announcer “is the new Prime Minister already in trouble”?
4/23:24	Music cue score under		The PM’s love is obvious—she has entered
4/23:32	Score continues under scene change	Interior, PM’s office, day, later	Time compressed
4/24:10	Hard cut	Interior day studio	
4/24:40	Music cue score under		Even in the most outrageous of circumstances, people can relate
4/24:49	Bang hard cut	Car exterior day	We hear car radio more than the street
4/24:52	Hard cut, continuous	Car interior day	Car radio continues over dialogue, but it’s lower level in the interior than the exterior
4/25:42	Hard cut	Office interior day	
4/26:20	Music cue score under		“hoping to be kissed”

Continues

Table 9-1 Continued

4/26:32	Score continues under scene change, then fades out	Interior, Liam's house, day	
4/26:49 – 27:32	Music cue score under	Camera cranes us to boy's room, then cuts back to live action	Mysterioso music: "what's he up to?"
4/28:08	Music cue score under, same theme as last one	Continuous	The scene ends with them going for the comfort food.
4/28:15 – 28:55	Score continues under scene change	Exterior, day, park bench overlooking Thames	Liam and stepson have the big talk
4/29:04 – 29:37	New music cue	Continuous	"Truth is, I'm in love" to "worse than the total agony of being in love"
5/29:44	New music cue	Continuous, but we are back on the wide shot	Helps get out of the scene
5/29:46	Score continues under scene change	Interior, night, office	Laura Linney at work—she's working up her courage to ask Carl out
5/30:27	Her cellphone goes off	Continuous	She's always being interrupted (as it turns out, by her incapacitated brother)—the music is bittersweet
5/30:38	Score continues under scene change		

might operate on the original sound, such as making an off-speed transfer, or one reversed in time, often performed to conceal its source. Editors were restricted to editing by the practicalities of the medium. Today DAWs have many processes available to alter the sound, and these range from the straightforward to the complex. Some of these are extensions to basic editing:

- *Time compression–expansion.* A complete sound may not fit within a given track length of time. By spreading out or shrinking the area concerned, a sound may be made to fit into a time sequence. This process is not perfectly transparent in digital audio, because small clips are being removed or inserted, so the more the process is stretched, the higher the likelihood of hearing artifacts. Also, time compression–expansion may be performed with or without accompanying pitch shifting.
- *Pitch shifting.* A related process to time compression–expansion, pitch shifting often proves to be particularly useful with sound effects. Lowering their pitch makes them seem bigger and more impressive than the original sound on which they were based. Pitch shifting may be performed with or without accompanying time stretching. Pitch shifting can be useful to dialogue editors in order to insert words from a different take or part of a documentary that don't quite match in pitch.

Note that a combination of both pitch shifting and time changing may be accomplished with an analog tape machine equipped with a varispeed control, and this can be much more effective for large speed changes than using digital techniques because analog doesn't suffer from the artifacts of splicing. However, the speed change and pitch shift are by necessity locked together in a one-to-one relationship. In an extreme example, Gary Rydstrom reports he used a U-Matic recorder set to move only at the rate necessary to prevent excessive tape wear—its nominal pause mode—and recorded it onto another machine overnight for odd sounds used in point-of-view sequences in *Strange Days*.

- *Reverse.* This takes an audio segment and reverses it heads for tails. It is useful in making sounds less recognizable, although if they have recorded reverberation, that can prove unnatural as the reverberation occurs before the event that caused it.

Reversing sound segments can be alternated with the same sound segment playing forward to improve the discontinuity at edits. Say that a sound effect contains an element with a rising pitch over its length. If it were looped, with the tail of one segment spliced to the head of the next repetition, the jerk in pitch at the edit would likely be noticeable. By alternating segments heads-to-tails, tails-to-heads, then heads-to-tails, and so on, there is no abrupt change in pitch at the edits.

- *VocALign*. This is a plug-in from Synchro Arts that uses a production track as a guide and slips an ADR performance on another track into sync. It does this by automatic microediting of the ADR track to match the original track in time.
- *Subharmonic synthesis*. Normally used with sound effects, such plug-ins find the lowest tonal frequency in a particular frequency range, such as 50 to 100 Hz, and produce an added output at half of the frequency, thus adding a new fundamental. The effect is to make larger and more impressive specific sounds, particularly explosions, cannon fire, and the like. One example of the use of such a device is on the giant voice of the tomb near the beginning of *Aladdin*. Subharmonic synthesis may be combined with slowing down the original recording, and equalization, to make prodigious amounts of bass.

These processes are performed on the clip/region of an audio file that is used by the project/session. A new audio file is generated and inserted in place of the existing one. The original file can be called back if the process is to be reversed for some reason. Some years ago, Lucasfilm produced a made-for-television movie, *The Ewok Adventure*. Randy Thom was the sound designer for the film, and he time-reversed an effect for an ogre that stuck out of a tree and uttered a screech. He thought it sounded better that way. Then he found out that a sound editor had already reversed the screech and that he had put it back the way it was recorded!

Editors also employ some processes that were traditionally done by mixers, or in transfer suites before they received the sound. In this sense, however, these are used editorially, not as a mixing function, as explained in the following.

- *Normalize*. This is a level-setting process that examines the audio clip/region and increases the gain until the highest instantaneous peak recorded level just reaches 0 dBFS. It is useful for under-recorded

content, although working on peak level, as it does, does not guarantee that loudness has been normalized. Normalization is typically done on an audio region-by-region basis and not on a whole track because it is used to get an overall smoother performance before mixing is applied.

- *Gain.* A level control applied uniformly throughout a specific clip/region. The segment is replaced by one with the altered level. This is especially useful in matching a particular word, for instance, from a different take, that might be different in level from those before and after it. Another reason to apply gain is to increase the level of under-recorded material without pushing it to the top of the range as normalization does, which could be too far for aesthetic purposes. Some DAW editors in mixing mode only have 6 dB of gain available above unity—defined as the point where the input and output levels are equal—and thus it is likely that more gain will be necessary at some point because of low-level recording.
- *Volume graphing.* An automation function that allows the editor to specify points in time for particular playback levels and to graph those over time. A mixer-like interface may be provided with automation so that a person can ride the gain to smooth out a performance, or an editor can take a view of a track that shows the level over time, which may be edited graphically. In this way volume graphing is different from gain, because it is dynamic, changeable throughout a region, and from region-to-region. Whereas volume graphing performs the same act as the mixer does when facing the console faders, there are reasons to volume graph certain kinds of events, while leaving others to the overall mix.

One reason to volume graph would be to match a region inserted with those surrounding it in level, similar to use of the gain function and accomplishing the same thing. Another primary reason is to tame certain features of dialogue recordings, like a cough or sneeze. An editor can volume graph around a cough more effectively than a mixer can mix around it, because the editing process takes place out of real time, whereas the mixer has to hit the cough exactly, or the duck of gain in front of it may be audible, for instance.

- *Noise gate.* A switch controlled by the signal that allows sound through when the level is above a specific, adjustable level, and otherwise

mutes it. A noise gate might be used to clean up dialogue that has some background noise. With a cleaned track, and an additional background track, bumps in the dialogue track caused by uneven presence across the edits may be concealed.

- *Noise reduction.* More sophisticated than a noise gate, a range of plug-ins are available that attempt to separate foreground sound such as dialogue from background noise and suppress the background, using a variety of strategies and at a variety of prices. Some of these learn the noise floor, frequency-by-frequency, and provide a multiband noise gating function at just above the ongoing noise floor.
- *DC offset removal.* Cutting together sound clips from different sources may produce clicks because the DC level of the two different converters in the two paths are different. In fact, this is a fairly common problem, because with 16-bit audio the smallest step is about 1/65,000th of full scale, and keeping the DC content to this small level is difficult. Thus, if there are clicks or thumps at edits, a DC removal filter is indicated. (If cross-fading has been done, this may not be necessary: a cross-fade usually takes care of DC problems.)
- *Clip removal.* (Note that this is not the same use of the word *clip* as in a segment in an audio bin of DAW audio region list.) A software device to find clipping distortion and draw in an appropriate peak. Because clipping of original sources is usually the worst type of distortion that occurs, use of this plug-in can save problematic clips.

These processes are used to smooth out bumps editorially as, say, the level varies for a sound insertion that otherwise matches. In this sense what is being done at the editing stage is relative, clip-to-clip, not necessarily the absolute final choice of level that will be applied in mixing. Any of these processes might be used on dialogue, sound effects, or music fairly early in the editing process, as they are useful to match regions. Other processes such as filtering, equalization, compression/limiting, and reverberation might be used too, but they are more generally thought of as mixing processes and so will be covered in the next chapter. Many plug-ins are available on the market, both from the original developers of DAWs and from a whole industry that has grown up to write specialized software. Plug-ins are written for particular sets of hardware/DAW software systems, as shown in Table 9-2.

Table 9-2 Audio plug-ins.

Name	For Environment	Notes
Audio Units (AU)	Mac OS X	Real-time native processing*
AudioSuite	Mac OS, Windows, Pro Tools	Non-real-time, operates on clips/regions to create new ones that replace the originals. Not automatable.
DirectX	Windows programs that implement it	Real-time native processing
Host Time Division Multiplex (HTDM)	Pro Tools	Performed by the host CPU in an environment that also has a set of DSP chips, such as Pro Tools. Automatable.
MOTU Audio Systems (MAS)	Mac OS host software synthesizer programs	
Premiere	Mac programs such as BIAS Peak, BIAS Deck, Logic Audio, Opcode's Studio Vision Pro	Non-real-time process performed by host CPU
Real Time Audio Suite (RTAS)	Pro Tools	Performed in real time by host CPU. Automatable. Limitations include being only 1 in, 1 out and restrictions on where they can fall in the chain.
Time Division Multiplex (TDM)	Pro Tools with particular hardware	Performed in real time by DSP. Automatable.
Virtual Studio Technology (VST)	Mac or PC, dependent on the host program	Real-time native plug-in format.

*Native means that the software runs on the host CPU, in this case as a part of the operating system.

Tracks and Channels

The organization of multiple simultaneous tracks has its own descriptive language, developed from the analog case. Tracks are what you see on the screen, individual time lines that contain the audio clips/regions. These tracks are analogous to the 24 tracks of a two-inch multitrack analog tape machine or to the various strands of cut magnetic film interleaved with fill leader to make each strand match all of the others in length, interspersing the sound where it belongs in the footage to match the picture. Tracks may be assigned directly to channels. Channels may be thought of as electronic pathways that may deliver particular inputs to particular tracks or, conversely, tracks to outputs.

Tracks may also be routed to busses instead of directly to input/output channels. Busses get signals from one place to another, and they also mix them together. Once mixed together, only hearing separates dialogue, music, and effects, so if you need separation for workflow reasons, separate busses for the various parts must be used.

Busses

The number of busses available on a given system determines how complex a mix can be accommodated by it. For instance, we need six busses to accommodate 5.1-channel mixing, but if we wish to perform the same mix but keep its component parts still separated, we need 18 busses, six each for dialogue, music, and effects. Such complex mixing is probably beyond the scope of what many people want to do, but it has a primary advantage: the preparation of music and effects (M&E) mixes is easy, for foreign distribution in a dubbed language. This style of mixing is called *stem mixing*, for keeping the ingredient parts of a complete mix separated.

There are three items that require the services of a buss:

1. *Mix busses*. These represent the final mix of all the elements together, or in stems. Typical mix busses would be Left, Right, Center, LFE, Left Surround, and Right Surround, connected directly to output channels labeled the same way.

2. *Track busses*. In multi-tracking of instruments in a scoring session, one keeps the various instruments separate and records them onto separate tracks, for which these busses are used.
3. *Aux busses*. The idea here is to pass a signal from a channel to another device or process, and then back to the main signal flow. It is a detour from the normal signal flow. Such a process might be reverberation, for instance, where the high complexity and quality of a reverberator may require a piece of outboard gear to perform it. You might, for instance, assign multiple channels to a bus called an auxiliary bus, which then could deliver the signal to an added channel called an auxiliary return that may contain some common processing for all of the tracks assigned to it, before the signal is passed on to an output channel.

Pan Pots

The outputs of channels may have pan pots (panoramic potentiometers) invoked so that a single input, the channel, may be panned among multiple output busses and thence to output channels, as in 5.1 panning. So channels can be hard assigned to specific output channels, or they may be sent through a panner to a variety of output channels.

Solo/Mute

Solo and mute are functions associated with individual tracks as well. Solo means that the output is muted of all other than soloed channels. Mute kills the output of a channel. These are arranged as buttons to activate in most DAWs. In addition, the mute function may be one that is automated through the length of a track. This is useful in the case where some sound is thought useful but not necessary at the moment. It is still there, available in sync, but it is muted. Automation systems can often record mutes, from a mixing screen or an editing screen.

Grouping Tracks

Tracks may be ganged together, called grouping for common editing or setting of processing parameters such as level, and the grouping may be

switched on and off. In addition to ganging whole tracks from head to tail, in some modes of editing software clips/regions may be ganged for editorial functions. Most systems permit grouping contiguous clips together and moving them as a block, and some systems perform the grouping even when the selected clips are discontinuous.

Differences Between Picture and Sound Editing Systems

Many picture editing systems are frame based, extending their picture thinking (only whole frames count) to sound, with edits taking place on frame boundaries, and the minimum nudge being one frame. This may be acceptable for a lot of work, but film sound editors prefer somewhat greater editorial resolution. One reason is that there is a phase issue between the start of a sound and edge boundaries of a frame. That is, a sound may start three-quarters of the way through a frame, and it becomes unclear where to cut. If you cut on the leading frame boundary, you have the potential of hearing a burst of undesired noise before the desired sound cuts in (which you may be able to cover up with a fade file), or if you cut at the trailing frame boundary, you have cut off the impact of the beginning of the sound.

Many digital audio workstations perform work with a time resolution down to just one sample, which we might think of at first glance as an embarrassment of riches. While single-sample resolution is never necessary for picture-sound sync, it is useful in certain instances. For example, if two tracks that have associated stereo content have a one-sample sync slip between them, this will be audible in the position of a phantom image between two speakers. Don't believe me; try it yourself. You can hear a one-sample offset, which is about 21 μ s, in its effect on the position of a phantom image formed between stereo channels because the minimum that can be heard is in the 10 μ s range.

Another potential problem solved by having single-sample editing and nudging resolution is that if two associated tracks become out of sync by even one sample, then are mixed together, the result is a frequency response dip, a part of the frequency range that is carved out. This is one reason that some sound editors like to work most of the time in a frame-based mode, so that all moves in time are in one-frame increments, on all tracks, unless otherwise needed.

Other differences between the mostly picture-oriented and the sound-oriented workstations are that the sound systems typically offer more tracks, greater flexibility with plug-ins, more provisions to handle multi-channel surround formats, more sophistication in bussing for auxiliary purposes, and so forth. This means that some kinds of shows can be fully postproduced on current editions of picture editing systems, while those requiring a little more of sound postproduction probably require a specific step on a sound workstation, at least.

Picture–Sound Sync Resolution

While the tolerance of many pieces of video equipment is ± 1 frame of picture–sound sync, in fact one frame of out-of-sync sound is noticeable to sound editors, and two frames is quite noticeable to virtually everyone. So if two pieces of video gear with a ± 1 frame tolerance are put in series, and both happen to be at the end of the tolerance, say both at +1 frame, the resulting two-frame error will be very noticeable.

Sound designers observe that having sound appear up to a frame early is better than if it is a frame late. This is curious, because in real life, sound is out of sync often because the speed of sound is so much slower than the speed of light. Watch a soccer match from the stands closely, and you will observe that the sound is out of sync, because at 1130 feet per second, sound is pretty slow. Interestingly, motion-picture prints routinely compensate for the speed of sound by printing the sound on the print one frame earlier than expected for dead sync so that the sound arrives in sync at a point about 50 feet from the screen. For most digital video work intended for television display or exhibition in a smaller space, such considerations are unnecessary. The most important consideration when editing is to watch the video on a monitor while listening to cut sound elements and ensure that the sound events appear in sync to the eye during playback.

Director's Cut

- Computer-based Digital Audio Workstation editing employs a non-linear, random-access, non-destructive, visual waveform-based editing model. These developments together have greatly increased editorial productivity. Cutting sound on film averaged about four sounds cut per editorial working hour; with DAWs the rate is much higher.
- There are two or three basic file types: the audio files themselves called audio files or clips, possible fade files constructed from the audio files for use in making smooth transitions, and project/session files. Project files control when the audio and possibly fade files are played out. Both or all three file types need to be available to play back.
- Sound editing of production sound occurs along with picture editing, but tracks are usually limited in number, and picture editors do not concentrate on the issues that affect fine sound editing. Thus editing may occur in two phases, first with just production sound and perhaps a few added effects, and then after export to a sound editing system, in fine detail.
- Track splitting of a continuous track is done when a mixer has to make a change at an edit, for example.
- Dialogue smoothing uses multiple sources: production sound, wild lines, and ADR, and is interspersed with presence, also called room tone.
- Sound effects libraries on hard disk, on optical disc, or available over the Internet are often used for effects, except for specialized effects needed for particular programs.
- Sound edits across scene changes can be of various types, which have different feels. Among them are the straight cut, the dissolve, the prelap edit, the postlap edit, score carrying one across the edit, and source music carrying one across the edit with appropriate change.
- Audio plug-ins are of two types: one processes entire clips and substitutes new copies with the effect desired, and the other works in real time upon playback. The first is less computer intensive because it can occur outside of real time, whereas the second can have automation functions.
- Many plug-ins represent processes that traditionally were the province of mixing, and still sometimes that is the best place to apply particular processes. On the other hand, fine surgery of sound may be done with plug-ins during editing, for which there may not be the time to deal with during mixing.
- Individual editing systems will have slightly different methods of routing signals, grouping tracks, soloing or muting channels, and the like, so these skills must be learned for the particular system.

This page intentionally left blank

CHAPTER 10

Mixing

There are two primary ingredients to mixing: performing a variety of audio processes on the tracks, and configuring those processes together, within and among the tracks. With the advent of the Digital Audio Workstation (DAW), plug-ins representing many different audio processes that traditionally had been the province of mixers became available to editors. So the question becomes the division of labor: how much equalization to do during editing versus how much to leave to the mix, for instance.

There are several primary differences between editing and mixing:

- *The training and experience of editors and mixers is different, although the disciplines are converging.* Nonetheless, on the highest levels of the Hollywood feature film, for instance, the roles are separate. Editors tend to concentrate on the trees (e.g., working within a track, especially at the edits, and getting blend within a track so that it plays across edits without bumps, and checking how the tracks are laid out), whereas mixers concentrate on the forest (e.g., seeing how the overall equalization of a track is working and how tracks are blending).
- *The environment of mixing is one with appropriately scaled presentation of picture and sound.* It also incorporates explicitly standardized calibration of the level and spectrum (frequency range and octave-to-octave balance over the range) and de facto standards on the direct-to-reverberant ratio of monitoring, picture-loudspeaker arrangement, surround layout, and others. Mixing in a calibrated environment has a much better chance of making recordings interchangeable with others with respect to dialogue level, timbre, and so forth.
- *Editing is almost always a solitary activity, preparing for the mix.* In a professional environment, the mix may occur with the producer, director, picture editor, or others in the room.

Even with just one person performing editing and mixing, there is some difference between what is best done editorially and what is left to the mix, for the best efficiency. Generally speaking, even though they use plug-ins, editors concern themselves with using them for matching within a track. Thus a gain plug-in may be used to match levels of a region inserted within a track, and time-compression used to fit sound to a specific clip length. On the other hand, extensive volume graphing to achieve balance among tracks at the editing stage may just hamper the mixing. So while an editor might use volume graphing for balance among tracks, these automation instructions might be erased for the mix. A better way to balance tracks during editing, if you have the hardware, is to run the various channel outputs to an external mixer, to set the relative balance among the tracks for the monitor, without affecting the level that occurs in the track.

Some clean-up activities are easiest for the editor to volume graph. Getting a cough or a sneeze into the right level range by tightly volume graphing around it is very valuable. That's because asking a mixer to catch a sneeze is problematic (and all that could be done some years ago), whereas today it is far easier for the editor to volume graph such an event. The normalize and gain plug-in functions aren't as good for this because normalize only fixes the insertion at the maximum permissible level, which is certainly not likely to be right, and the gain plug-in fixes the gain throughout the separated region at a fixed value. To use the gain function, one would have to (1) separate the cough from the original region into a new clip, (2) duck the level of the clip with the gain function, and (3) then likely have to do cross-fades at both at the head and the tail of the insertion, a total of about five functions. Volume graphing is simpler, but has a problem too: you might easily overwrite the volume graphing while you are mixing. For instance, some systems have touch-sensitive faders where, if you are touching the fader, you have the volume graph function under your control. If you are touching it during the cough, then the duck provided during editing by means of a volume graph gets overwritten. Avoiding this mistake is an example of the understanding of the grammar that it takes to operate these workstations effectively in editing and mixing modes. Each system is different, as is sometimes each revision of software, so the best advice is to do some testing of issues such as these with your system before committing a major mix to it.

If volume graphing is used editorially for items such as the cough duck, then the question becomes: What is the standardized gain setting of the volume graph, the gain to which we return during editing for all but short,

sharp items like coughs? It seems best to leave tracks at unity gain, 0 dB, and to adjust them for relative level in an external mixer for monitoring until the mix occurs.

Likewise, using equalization plug-ins may be useful for matching insert edits to surrounding regions, but determining overall equalization is best left to the mix, where, especially, the calibrated monitor system affects what is heard so that equalization judgments are best left to the mix room. This is because the frequency range and response of the monitors typically used in editing are poor compared to the mix room.

While hardware consoles with dedicated hardware for each audio channel were common for film and television mixing even in digital workflows several years ago, mixing for digital video increasingly takes place on control surfaces like the one shown in Figure 10-1, which effectively act as remote controls for DAWs running Pro Tools or similar software. The knobs and faders on a control surface update software parameters such as



Figure 10-1 A mixing control surface with faders for controlling level and knobs for adjusting other parameters in a digital audio workstation.

volume levels and plug-in automation in real time as a program is played back, allowing mixers to control in software a variety of processes that once required dedicated outboard gear. The processes and domains used in mixing audio today remain largely the same as they were in the era before such “virtual” mixing, so it is only the way in which the desired effects are realized internally that has changed with the advent of software control surfaces. One important caveat is that anything written using a control surface updates the DAW session file containing the audio and associated processing instructions, so volume graphing written during editing, for example, can easily be overwritten by automation performed during the mix. If important data is contained in the volume graphs made during editing, the mixing surface should be put into trim mode before writing automation so that any changes made to the underlying data are calculated relative to the existing volume graphs or plug-in settings on the DAW.

The sections that follow will focus on the road map for mixing audio and the specific audio processes that can be used to achieve the desired sound for a program.

The Mixing Hourglass

Your sound may be fairly simple, represented by just a few tracks. In this case, it may be possible to mix the tracks together directly to an edit master (see the next chapter) in just one step. However, for many mixes, things get more complicated because of the number and variety of tracks involved. In such a case, it is customary to first make a mix of like sounds together into what is called a premix or predub for each of the major ingredients: dialogue, ambience, Foley, hard effects, music, and so forth. Once these are done, they can be played together to make a final mix. A block diagram of the process is given in Figure 10-2.

Each of the premixes should contain all of the level balancing, such as fade unders for dialogue, that is needed for the final mix. This is accomplished by a technique called Mix in Context (Figure 10-3). In this system of working, the dialogue premix is made first, if at all possible, and then it is played back while another premix is being made, such as ambience. Then the ambience premix has already been balanced with respect to the dialogue. Doing this process for each premix in turn, and playing back all of the available ones at that point, allows for each of the premixes to be

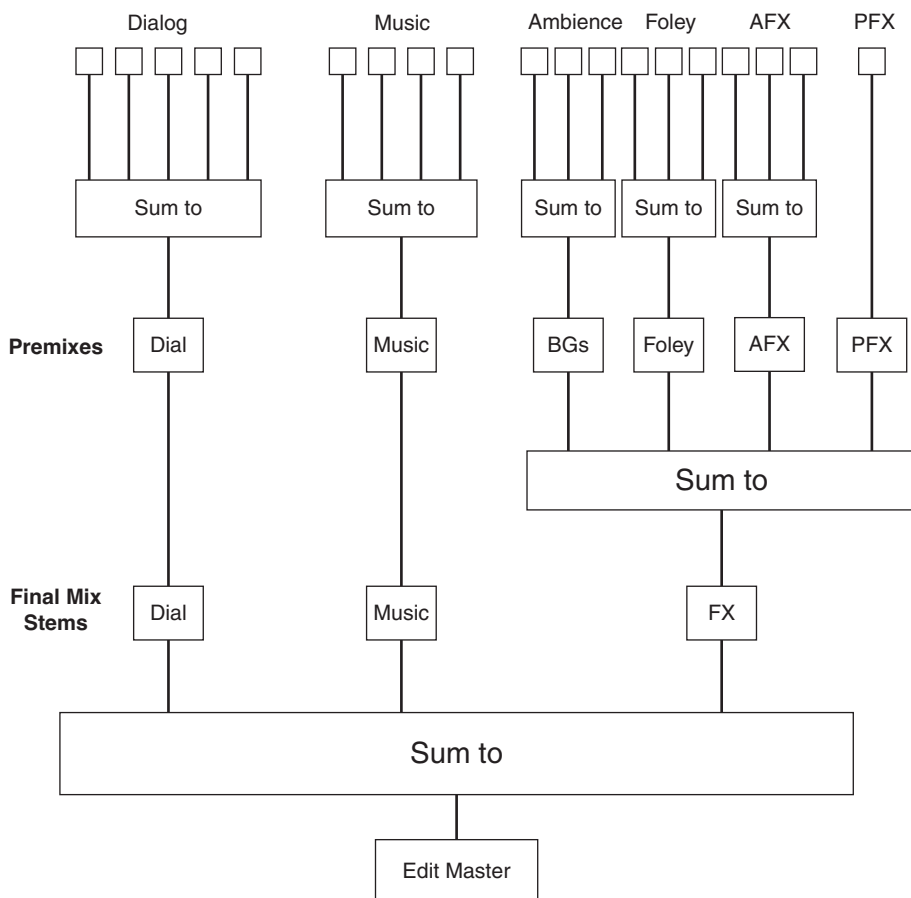


Figure 10-2 Block diagram of the workflow for mixing. Each line could represent 1 to 5 channels of audio.

correct in the context of playing against the others. Final mixing becomes rather straightforward, and, if everything has gone well, no adjustments may be needed at all.

Audio Processes

Audio processes may be broken down into the following domains over which they have the most effect: level, frequency, and time.

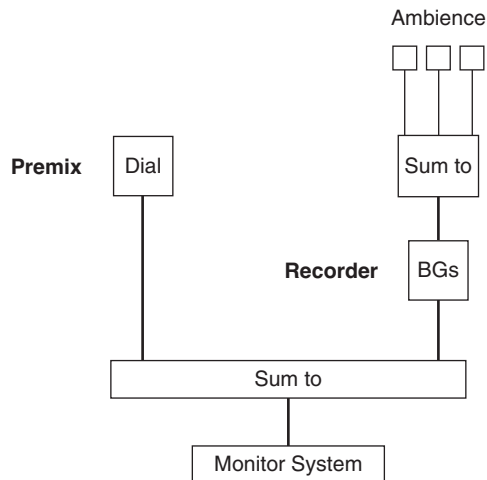


Figure 10-3 Block diagram of the signal flow for mix in context mixing.

Processes Related Mainly to Level

Level Controls

Among all processes, this is listed first, because doubtlessly the mixer's first job is to set both the absolute level and the balance of level among the tracks. In speaking of absolute level, it is remarkable how much agreement there is on the right dialogue level when even neophytes are confronted with the task of setting it, so long as the monitor environment is calibrated. Groups of students given the task of setting the right dialogue level routinely are able to do this within ± 2 dB from person to person, so we know that there is a right level for dialogue, and it's not all just to be left to taste. However, this experiment only works in an environment where the monitor level is calibrated and all meters are covered up. Mixing for the meter is one way to do it, but not a very good way, because peak meters especially don't tell you much about loudness.

Relative level among the tracks has a lot to do with the style of the program. An action-adventure movie might have peaks to full scale and passages that don't exceed -50 dBFS, whereas a dialogue-driven comedy might not use the top 6 to 10 dB of the dynamic range at all and rarely go below -40 dBFS. In the music industry, it would be considered bad practice not to use up all the bits (i.e., peak at full scale) for each piece of

program material (all of the time in some instances!), but the major problem with this approach is that it leads to program material that is not interchangeable: the loudness will vary dramatically from program to program as the peak-to-average ratio dictates. This means that with the highest peak of the program just touching 0 dBFS, but with different peak-to-loudness ratios of the program, the playback loudness will vary, potentially dramatically. A close-miked hammer transient at the start of a xylophone note exceeds the running ongoing level by some 40 dB. If we capture the hammer transient cleanly on a digital recording, then put its peak level at 0 dBFS, its average level will be around -40 dB, and it will play rather softly. This leads to a preference for recording xylophones, for instance, on analog tape machines, where the relatively benign saturation of the analog medium compresses the hammer transients and allows the average level to be raised without noticeable distortion.

Today it is all right not to use full-scale recording level anywhere within a program. In fact, some customers, such as television networks, may not even allow recording to full scale, because the nearly universally accepted reference level standard (SMPTE) is -20 dBFS, and the network may have only 10 dB headroom in some distribution channels. So some programs may contain no recorded level above -10 dBFS and still be considered good. In the early days of digital audio, this may have been a problem because keeping the recorded level up helped conceal low-level problems in converters, but 20 years into the digital revolution we can afford not to use the top of the range for some programs in order to achieve greater interchangeability of program material, without excessively exposing noise or low-level converter problems.

Take two movies: *Driving Miss Daisy* and *Terminator 2*. What they have in common is that the loudness of ordinary face-to-face dialogue is the same, but *Terminator 2* uses a lot more of the available headroom than does *Driving Miss Daisy* in its big action scenes. If we were to follow music industry practice and record the greatest peak of *Driving Miss Daisy* to full scale, then its dialogue would play back much louder than that of *Terminator 2*, a kind of absurdity. So we leave some unused headroom in *Driving Miss Daisy*, and then it is much more likely to get played back correctly in theaters and even homes, because both can be played at the same volume control setting. How to handle the wide dynamic range of T2 for home playback is described in the next chapter as a part of mastering dynamic range control.

Gain Staging

A concern in any complex system that includes multiple level controls is how the relative levels among them are worked out. If a gain control early in the chain is turned down for some reason, and the gain is restored later in the chain, the result is that noise between the two points is exaggerated. The converse is even worse. If a gain control early in the chain is set so high as to produce clipping, then no amount of subsequent level reduction will fix the clipping. This is more of a concern in production, where things are uncontrolled and unrepeatable without at least added expense, than in postproduction where there is an opportunity to optimize things.

Hand Compression

Observing some of the best mixers in Hollywood shows their hands to be always at work, smoothing out dialogue, fitting sound effects to it, and so forth. While compressors, limiters, and other dynamics processes are useful, there is nothing so good as a well-trained mixer with a fader in hand. While this seems not to be a purist approach (shouldn't we leave things alone?), in fact it might be more purist than not riding gain.¹

Good postproduction mixers ride gain, and they do a better job of smoothing out level variations than any device, because they can anticipate action and follow the intensity of the script in a way that no equipment can. That is not to say that these adjustments are large or noticeable: in fact,

¹ Here's why. Voice recordings are made at a single point in space, with a boom or lavalier microphone. Speech has a rapidly changing spectrum, with changing frequency elements from moment to moment. The recorded level picked up at a single point varies with the various frequency elements in the source, as a result of interacting with the standing waves in a room. In fact, frequency-by-frequency, from one point to another in a room, the level varies by ± 15 dB. When we are present in such a room, we hear nothing like this level variation, because voice is a broadband source and so strikes many frequencies simultaneously, which tends to average things out, and principally, because we don't listen at one point in space, but at two, with a thing between the points called a head. The statistical level variation is rather less for two-eared listening than for one-microphone listening. Thus the microphone hears the timbre variations caused by standing waves rather more strongly than natural listening, and the level variations are greater. At least this is our explanation why it is observed that a little compression actually sounds more like a talker than having none at all.

they are typically small, such as ± 3 dB, and not noticeable. That's the whole point. This does not take away from the dynamic range portrayed by the actor, because his or her performance is affected by many factors, not the least of which is timbre. Timbre relates only indirectly to level. It is why we can't turn up a whisper and turn it into a fully voiced speech. So the range from whispering to shouting may be actually rather less in level variation than it would appear to be at first glance, because the timbre of the performance says "whisper" to a listener as much as anything. This is well known in theatrical circles, where a whisper that is audible in the third balcony without benefit of amplification was possible for classical actors: it is more a matter of timbre than of level.

Compression

Audio compression may prove useful too, especially in particular situations. Compression is use of an automatic gain control. Such plug-ins or hardware boxes have a variety of controls that affect the range of level over which they are working, the amount of compression, the time effects of compression, and so forth. Here are some controls that may occur on specific models:

- *A threshold control.* This determines the beginning (i.e., lowest) level at which the system starts compressing.
- *A ratio control.* Above threshold, a ratio control determines the amount of compression, such as 2 dB in for 1 dB out, 2:1.
- *Input level, output level, and gain makeup controls.* Changing the input and output gains affects how much of the signal is above or below threshold, similar to the threshold control in function. Gain makeup controls are used because the natural course of compression has the overall effect of lowering the level. Gain makeup turns it back up after the process.
- *Attack and release time controls.* These are used to set the dynamic properties of how fast things work. With very fast times, short sounds that don't sound very loud control the level, but such times prevent any possible overmodulation of a following medium. Short release times demonstrate fewer effects that might become audible, such as a changing gain on a background after a main event

has passed. However, short release times also have the potential for distorting the signal, because any quick gain change actually changes the waveform. These controls are often variable, as they are widely used on things like drum kits to achieve a particular tight sound. On voice, the best attack and release times are likely to be in the tens of milliseconds for attack and the hundreds for release time, to be most effective. On background music, these times might be even longer in the case of keeping, say, classical music to a steady level. Short attack and release times may be most useful on Foley footsteps, for example, since we are trying to tame a short, sharp sound.

Compression is used, especially during mixing, in instances where there is no time to rehearse, particularly on dialogue tracks. It is also useful on tracks whose dynamic range must be tamed. Let us say we have an orchestral score that must play under a scene, but that has been recorded with a wide volume range. If played under, there will be times when the orchestra is too loud and interferes with the dialogue, and other times when it seems to drop out. Using compression is an aid to keeping it smoothly under the dialogue as accompanying music.

Compression used on specific tracks before they are combined into the mix is much more effective, and more easily hidden, than compressing tracks after mixing. That's because of artifacts that occur when multiple elements are simultaneously present at the compression controller. We hear dialogue pumping the level of background music, for instance, if both are present and the dialogue level range is controlling the compressor.

The amount of compression may be indicated by a display that shows the amount of gain reduction. Typical amounts are in the 6 to 10 dB range of maximum gain reduction, for good audible performance. Going beyond 10 dB usually leads to audible consequences, heard as gain pumping, a kind of breathing of the level that can be heard most readily on vowels in speech.

Limiting

Limiters have the same basic configuration as compressors, but their ratio control is set very high, like 100:1 or ∞ :1. With limiting, no audio is permitted through that exceeds the threshold level, because for every decibel of input change, the gain is turned down by the same amount.

Thus limiters cause an upper limit to recorded level on a track. They have three principal uses:

1. *To contain the maximum level, for use on a medium that may not be able to handle 20 dB of headroom.*
2. *To control the level of certain sounds that might otherwise interfere excessively with foreground action.* Such is the case for Foley footsteps where one or two stick out in the mix. They could be volume graphed down to the proper balance, but it may be easier to limit them. This could also be true of certain other kinds of sound effects.
3. *To put a lid on the maximum dialogue level.* Let us say we have two characters interacting, one at a normal level and one with a raised voice. A compressor might do to even out the performance (with the raised voice still sounding raised because of the timbre difference that raising the voice causes), but a limiter may be effective too. In fact, both might be useful simultaneously, as first the compressor works over a particular range of levels and then a limiter puts a maximum lid on the level.

De-esser

A special form of limiter is known as a de-esser. The “S’s” in speech cause problems for various media, usually downstream of the DAW. For instance, some older satellite broadcasting uses strong pre-emphasis and de-emphasis in the transmitter and receiver, respectively. This means that the high frequencies are turned up a lot in transmission and cut back in reception by an equal amount. The result is flat frequency response, but hiding a lot of noise that would otherwise be audible. The consequence of pre- and de-emphasis is that the headroom of the broadcast channel is much worse at high frequencies than in the midrange. Strong “S’s” will splatter and distort. Thus it is standard good practice to de-ess the dialogue. This is a limiter process that responds only to high frequencies in the content, so it does not affect most of the dialogue but just turns the “S’s” down. Actually, a good process will do this transparently, and you won’t notice a lack of highs in the dialogue. Controls are similar to a limiter. An amount of de-essing in the range of 6 to 10 dB is usually audibly transparent.

Noise Gate

A noise gate is a device that turns on the signal only when it exceeds an adjustable threshold in level. Noise gating may be useful to conceal

changes in background level from clip to clip in a dialogue channel, although using it alone to process sound, with nothing else to cover the resulting holes, would cause problems. Using noise gating to conceal the background changes, and then adding ambience to make a smooth background, may well work to ensure good continuity. Noise gates may have controls similar to compressors, such as principally a threshold control, among others. The object is to set the threshold control so it has minimal effect on dialogue but conceals the background noise. Attack and release times may be adjustable too, with the object being to conceal the noise but allow through the voice without hearing noise tails at each pause in the speech. The attack time has to be fairly fast, too, to not affect the timbre of the speech through altering its attack characteristics.

Downward Expander

Less abrupt than the noise gate are a variety of downward expanders, of varying complexity, all meant to reduce the effects of background noise. The simplest of these are like a noise gate, but instead of turning the signal fully off they simply reduce the lower level signals. The advantage is that while the noise is reduced, no separate ambience need necessarily be cut to cover the holes in the sound. There are also devices that work in both level and frequency dimensions, as described later.

Processes Related Mainly to Frequency

Equalization

Second only to gain controls are equalizers in their relative importance to most mixes. These are more sophisticated types of tone control, the basic idea of which is familiar from home and car sound systems. Equalizers are first characterized by the number of bands they control. Whereas typical home stereo tone controls are just bass and treble, professional equalizers are usually of four or more bands, such as bass, midbass, midrange, and treble. Thus equalizers are used to affect principally the timbre of sources by making separate adjustments to level in a particular frequency range. More sophisticated equalizers may have more frequency bands, adjustable frequency for each band, and adjustable sharpness for each band, called Q. All of these are meant to increase the precision over which control can be exercised over timbre.

Learning how to adjust equalization is a great deal of learning what mixing is about. One way to study it is to adjust a pink noise source (available

from the Generator function of most DAWs). You can arrange two different channels to have two different equalizers. Then have someone set one of them to serve as an unknown and conceal its controls. Now switch freely back and forth between the unknown channel and the second over which you have control. Equalize the known channel to match the unknown, by ear. This exercise is valuable in characterizing the voice with which each frequency range is identified. Using pink noise makes the job easier and harder: easier because it is a steady sound unlike the changing timbre of real sources, and harder because by containing every frequency and being continuously present, it is revealing of differences in timbre. Useful hours may be spent with this exercise.

Another way of looking at equalization is by its effect in subjective terms. A way to characterize the effect of equalizing various frequency ranges is given in Table 10-1.

A common form of equalization for well-recorded boom microphone work is to pull out some midbass. That’s because rooms emphasize this region through standing waves. This effect is most obvious in small rooms, like bathrooms, which is why people like to sing in the shower: the bass frequencies are emphasized. So cutting them back in postproduction by dipping this region can be effective.

If a great deal of windscreening has been used on a boom mic, like the combination of a silk windscreen and a Windjammer, highs may have been lost, calling for some high-frequency boost.

Table 10-1 Characteristic effects of equalization in various frequency ranges.

	Frequency Region			
	Low Bass	Midbass	Midtreble	High Treble
Increases in this range	Powerful, boomy, woofy	Emphasis on uhh uhh utterance; rooms often emphasize	Greater presence, honky	Bright, sizzly
Decreases in this range	Thin, weak	Thin on male voices; cut frequently needed due to room acoustics	Greater distance, lacks intelligibility	Dull

Equalizing lavalieres is probably the trickiest problem, especially if they have to be made to match a boom. It can be done, but it is a difficult task because several effects are simultaneously in play:

- The position of the microphone tends to lose high frequencies, as if the lips are shadowed by the chin. Most lavalieres have built-in high-frequency boost to overcome this loss, but the shape of this equalization may not be a particularly good match for any individual mic location chosen. Thus the high-frequency sound may be anything from dull to bright, depending on the interaction of the particular lavalier's frequency response and placement.
- For fiction filmmaking, the lavalier is often buried under clothing, resulting in a loss of highs.
- There is a resonance recorded by placement on the chest that emphasizes the region around 630 Hz. So a 630-Hz equalization dip is usually useful.
- The 2-kHz region is most responsible for the sensation of presence, yet this region can be lacking because of the lavalier placement and thus may need emphasis.
- Lavalier design sometimes de-emphasizes midbass frequencies, in a search for clarity. Thus male voices can sound a bit thin. It is uncommon to boost bass, because it often gets one into trouble with noise, but a little bass boost around 200 Hz can be useful to make males sound more like themselves.

There is an analytical way to equalize lavalieres. However, this method is not yet practical for day-to-day operations. It involves measuring the spectrum from a boom mic and a lavalier and comparing them, and then equalizing the lav to match the boom. An example of curves found this way is given in the figures 10-4 and 10-5, which, for several conditions, is an exact solution, with the timbre of the lav and boom matching after this equalization is performed. Note that the response shown is largely a function of placement on the chest and is not a reflection on the free-field response of the microphone. For raw microphone response curves for a variety of lavalier microphones including the one used here, consult www.microphone-data.com.

The curve in Figure 10-4 is the response of the microphone, so each dip has to be equalized with a corresponding peak, and each peak in the response

with an equalizer dip, to achieve the response shown in Figure 10-5. Even though this is just one microphone model on one person in one acoustic environment, nonetheless it offers guidance for where equalization will be the most useful. Other researchers have found the 630-Hz centered peak, for instance, which tends to make the sound chesty, and a corresponding dip in that range is quite helpful. Likewise, to a slightly lesser degree, the 3-kHz dip can be equalized with a peak. (The 3-kHz dip is not quite as audible as the 630-Hz peak because of the differing nature of peaks versus dips: for the same dB amount and frequency range, a peak is more audible than a dip.)

Filters

The effect of equalizers is generally broad across frequency, which is most useful for adjusting the overall timbre of an element. However, there are also sounds we want to exclude, such as low-frequency rumbles that change at cuts. For such clean-up work, filters are more appropriate because their change across frequency is usually more abrupt than that of equalizers, so they do a better job of separating close-by frequencies for inclusion or exclusion in a mix. The various filters are as follows:

- *The high-pass filter, also known as a low-cut filter.* This type passes all frequencies above its cutoff. A common frequency is 80 Hz for dialogue channels, in use virtually all of the time because it won't damage any but James Earl Jones type male voices, while it eliminates noise. If only female voices are heard in the channel, the frequency can be increased to perhaps 160 Hz. Besides frequency, the only other parameter might be slope, the amount of attenuation per octave. If this is available, the steepest slope is usually best for dealing with the issues of film sound tracks.
- *The low-pass filter, also known as a high-cut filter.* This cuts all frequencies above the stated one. A common frequency for dialogue channels is in the range of 8 to 10 kHz. It is useful because matches at edits at very high frequencies are problematic, and almost all of the voice energy lies below this frequency. It also helps, when an eventual optical track or television broadcast is going to be made, to limit the amount of very high frequencies in the material. It would be unusual, on the other hand, to use a low-pass filter on music. It may be useful on sound effects, to strip off undesired parts of the sound, where it may be used down to a fairly low frequency.

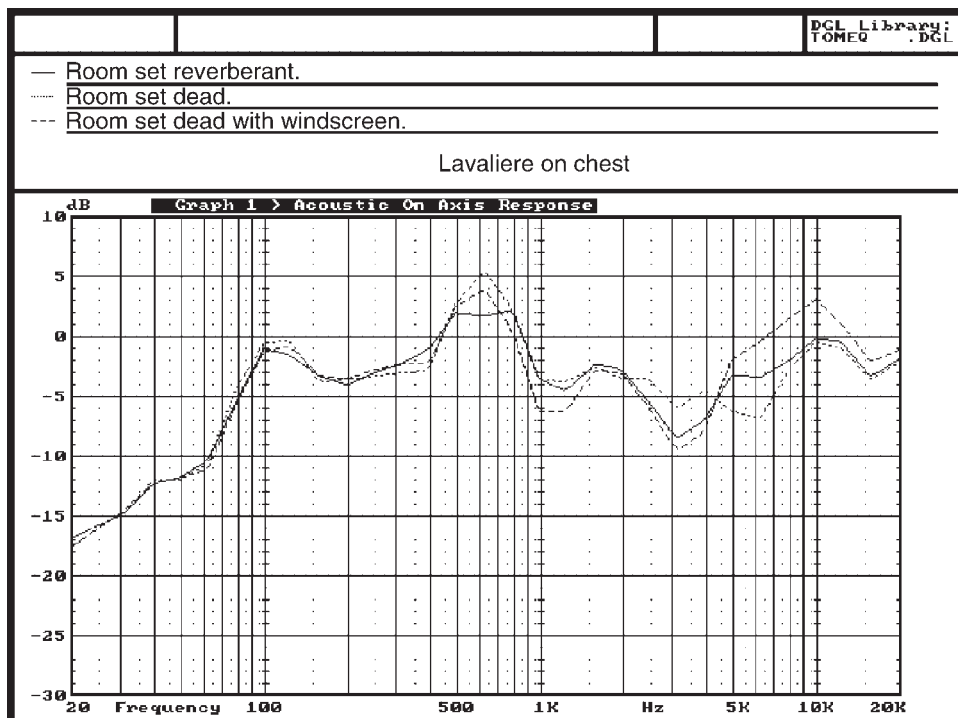


Figure 10-4 Unequalized response of a professional lavalier microphone (Tram TR50) on the chest. Curve made by comparing long-term spectral average response of this mic versus a reference measurement microphone at 1 meter on axis.²

- *The bandpass filter.* A combination of high- and low-pass filters commonly used to simulate a telephone. The frequency range limited to 250 Hz to 2 kHz sounds reasonably like a telephone, especially when compared to full-range voice.

² The variation in responses is caused by differing conditions in the room and wind-screening. This experiment was conducted with several microphone types and positions, and the room variations are far more important in more distant mic positions than for the lavalier. In this case, the variation seems mostly to be caused by small differences in mounting the lavalier from one section of the experiment to another, and possibly the angle of the head of the talker. The variation shows how difficult it is to get a match for lavaliers, but also shows that a certain fixed equalization seems useful, with some variability for specific conditions. This work was published as "Improving Microphone Equalization for Dialogue," *SMPTE Journal*, vol. 108, no. 8, August 1999, pg. 563–567.

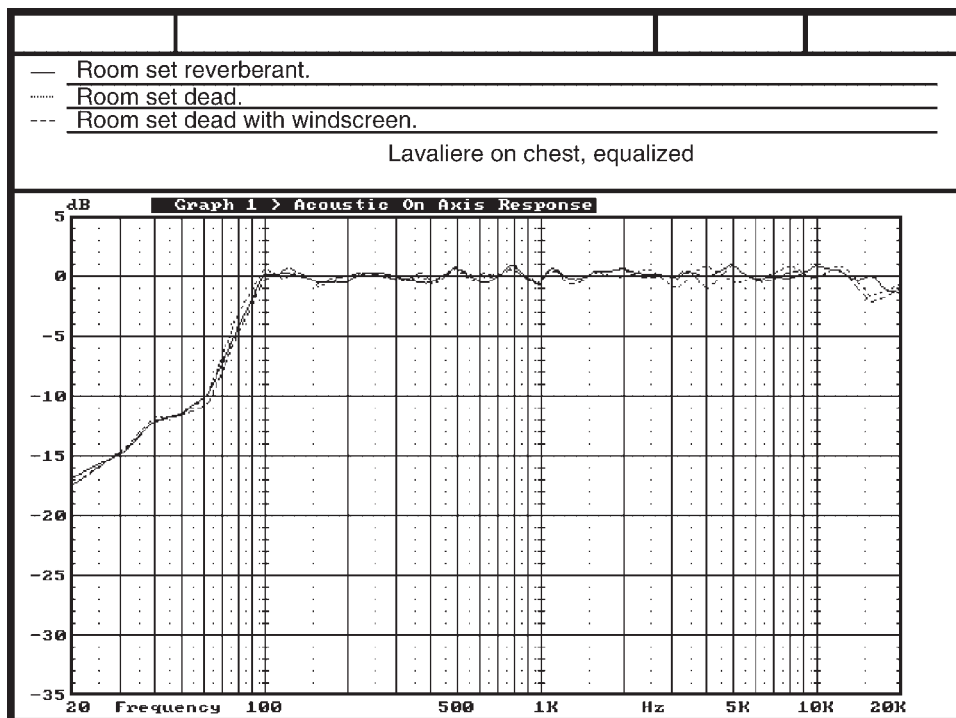


Figure 10-5 Equalized response of the same microphone. The peaks and dips over the range from 100 Hz to 15 kHz of the microphone condition shown in Figure 10-4 have been equalized out with a 10-band parametric equalizer plug-in (Waves Q10).

- *The notch filter.* This cuts out one narrow part of the frequency spectrum, which is useful to attenuate a tonal sound that could be annoying. Some equalizers have such a range of adjustment that they can be considered to be notch filters when turned fully to their sharpest setting (highest Q).
- *Hum and buzz reduction.* In the United States, 60-Hz hum and its harmonics at 120, 180 Hz, and so on are bothersome because they reveal stray electrical fields that may have been induced into the signal along its way. Also, devices that run from the power line, like fluorescent lights, can make mechanical noise at these frequencies. A single-notch filter will only attenuate one frequency component, which is only occasionally what is needed for such problems. A multidip filter at the fundamental and its harmonics is better.

Combinations of Level and Frequency

Some plug-ins combine features working on both level and frequency domains simultaneously. Many are aimed at specific problem areas, particularly in production sound recordings. Among them are the following:

- *Broadband noise reduction systems.* Utilizing multiple frequency bands, these systems rely on the fact that the desired signal is above the noise level, and discriminate between the two, applying downward expansion to reduce the noise. Some of them can be fed a background noise passage for analysis and will then automatically set the threshold level in multiple bands.
- *Click and crackle removal.* Phonograph records exhibit ticks from dust. Bad mic cables may crackle and leave gaps in the audio. These software plug-ins attempt to smooth over such abrupt transients and substitute best guesses, perhaps human aided, at the underlying desired signal.

These kinds of device range from relatively simple plug-ins, through ones that cost many thousands of dollars, to outboard pieces of equipment such as the Cedar DNS-2000. The effectiveness of each of the methods varies, depending on the type of background noise to be reduced. Use of such devices is widespread in high-end postproduction. In fact, over the years there has been a steady increase in using tracks that would previously have been considered too noisy to fix.

Time-Based Devices

Reverberation

Reverberation plug-ins and outboard equipment to generate reverb are used to add to original recordings to set the space for sound to exist in. Various plug-ins and devices offer several different adjustments, such as the following:

- *Reverberation time.* Classically defined as the time it takes for an abruptly stopped signal to decay by 60 dB (called RT60), or, for practical purposes, to inaudibility. This is the basic adjustment that we associate with size of a space being portrayed. Ordinary small rooms

might have 0.5 s RT60, whereas a concert hall could have 2.0 s and a cathedral 5.0 s or more.

- *Initial onset time.* This is a delay between the incoming sound and the start of reverberation, and it also tells the ear something about the size of the space, longer times indicating a larger space.
- *Adjustments across frequency.* Most rooms have longer reverberation time in the bass than the treble, and this factor becomes larger in bigger rooms. Most reverberators thus provide means to adjust a difference in RT60 across frequency.

Reverberators can be used on each element of a sound track, although not all reverberators are set to the same parameters. From the driest type of recording with no reverberation, to the wettest, the most reverberation, here are some items that use varying amounts and reverb times:

- *Dry sound.* Utterly without reverberation, whispered narration (*sotto voce*: the whisper that reaches the back of the house in theater). Example: voice-over in *Apocalypse Now*.
- *Room sound.* Production recordings contain some room sound according to the location. A task in postproduction is to come up with a reverberator that matches that reverberation, for use in matching lavalier to boom recordings and ADR to reference boom miked recordings, depending on the coverage of the scene. Foley recordings and other on-screen effects may also use a reverberator to make it seem more as if the sound is associated with the right setting of the picture.
- *Off-screen effects.* Occasionally, off-screen effects are more reverberant than on-screen, as though heard from a distance. Example: Indy meets the girl in a portico above a stadium in which a book burning is being performed by the Nazis in *Indiana Jones and the Last Crusade*.
- *Music.* Traditional large-scale orchestral score is often the most reverberant element of a sound track, because classical music is traditionally heard in concert halls with fairly long reverberation times.

By automating the plug-in's reverberation time and/or the faders controlling the amount of reverberation reproduced along with the dry, non-reverberated signal, an acoustic zoom can be performed. Note that the use of reverberation is a way to introduce a sense of space even for monaural productions. On the other hand, for multichannel productions each reverberated signal has potential reverb returns to all the channels, so the question becomes the balance among them. There are two classical ways to do this: return the reverberation device's outputs just to the screen channels, or to both the screen and the surround channels. The former gives us a view of listening into a space portrayed in front of us, and the latter envelopes us and places us in a shared space with the screen. There is no one right answer, and the answer may well vary from time to time within a show, but it is always a consideration.

An unusual example of the use of reverberation in storytelling is the opening of *Midnight Cowboy*. It starts with a reverie by Joe Buck of being in an open-air drive-in and singing. Although it is open air, the sound is reverberant—this is a dream, how he remembers it. When the picture cuts to him in the shower singing the same song, in continuous time, the reverberation is cut out. This reversal of the normal order of things—outside reverberant, singing in the shower less reverberant—helps establish this sequence as montage, before the title. Interestingly, the idea is carried through with audio panning a moment later, when a chorus of voices asks: “Where is that Joe Buck?” from one direction and another, an unusual thing for dialogue to do. When the real dialogue starts a minute later (after the song), it is normal, centered on screen.

Other Time-Based Effects

A wide variety of effects is possible once a signal has been digitized. A simple, single delay may cause an added thickness through comb filtering up through a discrete echo, depending on the time of the delay. Multiple delays lead to curious effects, potentially useful to modify voices or effects. Some outboard devices, operating mainly in the time domain, offer many dozens of effects that are useful in modifying voices, sound effects, and music. Plug-ins are available in the hundreds, with active online communities built up around them. The areas covered include synthesizers (making sound from scratch), samplers (capturing, storing, and manipulating sound), and DAW processing plug-ins. The issues to be concerned with in order to use these are the environment for which the plug-in was designed, such as AU, AudioSuite, DirectX, HTDM, MAS, Premiere, RTAS, TDM, or VST. See Table 9-2 in Chapter 9 for more information.

Other Plug-Ins

A wide range of other plug-ins is available, many of which may see only occasional use in editing/mixing for picture. The following are ones that are highly valuable when their specialized functions are needed:

- *Dither*. Low level signals may suffer audible artifacts caused by listeners hearing the steps composing the digital signal--a particular kind of distortion that increases with decreasing level. This problem can be completely overcome using dither, a small amount of deliberately added noise. With the right type and frequency spectrum of noise (with equalization applied to it), its addition is benign insofar as audible noise is concerned, and low-level sounds remain undistorted. Although it is a complex topic, most sources that are used in sound for picture have enough associated noise to dither the output conversion. For wide dynamic range sources, it is a good idea to use a dither plug-in as the very last thing in the chain (such as used as a plug-in for the output channel master) with exactly correct gain (fader set to 0 dB [unity gain]) following the plug-in. The dither amount should be set to the number of bits expected in the least resolution system. For instance, if the output recording will be made to CD, then dither should be set to 16 bit.
- *Generator*. Useful in calibrating systems, they usually offer a variety of test signals, such as sine waves for head line-up tones of edit masters and level setting throughout a system as described in Chapter 11, and pink noise for aligning sound systems acoustically. Defining the level of pink noise is a bit tricky. You want noise that's at exactly the same long-term level as a sine wave, say, of -20 dBFS. However, such pink noise will read about 10 dB higher on a peak meter than the sine wave of the same long-term or rms level. The Pro Tools generator function does this correctly, setting the level for long-term value rather than short-term. Thus the peak level seen on its meters goes up when switching from sine to pink, but this is correct; the pink noise is at the right level. Also interesting is that the pink noise will sound louder, which seems to correspond to the meter reading, but this is actually true for a more obscure reason. By reproducing a wider frequency range, pink noise stimulates hearing more, and thus sounds louder than a sine wave of the same level.
- *Pitch correction*. By feeding in a MIDI file, typically used by synthesizers for control, a performance on one track can be compared to the correct pitch and pitch-shifted moment-by-moment to get it in tune.

Other plug-ins are described in Chapter 9, those that are perhaps more frequently used by editors, although they may be necessary for mixing too.

Panning

Panning bridges the domains processing and routing, which is described next. Panners take in typically one channel and redistribute its signal among multiple output channels, such as the five: left, center, right, left surround, and right surround. The panning law or rule by which the energy is redistributed is that exactly half way between any adjacent pair each will be fed the same signal attenuated by 3 dB (equal power panning), in normal operation. An additional control called *divergence* tends to spread the content out among all of the channels, not just the adjacent pair where the sound is panned, with it ultimately leading, when fully turned up, to the same signal in all five channels. The idea behind divergence is to thicken up or broaden a source, but in fact it doesn't work very well to do that, creating comb filters, and often a sense that it's coming at you from multiple directions, not a fuzzy cloud that's expected when divergence is fully up. Some chorus and other programs of reverberation devices or plug-ins do a better job at thickening up sound, and a surround decoder may be useful in decoding two-channel stereo into more channels. For some kinds of non-tonal sounds, slight pitch shifting inserted into multiple channels will make the source seem bigger.

Routing, and Limitations Caused by Routing

Busses, Channels

The basic idea of busses is described in Chapter 9, as well as their use in mixing and monitoring, and auxiliary uses. Hardware inputs are connected to edit workstation tracks through input routing by means of input busses (Fig. 10-6). Tracks are connected to the outputs by way of output busses, or they may be routed through an intermediate path to do more processing by way of an auxiliary bus. An aux bus has its input from one or more tracks, and outputs to an output bus, providing common signal processing for several channels by way of plug-ins in the aux track, for instance.

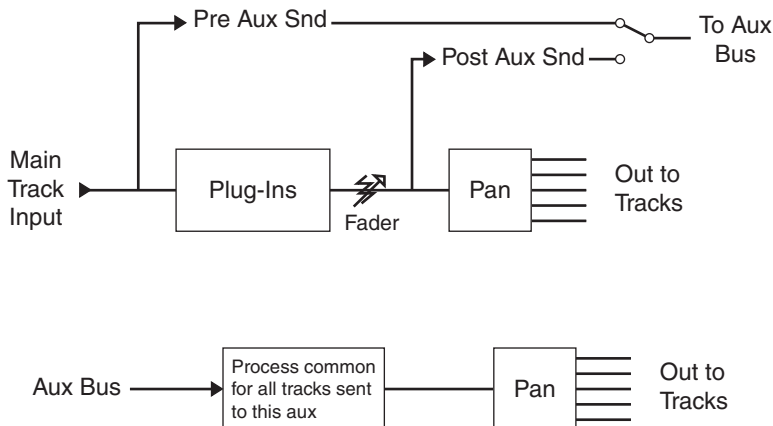


Figure 10-6 A block diagram including inputs, outputs, and main and auxiliary busses.

An aux bus may also be routed to external equipment, such as more complex reverberation devices than can run as plug-ins on conventional workstations. In this case, it has an input from the tracks, a send (by way of an output channel, either analog or digital) to the outboard equipment, a return from the outboard equipment (by way of an input channel, either analog or digital), and an output assignment to a track.

Delay Compensation

A consequence of digital audio processing is associated delay. Audio conversion and signal processing both cause delay. If a signal is routed by two routes, one for extra processing such as in an aux-send/aux-return situation and one more directly, the signal encountering the added path will be delayed relative to the shorter path. The consequence of adding two signals together with one delayed relative to the other varies, depending on the delay time. For short delays, a notch filter will be caused at high frequencies. For longer times, comb filtering occurs, which once heard is an obvious coloration. System designers know about such delay and take various measures to overcome it. One is to put in all the delay one would ever think necessary in every path and then remove the delay as plug-ins, bus routes, and so on are patched in according to the length of delay that the item being installed represents. This system works all right, but it consumes system resources, at the least.

Voice Limitation

Only a limited number of tracks, busses, and input/output channels are available for use at one time. Another limitation is caused internally by the number of sounds that can be reproduced simultaneously. In a system where almost everything is virtual, this limitation is a little more obscure than the others. Called the number of *voices* available, a term borrowed from synthesizer design, this number represents the number of sounds that can be played simultaneously. This may be fewer than the number of tracks cut or the number of input/output busses available. It depends generally on the amount of computer horsepower available in a given configuration of machine.

How to Mix

While there are many ways to mix, one way to approach the problem of conventional dialogue, music, and effects mixing is as follows. Clearly this is an area where the particular type of program can change the order of doing things, the time available to do them, and the specificity with which they can be done for a given budget. It is assumed that you have calibrated your monitor system according to the instructions in Chapter 11. A valuable exercise is then to listen to good feature-film mixes of the style you are working in over the system, to begin to understand the desired timbre for speech, its balance with other elements, and so forth.

That having been said, here is guidance on a way to proceed, perhaps most of the time:

1. *Mix the dialogue first.* It usually tells the story and is essential for the understanding of the audience. Remember that your audience usually hears the program only once, whereas you may have heard it hundreds of times. Something that happens in feature-film mixes is that the directors have heard the dialogue so many thousands of times during editing that they can recite it in their sleep, and what they are looking for in the mix is a new, heightened experience; so, left to their own devices, they will slather sound effects and music liberally over all of the dialogue. The competent lead mixer will resist this trend, making sure that he or she stands in the shoes of the audience, which is trying to understand the words on the first pass.

Mixing dialogue means mixing the dialogue stem. That includes intercut presence, wild lines, ADR, and so on. The editing should already have provided a complete track, with no holes, or at least without holes that won't be covered up by other sound effects. There should also be no bumps, undesired discontinuities in the tracks. Editorial principles such as filling out split track with presence should have been followed.

It does not necessarily mean mixing production effects, called PFX, which, although recorded along with the dialogue during production, may have been pulled out to sound effects tracks in order to exercise more control in the final mix over the parts.

With SMPTE reference level of -20 dBFS, and video level calibration, you are likely to see dialogue read peaks of about -15 to -10 dBFS on peak meters.

The order of signal processes matters. While conventional processes such as equalization and filtering can be exchanged in order without consequence, dynamic processes perform differently, depending on their order. Here is an order that will normally work:

- a. Clean up first. Apply high- and low-pass filters to avoid undesired noise and changes at edits in production sound tracks. A normal high-pass filter that won't damage most male voices is 80 Hz. A low-pass filter is a little less necessary, but can be valuable, say, set to 8 to 10 kHz. Apply a notch filter for a discrete tonal noise (fairly rare), or a hum/buzz plug-in if that is a problem.
- b. Next perform broadband noise reduction with a plug-in or by routing to an outboard signal processor and back into the signal flow. Occasionally you might want to follow this with a noise gate. These devices will help reduce noise from traffic, airplane flyovers, camera noise, and other sources of broadband (non-discrete tonal) noise.
- c. Equalize for good timbre on voices. Lavalier tracks will typically require more equalization than boom mic ones. If the editor has separately equalized clips or regions so that they work more smoothly, be certain that the automation of the plug-in you are using does not interfere with this equalization. For instance, in Pro Tools, AudioSuite plug-ins applied to whole regions create new audio regions that are inserted instead of the original during editing; then, during mixing,

an automated TDM plug-in can be used to set the overall equalization, and even dynamic equalization, as required.

- d. The level control is here in the chain typically, but is the last used after all of the others have been adjusted correctly. Perform the mix function here. Usually the adjustments will be fairly small, say ± 3 dB if the editing job has the matching right from clip-to-clip.
- e. Compress lightly for better intelligibility in view of other sound sources that will interfere with intelligibility, poorer listening conditions, and the like. This is easy to overdo. Peak gain reduction value should not exceed 10 dB typically. The attack time can be 10 to 50 ms, and the release time 100 to 500 ms. Input level, output level, threshold, and gain makeup controls are adjusted to get a peak gain reduction of 10 dB and a peak recorded level of -15 to -10 dBFS maximum.
- f. De-ess lightly for best compatibility with all downstream recording processes. This is audibly benign but will prevent splattering “S’s” later on in the program’s life. Adjust the de-esser for the correct frequency range of a particular talker’s “S’s,” and adjust the levels, threshold, or other such controls to produce a maximum amount of reduction of about 10 dB. Some plug-ins provide a way to listen to the sound that is controlling the de-esser, that is, the bandpass filtered “S” sound, so that you can tune it to the particular “S” range for a given talker.
- g. Apply peak limiting adjusted to create a maximum level of -10 dBFS, with no more than 10 dB of gain reduction. Excessive peak limiting will probably bring up the effect of background noise as the louder sounds are constrained in level, meaning the softer ones are actually being exaggerated and will cause funny, hard-sounding dialogue.

Some of these processes are likely to need dynamic adjustment, particularly level, equalization, and broadband noise reduction. Thus they need to be dynamic plug-ins (RTAS, TDM, HTDM) in Pro Tools, not the region-oriented AudioSuite ones. Often the compression and de-essing can be set once and do not need to be dynamically varied throughout the mix.

In addition, occasionally reverberators will need to be applied, either as plug-ins or outboard, particularly to ADR tracks to get them to match production sound tracks.

2. *Set up the bussing and routing for mix in context so that you can hear the dialogue premix you’ve made, while working on the next ones.*

3. *Make an ambience premix from the ambience elements.* Ambience may use equalization (with more extreme values than dialogue often useful), level, and so on but would rarely have so complex a patch as a dialogue mix.
4. *Make a Foley premix from the Foley elements.* Here again, stronger equalization may be useful because of the skin, grass, and sky analogy (we only know what color those are for sure, not other objects; we only know what dialogue and some musical instruments sound like, on the whole). Peak limiting can be helpful so that particular Foley events don't stick out too much when the average level of Foley is right.
5. *Make one or more hard effects mixes.* If you make more than one, the principle to apply is to checkerboard the several mixes (i.e., with sound first in one, then another, along the timeline). This helps in going back and making corrections later. PFX is one of the multiple hard effects premixes. In simpler productions, all of the hard effects can be mixed together.
6. *In performing the foregoing, you add each of the finished mixes to the mix in context monitor busses as they are complete.* At the end, you have a master mix, with each of the premixes having the right level relationship to the dialogue premix.

Mixing is affected in the number of tracks, routing, and panning, by the required format of the final master, two-channel or 5.1, as described in the next chapter. Chapter 11 also has a great deal of information on mastering for overall level, which is a part of mixing. It is covered under Mastering because the overall recorded level often depends on the target delivery medium.

Several operational matters principally affect mixing. For many years, sound systems for mixing have been slaved to picture sources, such as tape machines. The whole speed of operations was controlled by the dynamics of the deck transport, how fast it could rewind, and so forth. Once rewound and playing, time code from the source machine was used as control to lock up the audio system and slave to the picture. This process is fairly slow. In recent years, the use of digital picture for mixing has become prominent, so there is essentially no waiting time, and the DAW can contain a picture representation good enough to mix to as a video track, so there is no time spent waiting for sync.

A feature derived from film mixing is not widely available yet on workstations but would be a valuable addition to the working environment, and

that is the ability to mix in reverse. Winding over and over a particular passage permits comparisons at its in- and outpoints with what preceded and follows it, which is particularly useful for equalizing. Although today's systems can loop quickly over a passage, they do not typically run in reverse, which old film dubbers could do. Listening to a voice in reverse, for example, allows one to concentrate on the timbre rather than the content, and this may be a valuable addition in the future to DAWs.

Director's Cut

- The most important part of mixing is getting the various tracks into the right balance of loudness. It is usually useful to start with the dialogue, and then add to it.
- Stem mixing is valuable, because it allows balances to be changed until the end of the process. Stem mixing is keeping dialogue, effects, and music tracks separate until print mastering.
- Mixers employ a wide variety of sound processes: equalization, compression and limiting, noise clean-up devices, and so forth. In some cases the order that the processes are applied matters.
- Today's debate is over how much processing to do editorially and what to leave to the mix. Generally, overall equalization and balances should be left to the mix because that is where the best monitor environment exists, and the training and experience of a mixer is different from that of an editor. What is most useful editorially is to be able to operate on, say, a cough, which will be harder for the mixer to jerk down smoothly in level than it is to do editorially.
- It is most useful to develop your ear for timbre, particularly for dialogue, because that is so important.
- There are issues regarding sound localization within a surround-sound field. Normally dialogue will be centered, unless it is supposed to be off-screen or at the extremes of a frame. This is because dialogue edits that jump around are distracting from the story.

CHAPTER 11

Masters and monitoring

The final output of the postproduction process is a delivery master. While there is a desire among program producers to make a “one size fits all” master, in fact, if the program is intended for both theatrical distribution and video distribution to homes, the differences between the sound systems and environments are large enough that different masters are required for optimal sound. Using a cinema master on a home system requires special playback techniques to be optimal,¹ and while these are available, they aren’t accessible to everyone. Using a home master on a theatrical sound system also requires special playback, which is virtually never available. Thus masters and monitoring are covered here in one chapter because they are interwoven.

Master is a slippery word in film and video production, because just what does it mean? Basically, standing alone it means just about any postproduction recording, anything other than an item of finished goods. In postproduction the word *master* has lost its meaning through overuse—almost everything in sight is a master of one kind or another. To differentiate among the various items used by program makers, we can better use the following terms:

- *Camera original*. This states, more clearly than using the word *master*, that this piece of tape or hard disk file is one onto which the original capture from the camera to the medium was made.
- *Camera original clone*. A digital copy at the full resolution of the original recording. That is as opposed to *camera original copy*, which could

¹ Home THX® uses re-equalization and timbre matching to equalize the octave-to-octave balance of a home system to make it better match the cinema experience when playing movies that have been mastered directly from the theatrical masters without alteration.

even be an analog copy of a digital original; the use of the word *clone* is specific in meaning a bit-for-bit copy.

- *Intermediate master.* This could be a tape or disk file containing typically all of the content at just about any stage of postproduction that is not a camera original or a delivery master. For instance, it could be a work copy of a cut used to export picture from a picture-oriented editing system to a sound-oriented one.
- *Edit master.* The finished program on tape or disk. An edit master typically contains a header consisting of identifying information, color bars and tone, a countdown leader with a 2 pop, followed by the program with accompanying master audio. In the case of 2.0-channel masters, the audio is on the medium. In the case of 5.1-channel masters, only some of the highest-end media have space for the audio, and if there is not enough space, two methods may be used for it. One is to record double-system by putting the audio on an accompanying medium such as DTRS² tape, with time code matching that of the picture so that the two can be locked up by a synchronizer and played simultaneously. The other method is to use a special bit-rate-reduction scheme for such masters, Dolby E (E for Editable), to fit 5.1 channels into the space normally used by two channels of conventional digital audio, and record it on the medium.
- *Delivery master, also called studio master.* This is a master tape or disk file that contains the program in its finished form, plus potentially added content intended for distribution through one channel. Compared to the edit master, it may contain studio trailers, identifying logos, MPAA rating card, FBI warning, and other such types of content at the heads and/or tails of the principal program. The term *channel* in this case means a marketing channel, not a television one, with particular response and level characteristics intended for a specific market, as described later in this chapter.
- *Protection dub.* Any one-to-one copy of any of the above. *Protection clone* is a more descriptive term for a bit-for-bit verified digital copy.

² Special digital audio tape housed in the same type of cassette as Hi-8 tape. Such tapes are often called DA-88, for the model number of the machine that popularized the format, but the more general formal name is DTRS.

Choice of Sound Format on Edit and Delivery Masters

One big question facing digital video producers is what form of spatial sound to choose: mono, stereo, matrix encoded surround, or discrete surround. Mono is used today generally for informational programs aimed at cable and network television and what used to be called the industrials market. Because virtually all media are at least two-channel today, for mono programs the content should be recorded identically in the two channels available on the medium. This will permit stereo systems to reproduce them from both loudspeakers, and will thus render the content acoustically as a phantom image.

Recording mono programs to both channels is good practice because a phantom sound image can be arranged to appear centered on the picture if left and right loudspeakers have been placed to the two sides of it. It sounds strange when, for instance, the output of an Avid workstation set to Channel 1 is recorded on a DVD to two channels. The resulting left-only speaker signal doesn't permit matching the position of the picture—the sound image is always off center to the left. For instance, listen to the alternative ending in the supplementary materials of *Ronin*. Here is a glimpse at the postproduction output from a picture editing system raw, before any sound postproduction has been done, and one of the problems is the left-only recording.

Two-channel stereo can now be seen as a detour in the history of sound recording and reproduction. It arose in the late 1950s because of the two groove walls of the phonograph record, the first way to distribute stereo widely, and thus was cast as the home system, with other media (e.g., FM, cassette, CD, television broadcast, VHS tape) just meeting the requirement to match what was already available. Two loudspeakers became the norm at home, and the chicken-and-egg problem applied to all subsequent media for many years. Actually the beginnings of stereo were multi channel, with three front channels in 1933, to which surround was added about six years later. While early multi channel systems could be chalked up to experimentation, they became widespread for movies in the early 1950s. When the stereo LP was introduced in 1957, practicality called for just two channels to get stereo launched into the home. Today it is hard to buy a stereo receiver—they're virtually all multi channel. So the main reason two-channel stereo survives is as a legacy format for the many existing programs recorded that way. The problems with it include phantom imaging problems (i.e., image pulling for off-center listening and timbre of

centered sound for every listening location) and the fact that it lacks the capability for the enveloping sensation that is a major ingredient of surround sound. Today, to distinguish two-channel stereo from two channels used to convey surround-sound information, the left/right or stereo designation has been renamed LoRo for left only, right only. So a proper labeling for a two-channel stereo master is Program Name—Edit Master—2.0 LoRo. The 2.0 designates there are two front channels and no low-frequency enhancement channel.

Arising from the need to deliver a surround-sound experience over just two channels, amplitude-phase matrix technology first became employed in the ill-fated quad era. It was reworked for film sound by Dolby Labs in the middle 1970s and remains to this day a widely distributed format for film sound. Such tracks are called LtRt, which stands for left total, right total and indicates the fact that this two-channel delivery is intended for decoding into left, center, right, and surround channels. The advantage of this format is that distribution can be by way of the stereo infrastructure built up since the 1950s for both broadcast and packaged media. The disadvantage is that the channels are not in fact perceptually discrete; if they have complex content, your results may vary. For instance, trying to record four voices in four channels and have them come out in the right places continuously just isn't possible with a matrix. On the other hand, for most sounds most of the time, the matrix technology, which has seen further development in recent years, is adequate to convey the surround impression, and mixes made to it fit in a lot more paths to market than those in 5.1. Table 11-1 compares the two systems.

Discrete multi channel sound in the 5.1-channel format has to be said to be the ultimate widely observed standard today. While available on fewer media, its use has come to dominate thinking about film programs, high definition television, and by extension, in some cases, digital video. The advantages of discrete sound are that much more complex mixes can be heard more clearly because there is less spatial interaction of the component parts of the mix, no sound ever winds up in a position where it wasn't put by the mixer, there is a left-right sensation in the surround field that is at least difficult to achieve with matrix techniques, and there is the 0.1 LFE channel to handle higher levels of bass.

The trade-off for most video producers among mono, stereo, matrix surround, and discrete surround probably falls on matrix surround a good deal of the time. The simpler distribution to most listeners, many of whom

Table 11-1 Comparison of surround produced by way of LtRt matrix with 5.1 discrete technologies.

Surround Format	Two-Channel Matrix-Encoded Surround	5.1-Channel Discrete Surround
Tradenames of encoding processes	Dolby Surround	Dolby Digital, AC-3
	Dolby Pro Logic II	dts
	SRS Circle Surround	SDDS (cinema only)
Tradenames of decoding processes	Dolby Surround	Dolby Digital
	Dolby Pro Logic	dts
	Dolby Pro Logic II	
	Dolby Pro Logic IIx	
	dts Neo:6	
Application area	Logic 7	
	Circle Surround	
	To deliver surround sound to the widest possible audience. This means delivery over media limited to two channels, which includes a broader range of possibilities than the more limited distribution formats for 5.1-channel sound, with some compromises.	To deliver a more complex surround sound field over more specialized media than two-channel delivery permits
Primary enabling technologies	Amplitude-phase matrix encoder and decoder: Encoder and decoder use sum-difference matrixes with phase-shift networks to produce a 4:2:4 basic matrix (four input channels to two recorded channels, decoded by the end user's equipment back to four channels), and the decoder typically includes playback delay of the surrounds by 20 ms to help prevent crosstalk. Systems introduced more recently, such as Pro Logic IIx, derive up to 7.1 channels from the LtRt.	Bit-rate-reduction of original linear PCM audio to fit into the digital space available on restrictive delivery media, typically by a factor in the range of 8.6:1 to 12:1. This is accomplished by using frequency- and time-masking perceptual algorithms to discard signals that are said to be inaudible to human listeners. Extensions include using matrix techniques on the LS and RS channels to derive separate back surround, for instance.

Table 11-1 *Continued*

Surround Format	Two-Channel Matrix-Encoded Surround	5.1-Channel Discrete Surround
Delivery media	Theatrical film analog, DV, HDCAM and other digital tape formats,* DVD, Blu-ray, NTSC TV broadcast, most satellite cable broadcast, VHS tape	Theatrical film digital (SR-D, dts, SDDS),** DVD, Blu-ray, professional digital tape formats utilizing Dolby E compression,* used largely for digital satellite and television.
Encoding requirements	Hardware from manufacturer or software such as DVD Studio Pro or others	Hardware from manufacturer or software such as Avid's Dolby Surround Tools
Minimum monitor requirements	Decoder such as a Pro Logic II decoder in a consumer receiver, calibrated level monitor	Proper hardware decoder corresponding to encoder, bass management, calibrated monitor level

*Many professional tape formats including HDCAM have four channels, which is not enough to do a 5.1 mix directly. Thus, they are used for two sets of matrix-encoded surround mixes with differing headroom depending on their usage, or they are used with Dolby E bit-rate-reduction to fit 5.1 channels into a two-channel space.

**There is an analog backup of two-channel matrix-encoded surround on digital prints.

are listening to decoded surround sound, wins the day, and for those who are listening in stereo, there is backwards compatibility. This means that a stereo listener hears all of the sound, and most of it in the right place. Left and right alone are easy: they come out of the respective loudspeakers. Center is rendered as a phantom image. Surround is rendered as an out-of-phase signal from both the left and right loudspeakers, which tends to make it at least a little more spatial and harder to localize. For backwards compatibility all the way to mono, a problem is that the mono listener does not hear the surround component of the sound field, because it is transmitted as a left minus right signal, which results in zero output when summed into mono. So it is important not to transmit essential storytelling content by way of the surround channel in a matrixed system.

In most studio environments a hardware encoder/decoder pair is used for matrixed stereo (Fig. 11-1). These are placed in the signal path for making the Edit Master in the LtRt format. There is also a plug-in available for Pro Tools called Dolby Surround Tools, which may be rented or purchased from Avid's web site (www.avid.com). In addition, if you own the hardware, it is possible to listen to the effect of the encoder and decoder

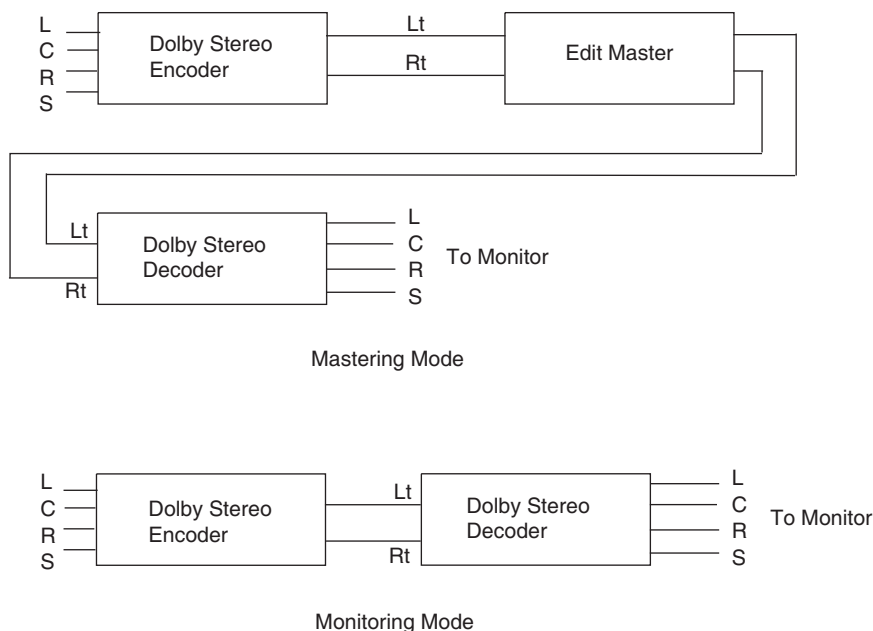


Figure 11-1 Mastering versus monitoring with a matrix encoder/decoder.

continuously while editing and mixing, by putting the encoder and decoder in the monitor path, and thus understand better the effects that the process has on the surround-sound field, and prevent errors. For instance, a common audio problem is called a *phase flip*. Using balanced lines, it is easy for this to occur because a single pair of conductors anywhere in the chain just has to be reversed from others, and you will have a phase flip. A phase flip in a matrixed surround system will cause center and surround to be reversed, so things panned center front will come out of the surrounds, and vice versa. In a multi channel environment, it is easier to hear this effect (with an encoder/decoder in the monitor path) than it is to diagnose the problem just about any other way. So the encoder/decoder has two jobs: actually encoding the master from L, C, R, S channels into LtRt and corresponding decoding at a mastering stage, as well as being placed in the monitor path to hear its effects at earlier stages.

Possible negative effects of the amplitude–phase matrix include the following:

- *Static narrowing of stereo sound fields.* If there is a strong center component to the sound field (i.e., if left and right have a lot in common), the matrix systems will tend to emphasize this and reproduce the content with a dominant center channel. This can reduce the apparent width of the sound field.
- *Dynamic narrowing of stereo sound fields.* In an early scene in *Young Sherlock Holmes*, the teenage Holmes and Watson cross a courtyard on their way to class, and there is a music cue underneath their dialogue. A poor surround decoder will modulate the width of the music with their dialogue, making the orchestra sound narrower when they talk and wider when they don't.
- *Magic surrounds.* If two-channel stereo sources have a strong difference between the two signals, expect a high level of surround output. Because this is often unexpected, the result is, well, magic.
- *Image pulling.* Strong content in one channel may pull content following it in time to its location. If a sound effect in, say, the left channel comes just before a line of dialogue, then the first syllable may be pulled left, and then the subsequent dialogue may center. This is generally most audible for misplaced dialogue, but it also occurs with sound effects and music.

- *Uneven pans.* Some systems can reproduce sound panning smoothly, and some may appear to stick for a long time at a particular location as sound is panned, and then jump all at once to the position planned.
- *Dialogue leak into surrounds.* Because dialogue is almost always anchored with a screen image, to hear it from the surrounds breaks the sound image.

The extent to which these problems are manifest depends on the particular encoder/decoder system chosen. Because what is being done cannot be strictly done scientifically (you can't really get four discrete channels from two recorded ones, as that violates information theory), there will always be subjective judgments involved in the design and use of these systems, and it takes lots of program material, not test signals, to understand what they are doing. The market is in the many tens of millions of decoders, so it is a hot area for competition. About the best a user can do to make a decision is to compare the encoder/decoder systems in a fair evaluation, with matched levels, over the same sound system, with content with which you are familiar.

The 5.1-channel system is the current leading one in providing discrete, multi channel sound. By discrete is meant that the channels are separate for all practical purposes.³ They are Left, Center, Right, Left Surround, Right Surround, and Low Frequency Effects (LFE), which is also known as the 0.1 channel. It is a low-frequency-only channel with a frequency range up to 120 Hz, and it has 10 dB greater headroom than any one of the main channels. The reasons for the channel are psychoacoustic: greater sound pressure level is needed at low frequencies to sound equally as loud as in the midrange, and only a single channel is necessary because localization of sound is increasingly poor as frequency goes down.

With multi channel sound there are two principal ingredients of surround sound to consider: imaging and envelopment. Imaging has to do with hearing the direction and size of a sound source. A more limited case of imaging is localization: pinpointing just where a sound source is located in space. The three front channels of 5.1 provide a means to locate sounds well across the front sound field. The center channel provides an "anchor"

³ Although joint low-bit-rate coding of the channels is done; thus the five channels cannot be used for five separate mono programs, for instance.

so that centered sound is always reproduced from the center of the sound field no matter where you are sitting (unlike stereo), and with sounds anchored at left, center, and right positions, frontal imaging is very solid. To a lesser extent, imaging works across the space between left and right surrounds. That's because phantom imaging more or less works across the listening area (in a cinema with an array of surround speakers) or behind a listener (in a home setup with the speakers somewhat behind the listening location) as well as in front. However, because of the lack of a center back channel, off-center listening will favor the closer side when sound is panned precisely between the two, just as in front. Panning between the front and surround channels on the same side, on the other hand, creates poor phantom imaging, because with our ears at the two sides of our heads, humans aren't built for such phantom imaging. This disposition of the channels, with more in front than behind, is thus tuned to perception and to the fact that sound is meant to have better precision of imaging where there is a corresponding picture than where there is no picture.

Uses of Surround Sound

With the foregoing perceptual discussion, it can be seen that various parts of the sound field are reproduced differently by the 5.1-channel system. This also corresponds to the aesthetics of surround-sound production. Generally speaking, sound objects with hard sound images, such as dialogue and hard effects, usually appear on the picture, which means in the LCR channels. Sound without such a hard image may appear in the front as well, or include the surrounds. Such sound includes ambience and reverberation as prime elements. One problem is that most sound effects libraries are two-channel stereo, and yet we'd like to have sound light up the surround channels. One way to do that is to use a two-channel amplitude-phase matrix decoder to produce an LCRS output from a two-channel library source, and then decorrelate the surround output into left and right channels. Decorrelation can be done by adding reverberation, by slight pitch shifting between the two channels, by using the chorus program of a signal processor, or by other custom means.

A special case that violates the norm of having only reverb and ambience in the surrounds is the transient fly-by. A famous one is a bird that flies from screen left to surround right in the jungle scene in *Apocalypse Now*. This effect alone sold the idea of two surround channels over one to

international standards bodies. Fly-bys may be enhanced with added Doppler shift and perhaps varying reverberation, along with panning.

The ultimate usage of surround sound is to engage the audience. This is often found in some of the best movie sound tracks by not sticking to just one recipe throughout, but by aesthetically determined variation throughout the piece. That is to say that there are times when a fully rendered sound field all around the listener is appropriate, and times when it is not, and variation between these is engaging.

Mastering for Level

As anyone can tell you who has tried to run a DV film festival, audio level is the single most annoying part of running things. That's because the levels are so inconsistent among the sources from various producers: one will blast while the next one will be too soft. One producer will have the monitor system they are using turned way up, and so they make the recording level low, while another will run the monitor system with the volume down, causing them to master hot. The problem is complicated by the fact that there are so many level controls in the system. Even within a computer, there may be playback level controls for the application software, for the operating system software, and even possibly for the digital-to-analog output board in software and/or hardware. To this must be added any external mixer used for monitoring, with perhaps three level controls in the signal path and those on any external-powered loudspeaker. With so many level controls—each of which affects what you hear and some of which affect what gets recorded—confusion reigns. Each developer of a product in the chain thinks it has to provide flexibility for customers, so provides such controls, yet the sum of them all in series actually greatly complicates matters for getting things right on the master. What is needed is an orderly approach to setting each of the controls so that what you record, and what you hear, are related correctly, and such an approach will be explained below under the heading "Level Calibration". This is reminiscent of the production sound problem with multiple pieces of equipment in the signal path described in Chapter 6.

At the very least, people use both -12 dBFS and -20 dBFS as reference levels as described in Chapter 3, and these are a long way apart psychoacoustically, the ratio in loudness approaching 2:1. They then place program

material at various peak levels with respect to the reference level, further confusing the issue. And even if peak levels were standardized, they don't correlate very well with loudness because peak level is measured in the very short term, by definition, whereas loudness accumulates over time. It takes some 80 ms or roughly 2.5 frames to reach more or less full loudness, yet a digital peak meter will read a 1-ms glitch at the same level as it does a full 2.5-frame sound that sounds much louder. So a program filled with 1-ms events like clicks will sound softer than one filled with longer events, even though they have the same peak level. (It is argued that the meter is set this way in order to capture the peaks that will cause audible distortion rather than indicating loudness. This is a fair way to look at it; just realize the peak meter is all about preventing distortion rather than reading loudness.) Thus the first thing that needs to be done is to standardize the settings of all the level controls in the signal path. With such standardized settings, much greater consistency can be achieved among programs.

By the way, the best contemporary answer to the DV film festival problem, before monitor calibration is widely accepted and done the same way by many producers, is for the festival operator to screen everything in advance and set a level for each program, with a chart for the operator/projectionist to use to set the volume in playback in front of an audience.

Background on –20 versus –12 dBFS Reference Level

The motivation behind the –20 dBFS reference level was for digital to more or less match the headroom of 35 mm analog film masters so that movies could be transferred to digital with little hassle, especially given that Hollywood postproduction studios were among the first to use digital video, and thus their viewpoint set the early standards. Unfortunately, a lot of conventional systems located downstream from the digital master could not handle 20 dB of headroom, which gave rise to –14 and –12 dB as reference levels. Digital masters made with a –12 dB reference and thus only a maximum of 12 dB of undistorted headroom provided a better fit into the capacity of such media as broadcast television, interstation microwave relay links, analog satellite distribution, VHS tape, and so forth. Interestingly, more recent media developments such as DVD have no problem with using –20 dBFS as a reference with 20 dB of potential headroom; only legacy systems have to be protected. Thus the two sets of reference levels are at cross purposes, and in some sense never the twain

shall meet. However, there is a way to solve this seeming conundrum by untying the Gordian knot, to be described later.

A field that exhibits much better level consistency than general DV is movies. That's because a standard is in place in postproduction for the reference level, both on the medium and acoustically in the room (Table 11-2). Thus the volume control of the dubbing stages that make movies is achieved by technical means, and that makes for better consistency.

Table 11-2 Monitoring reference levels for various playback environments.

Field of Application	Room Size	Digital Sine Wave Reference Level	Acoustical Reference Level	Monitor System Frequency Response
Cinema	Large	-20 dBFS	85 dB SPL	X curve
Cinema	Small	-20 dBFS	82 dB SPL	X curve with small room adaptation
Video with high headroom	Small	-20 dBFS	78 dB SPL	Flat to 10 kHz, then natural roll-off
Video with low headroom	Small	-12 dBFS	70 dB SPL	Flat to 10 kHz, then natural roll-off
How measured		Console VU meter set for 0 VU on -20 dBFS rms pink noise (ignore peak meters as they will read higher on the same signal!)	Sound level meter at primary listening location set to C weighting and slow reading	Real-time spectrum analyzer

Level Calibration

Calibration signals are required for the best level calibration of your monitor system in conjunction with a simple sound level meter such as a

Radio Shack 2055 (\$49.99). The necessary audio files can be downloaded from the website accompanying this book. To calibrate, do the following:

- Import the files digitally into your editing system.
- Be sure that all of the editing system's playback level controls in the chain, such as volume graphing, are set to unity or 0 dB gain, and that channel 1 is panned left and channel 2 right. In the case of 5.1-channel monitoring, such as with an external Pro Tools system, copy each region of the file to each of six channels and perform routing to six outputs. Equalization should be set flat or switched out. Since the file is stereo, you can do this in stereo pairs: left/right, center/sub, left surround/right surround.
- Note when playing the sine-wave tone that it reads exactly -20 dBFS on the editing system's software and, if present, hardware meters.⁴ Some digital meters have such coarse resolution that you cannot read -20 dBFS accurately, so for these systems just be certain that the volume graphing is reproducing the tone at unity gain.
- Set the output controls in your computer's system software to a standard position that's repeatable, such as at the 80 percent point for the slider, which is the fourth of the five fixed markers on Mac OS X, Apple Menu, System Preferences, Sound, Line Out: Built-in Audio, Output Volume. If you have control over the level of the hardware, such as with a Pro Tools system, it is also desirable to standardize the electrical voltage level by measurement with a voltmeter at each output channel, setting the output potentiometers for a nominal value such as 775 mV when playing the -20 dBFS sine wave.
- If you are using an external mixer as a control device, set its relevant channel level controls for unity gain ("U" on a Mackie, 0 dB if the scale goes above and below 0), and its master level control for unity or 0 dB. Set the mute/solo controls to be not engaged, and set the equalization flat or out on the external mixer.

⁴ If other level systems are in use than referencing directly to full scale, you may have to make an adjustment for where -20 dBFS lies. For instance, on the individual channel track meters in Final Cut Pro, reference level is -12 dBFS, and db numbers are scaled from there. Thus -20 dBFS is -8 dB on this scale.

- Play the sine-wave tone into the external mixer. Set the input trim controls so that the -20 dBFS tone reads 0 VU on VU meters (usually those that have 0 somewhere near the top, but with numbers above 0 VU) or -20 dBFS on digital peak meters (those with 0 dB right at the top, and a descending scale of numbers below it). Note that some digital peak meters neglect to use the minus sign ($-$) for dB below 0 because of space limitations. If you are summing channels to get from 5.1 to 2.0 for monitoring, mute the channels unused and proceed to check level in pairs.
- Roughly set the monitor volume control level on the external mixer and any input level controls on the loudspeakers for a comfortable level playing the tone (for now). These should be at nominal, repeatable settings.
- Play the band-limited pink noise. This sounds like a kind of filtered waterfall. Ignore all of the meters. Peak meters will read something around 10 dB greater than they did on the sine-wave tone, but ignore this. The higher reading is caused by the time properties of the noise and does not reflect its long-term value. A VU meter will read it much more closely to its long-term value, but because it bounces, the exact level will be difficult to tell. Trust that the file has been made correctly and that the level is in fact precisely -20 dBFS measured on an rms basis, which is how engineers measure the long-term value.
- Use the mute and, if necessary, the pan function of the editing system on all unused channels to play the noise from each channel in turn.⁵ Place the sound level meter in the center of where your head would be while listening and stand off to one side while holding it, not behind it. Set the meter to C weighting and slow reading. Adjust the level control of each powered speaker or power amp level control in turn so that the noise reads the value given in Table 11-3.
- If you are using a power amp and set of loudspeakers with no gain control, and your console only has a common, ganged monitor volume control, you'll have to check that each of the channels produces the same sound pressure level for one level control setting.

⁵ Note that while the solo function might work, especially on external consoles, solo may change the level or position; thus using mutes is more certain.

Table 11-3 Reference sound pressure levels for monitoring.

For	Reading
Monitoring for video release in edit rooms, mix rooms, and control rooms, with all channels set each in turn	79 dB SPL (–1 dB on the 80 dB scale)
Monitoring for theatrical distribution masters in edit rooms and small mix rooms, each channel in turn	
L, C if used, R	81 dB SPL (+1 dB on the 80 dB scale)
LS, RS	78 dB SPL (–2 dB on the 80 dB scale)
Monitoring for theatrical originals being redistributed into the home (DVD) in edit rooms and small mix rooms, each channel in turn	
L, C if used, R, LS, RS	81 dB SPL (+1 dB on the 80 dB scale)

The alignment should be within 0.5 dB. If you can't achieve this degree of balance, a trim should be made to each channel's level to reach the correct SPL.

In film dubbing operations, these settings are rarely changed. While editing, you may find a need to increase the level to hear a background, and possibly while mixing occasionally to listen to the backgrounds. In that case, doing so on a calibrated control, such as the main channel level faders, allows you to reset it to the calibrated level once your increased gain listening is complete.

When You Can't Calibrate with Test Signals

Although the aforementioned procedure looks rather long, it can actually be performed quickly once you've done it a time or two. But what if you've lost track of the file and don't have a sound level meter? The first thing to know is that, typically, dialogue controls the volume level that people set. For programs like action-adventure movies, extra headroom is needed beyond the level of dialogue for momentary loud sounds, so dialogue cannot be recorded up to full scale. It is typical for normal dialogue to

peak around -10 dBFS, again, on a digital peak meter, which is not very reliable for reading loudness. Because of the relative sluggishness of VU meters, the dialogue will normally peak somewhat below 0 VU, reaching 0 on the maximum dialogue level. This recorded level allows then for 10 dB peaks above the running level of the dialogue for those big scenes.

Set the system up in a similar way to that described under calibrating: every control should be run around its nominal level. If one control in the chain is turned down, and a subsequent one turned up to make up the gain, the result is excessive noise that is increased from the stages between the two controls. If one control is turned up relative to normal, and a subsequent one turned down to get the gain right, the potential is for clipping distortion to occur between the two controls, just as bad as digital clipping. What you do in setting all of the controls around their nominal values and then trimming to get the overall system level correct is called *gain staging*.

Then, having set the controls nominally and with dialogue that's recorded well as a source, set the monitor volume control to the right level. The surprising thing about this is how close even untrained people come if they just let themselves go, don't look at the meters, and set the level. I ask beginning mixing students to take a blind vote on the level of dialogue, and the agreement among them is remarkably good, certainly in the range of ± 2 dB—this seems to be one experience that's universal.

Best One-Size-Fits-All Approach

Once the monitor system has been calibrated, we can get on with recording. Interestingly perhaps, this is the point at which I cover up all of the meters and make mixers respond to what they hear rather than what they see. After all, mixing is experiential in the end. However, we still have to peek at the meters, because there are still delivery requirements to deal with, set by the company to which you are to deliver.

For a program that is completely dialogue driven, the temptation is to master the program so that the highest peak in the entire program just reaches 0 dBFS; you might then say that the program has been normalized, in the plug-in sense of the term. However, this leads to an absurdity

when your product is then intermixed in an overall film festival or broadcast day that contains an action-adventure film mastered with dialogue peaks hitting -15 typically—there's a 15 dB difference! The answer is: *do not* do this! Do not master the dialogue-driven program at peaks of 0 dBFS. This is a bad habit that the music industry is in, and it shouldn't be copied into the audio-for-picture industry because it only leads to badly non-interchangeable programs. You might actually get rejected by a network as well, because some of their distribution paths only have around 10 dB headroom, and their reference is -20 dBFS (as you can see on virtually any digital television tape machine). Instead, peak at a maximum of -10 dBFS with a reference level of -20 dBFS and you will have made one universal product. Losing about a bit and a half of dynamic range is not too much of a price to pay even in a 16-bit environment, and the world is evolving to 24-bit mixes, where it just doesn't matter that you are under-recording by 10 dB.

Considering this argument in terms of the two widely used reference levels produces the following prescription: The reason the -12 dBFS reference level exists is that there are downstream media and equipment from the master that can handle only 12 dB of headroom. Yet intermixing media referenced at -12 and -20 dBFS leads to big problems. There is a solution: reference to -20 dBFS and use only about 8 to 10 dB peaks above that level, thus keeping to the headroom requirements of those paths having low headroom, but producing better interchangeability with sources referenced to -20 dBFS.

This goes against the grain of those who have been trained that to get the best dynamic range in digital audio you ought to put the highest peaks at 0 dBFS, because it leaves unused the top 10 dB of the dynamic range. As with all folklore, there is something to be said for mastering to the highest levels: it's supposed to keep you well above the noise floor. And in the early days of digital audio, the low-level region of converters was quite bad, turning low-level smooth sine waves into coarse square waves. These problems are long gone with good equipment today, so the prescription passed down for years does not apply. Furthermore, we have seen that the actual limitation on dynamic range is not typically the basic medium, but is the actual implementation of the camera electronics, source recordings, and so on, so the idea of recording hot to prevent low-level problems in converters isn't even a problem in practice—the program controls the noise, not the converters.

The bottom line on mastering for level is the same as the mixing prescription: start with the dialogue and then scale the backgrounds, Foley, effects, and music to it. In music mixing, the equivalent is to add a track at a time. Because of the masking effects of one track over another, music mixers will compress and equalize each track in turn so that the mix elements read through one another, which is one reason music mixing takes a long time. Mixing for accompanying a picture, with dialogue, music, and effects, has two things going for it that music mixing doesn't have. First, there's the relatively strong difference among the types of sources; the elements are less likely to cover one another up when presented together, and good editing involves leaving "little holes" for the dialogue to shine through that music does not do as often—although music composition can leave room for a solo to shine too, for example. Second, there's the fact that much mixing for picture is in surround sound. Spreading sources out around a surround-sound field allows the individual tracks more clarity than having to squish them into two-channel stereo.

You might need to apply some peak limiting to stick to the prescription for a universal master of peaks at -10 dBFS. Let's say you've first prepared a wide dynamic range LtRt mix for DVD release, with 20 dB of headroom and -20 dBFS reference. Now you have to make a new delivery master for, say, a demanding network. Applying SMPTE reference of -20 dBFS, you should peak at -10 dBFS. So one way to do this would be to set a limiter with an infinite compression ratio to -10 dBFS. However, substantial limiting may be audible as unnatural ducks in the content. So the best way to proceed is probably to turn the level of the original mix down by about 5 dB, and add 5 dB of peak-limiting gain reduction, thus limiting the 20 dB headroom master to 10 dB of headroom.

Mastering for DVD, Blu-ray, Digital Broadcast, and Digital Satellite Television

When digital video files or tape masters are transferred to DVD or Blu-ray disc, another mastering stage occurs in order to translate the linear PCM conventional digital audio used in postproduction to the bit-rate-reduced form necessary to fit on the DVD or Blu-ray. In addition to simply compressing the bit rate, Dolby Digital, Dolby TrueHD, and DTS-HD have features called *metadata*, or data about the digital audio. Metadata in encoded audio bitstreams for distribution contains, among other things,

Table 11-4 Digital audio bitstreams commonly used on DVD and Blu-ray disc.

Bitstream Format	Channels (Typical)	Bit Rate	Disc Format	Encoding Software
Dolby Digital (AC-3)	2.0, 5.1	up to 448 kbps (5.1-channel DVD) or 640 kbps (Blu-ray)	DVD Blu-ray	A.Pack Neyrinck SoundCode Minnetonka Audio SurCode most DVD authoring software
DTS Digital Surround ⁶	5.1	754 kbps to 1.5 Mbps	DVD Blu-ray	Minnetonka Audio SurCode DTS Encoder
Dolby Lossless (TrueHD)	5.1	up to 18 Mbps	Blu-ray	Dolby Media Encoder
DTS-HD Master Audio	5.1	up to 24.5 Mbps	Blu-ray	Minnetonka Audio SurCode* DTS-HD Encoder
Linear PCM	2.0, 5.1**	1,536 kbps (2-channel, 48 kHz, 16-bit)	DVD Blu-ray	any (Pro Tools, Logic, Audacity)

* Compressor plug-in enabling creation of DTS-HD Master Audio streams for Blu-ray content.

** Because of its high bit rate, LPCM would typically be used for 2-channel audio on DVDs when delivery of an LtRt or 2-channel stereo master audio recording is desired.

two different level-affecting mechanisms: dialogue normalization and dynamic range control. So, in mastering for optical disc, a measure of standardization can be added to that present on delivery masters.

Table 11-4 lists the standard audio formats in use on DVD and Blu-ray video media, along with information on how to go about encoding audio to each format using video or audio editing software plug-ins. The program used by DVD Studio Pro to encode Dolby Digital is called A.Pack. Plug-ins for digital audio workstations, such as Minnetonka's SurCode plug-ins and Neyrinck's SoundCode plug-ins for Dolby and DTS formats,

⁶ Home release format, not to be confused with the DTS theatrical sound format.

also allow PCM audio to be encoded for inclusion on DVD and Blu-ray release media, and they operate right out of the DAW to produce a single bitstream file ready to be imported into authoring software. Apple's Compressor can also be used in conjunction with external plugins, such as Minnetonka's SurCode plugins for Dolby and DTS audio formats, to encode audio for DVD and Blu-ray release. Some media such as Blu-ray, DVD, and digital broadcast platforms have space for more than one audio service, with Complete Main being a dialogue, music, and effects main mix but also capable of handling special content for hearing impaired (compressed dialogue only), visually impaired (description of the picture), and so forth, as separate, parallel bit streams. Alternate audio tracks such as filmmakers' commentaries on Blu-rays and DVDs also make use of the available space for multiple audio programs on a single disc. Each of these different services needs a separate coding pass to produce its own files, which are then multiplexed in authoring to produce the final bitstream to be written to disc.

For Dolby Digital 5.1 encoding using A.Pack, the parameters to be set for each service before encoding can occur are as follows:

Under the Audio Tab of A.Pack:

- *Target system.* Whether DVD Video, DVD Audio, or generic AC-3, depending on where the master is to be used.
- *Audio coding mode.* This specifies the number of channels and their distribution. The possibilities are 1/0, 2/0, 3/0, 2/1, 3/1, 2/2, 3/2. The two numbers are the number of front and the number of surround channels, respectively: 3/2 is the five-channel system, because it stands for three front and two surround channels. Note that those with one transmitted surround channel would have that channel reproduced by all of the surround speakers in a system: it's mono surround, not single-speaker surround. Also, of these, 1/0, 2/0, and 3/2 are the popular choices.

Beneath the Audio Coding Mode pulldown menu is a box labeled Enable Low Frequency Effects that should be checked if an LFE channel is available. Note that most, if not all, set-top cable/satellite boxes ignore the LFE channel when mixing down 5.1 content to their two-channel outputs, so no essential storytelling information can be put only in the LFE channel. It is meant to be an enhancement

channel, not an essential one, in home distribution. For this reason alone, there may need to be different delivery masters for cinema and home, because in the cinema the use of a dedicated subwoofer system is assumed, and so its content will be heard. For corresponding delivery masters aimed at the home, some studios choose to redistribute the LFE content from cinema masters into the main channels and LFE channel to make masters wherein the LFE content will survive mixdown to two channels, but other studios do not. This is one reason why separate cinema and home masters may be needed—other reasons are described later.

- *Data rate.* A variety of bit rates are selectable for this specified service, and the range of bit rates available depends on the number of audio channels to be coded. Higher rates produce a greater likelihood that no coding artifacts will be heard. The maximum coding rate is typically 448 kbps, but note that multiple streams are possible simultaneously on DVD for separate languages and that each one uses up some of the bit budget, and the bit rate must be traded off against program length, number of audio services, and picture quality. Typical would be two audio services in one language, an Lt/Rt and a 5.1, perhaps coded at 192 kbps and 448 kbps, respectively, for a total of 640 kbps, a fraction of the average bit rate of the DVD of typically 3.5 Mbps and only about 6 percent of the peak rate of 10 Mbps.
- *Dialogue normalization.* The default value for movies is -27 dBFS, and if you have followed the calibration in this chapter and recorded dialogue at normal levels, this is the setting you should use.
- *Bit stream mode.* A variety of audio services may accompany one video program on broadcast and satellite digital television, and on DVD. Complete Main is the most common setting, but ancillary services such as those for the hearing impaired or visually impaired are also possible audio services. They would be coded in a separate pass of the encoder and combined only on the final medium.

Under the Bitstream Tab of A.Pack:

- *Center downmix.* This applies to the backwards compatibility of three front versus two front speakers. In the case where the end user has only two speakers, how much of the center signal should be delivered

to each of the channels? The goal is to get a sound source that is panned across the front channels to remain at constant loudness throughout the pan. Classical acoustics says that if you are a healthy distance from sound sources in a room, two equal-level sources will add by 3 dB, so the correct downmix level for this condition is -3 dB, which is the typical setting. However, you might expect the user to sit close and be dominated by the direct sound field, in which case a better downmix may be -6 dB. The value -4.5 dB was meant to be correct in all cases, ± 1.5 dB.

- *Surround downmix.* This answers the question “how much left surround should be mixed into left front if there is no surround loudspeaker, and the same for the right channels?” The answer here depends on how prominent the surrounds are. If it is felt that a mixdown would be too great at the normal -3 dB, then -6 dB and off values are possible. These might be necessary for a sports game with loud crowd noises, for instance, where the two-channel and even mono viewer could wind up with too much crowd noise.
- *Dolby surround mode.* This is used to distinguish between LoRo (two-channel stereo) mixes and LtRt ones intended for amplitude-phase matrix reproduction (by way of Dolby Pro Logic II, Neo:6, Circle Surround, Logic 7, etc.). A third value is Not Indicated, but readers of this book should know enough by now to pick one of the other two indicated alternatives.
- *Copyright exists* checkbox. Check with program producer.
- *Content is original* checkbox. Check with program producer.
- *Audio production information* checkbox. Check the box to indicate that the following information has been provided. Again, readers of this book are in a position to understand and use these features.
- *Peak mixing level.* Dialogue normalization is relative. To get the right absolute SPL calibration, future controllers would be able to reproduce the sound at exactly the same level as the producer or at a fixed offset from the producer’s level (such as always 6 dB quieter than “Hollywood”). The number indicated is the sum of two factors: (1) the headroom, which is the amount between the reference level and 0 dBFS, such as 20 dB, and (2) the sound pressure level that you

have set the front channels to. So, if you are using 79 dB SPL and -20 dBFS as a reference, set the number to 99 dB.

- *Room type.* The choices are “Small Room, Flat Monitor,” which is what it is likely you are using, “Large Room, X Curve Monitor,” which is used on film dubbing stages, or “Not Indicated.” Setting this parameter to “Small Room, Flat Monitor” will likely best duplicate your conditions of monitoring. This is more fully explained later.
- *Compression preset.* Many people observe that films mixed in large dubbing rooms for listening to in cinemas sound as though they have too much dynamic range when played at home. They say the dialogue is buried, but in fact this usually means that the entire mix has been turned down compared to the original, and the dialogue is the most affected. To reduce this problem, the listener can add audio compression optionally, and this parameter sets what program type and thus how the parameters of the compressor are set. Note that it is up to the end user’s controller whether this is ignored, as in a home theater application, used all the time as in a set-top box, or switchable in and out, sometimes called a “night switch,” in a receiver. The settings are program dependent: Film Standard Compression, Film Light Compression, Music Standard Compression, Music Light Compression, Speech Compression.
- *General.* RF Overmodulation Protection is for the case where the output of a receiver must be retransmitted as a television channel on an RF (antenna) cable within a home. Many VCRs had this feature, so modern receivers sometimes have it too: you can switch the VCR or set-top box typically to put out a standard NTSC television channel 3 or 4. In this case, the wide dynamic range of modern-day programs can overload the simple mono modulator, and therefore the RF Overmodulation Protection should be checked.
- *Apply digital deemphasis* is used only for cases where the PCM (whatever form, WAV, etc.) has been preemphasized, such as in a very few CDs. It is practically never used.
- *Full-bandwidth channels.* The Low-Pass Filter box should be checked in the case of operating at any lower than the maximum bit rate for the number of audio channels. It prevents (probably rare) problems with

the highest frequency audio signals. The DC filter may be useful in the case where there is no other such filter in line, preventing wasted coding of inaudible signals and eliminating potential thumps upon changing programs, for instance.

- *LFE channel.* Apply Low-Pass Filter restricts the bandwidth of the LFE channel appropriately and would normally be used. This is because editing and mixing software often lacks this function, and the LFE bandwidth needs to be limited, because if it isn't upon recording/transmission, it certainly will be in playback/reception. This can create major surprises as the attack of a kick drum, for instance heard over a monitoring system with relatively unrestricted subwoofer bandwidth, becomes band limited in transmission. The hard part—the transient—of the attack is gone, and this reveals a mixing problem (the content should have been split between a main channel and the LFE one). Or, better yet, the entire bass drum should be in the main mix where the time relationships across the various frequencies will be better preserved. Save LFE for when it is really needed, like the cannon fire in the “1812 Overture.”
- *Surround channels.* Apply 90° Phase Shift would normally be checked so that matrix surround decoders work correctly. Apply 3 dB Attenuation is used only in the case of theatrical cinema masters being transferred to home video, because the calibration levels are different between these two domains by the same 3 dB.

Broadcast terrestrial and satellite digital television also uses metadata, and one of the metadata elements is absolutely standardized as required of all digital television receivers: Dialogue Normalization. For these media, the mastering may be performed by a professional postproduction house, which today will supply the metadata typically in written form, so that the parameters may be set upon ingesting the source program into the broadcast channel. The label that follows is to be used to label the source media, such as a DTRS tape. The resulting files, and recordings by way of Dolby E, should additionally be labeled with metadata settings as described above.

Multichannel Audio Postproduction Media Label***Postproduction Studio Info***

Studio Name _____

Studio Address _____

Studio Phone Number _____

Contact Person _____

Date Prepared (e.g., 1999-01-12) _____

Program Info

Producing Organization _____

The following items are used by the ATSC Program/Episode/Version identifier.

Program _____

Episode Name _____

Episode # (1-4095) _____

Version _____

Version # (1-4095) _____

First Air or Street Date (e.g., 1999-01-12) _____

Program Length (time) _____

Contents Info

Track layout (check one):

	1	2	3	4	5	6	7	8
☐	L	R	C	LFE	LS	RS	Lt	Rt
☐ Other								

Leader contents ☐ 1 kHz sine-wave tone at -20 dBFS, 30 s (required) from heads:
☐ Pink noise at -20 dBFS rms, 30 s (optional)
☐ Silence, 30 s (required)
☐ 2 Pop on countdown leader (optional)
☐ Other: _____

Program starts at ☐ 01:00:00:00 ☐ Other _____

Program ends at _____

☐ Multiple program segments (detail):

Time code ☐ 29.97 DF ☐ 29.97 NDF ☐ 30.00 DF ☐ 30.00 NDF

Sample rate ☐ 48.000 kHz ☐ Other (why?): _____

Tape Geneology ☐ Original ☐ Clone ☐ Dub ☐ Simultaneous Protection

©2005 TMH Corporation. This form may be duplicated with inclusion of this notice.

Monitoring

Perhaps the most common cry heard from editors in mixes is “it sounded all right on the Avid.” This means that in the editing room, with monitor speakers that don’t cover the full frequency range and are lumpy in frequency response across the range, with masking by lots of computer noises, and in a hard, small room with awful acoustics, the sound was acceptable. When subsequently heard in a quiet dubbing stage or control room, with controlled reverberation, and over a system that correctly spatializes the sound for the given recording, it sounds awful. This cry is often given in frustration: “Why do sound people have to be so picky?” but the problem of perception nonetheless exists.

The story is told about why “it sounded all right on the Avid,” and when engineers investigated it was found that the tweeters, the high-frequency elements of the monitor loudspeakers, were burned out. So the editor missed all of those subtle background changes that make for bumpy edits, because the sound system wasn’t putting them out.

It is perhaps inevitable that edit room monitor systems will always be less expensive than sound control rooms, but that doesn’t mean that they have to be put behind the picture monitor and facing away, bouncing off the wall behind. Most edit rooms will be restricted to two-channel monitoring, and this is usually adequate for cutting because most elements are delivered as mono or two-channel. So at least the two channels can be set up to span the left and right sides of the main picture monitor and, with centered listening, produce a phantom image on the screen for dialogue.

The next most important thing is the frequency range of the monitor. An old saw in audio is that the product of the low-frequency and high-frequency limits of a sound system should equal 400,000. Thus 20 times 20,000 Hz equals 400,000, and so does 100 times 4000 Hz. As the low-frequency limit goes up, which is what costs the most in size and money, the high-frequency bandwidth also should come down, but it often doesn’t, leading to trebly sounding audio from many small monitor loudspeakers. Most edit room-style monitors go down to the 60 to 100 Hz range and miss one to two octaves of importance at the low end. Then, if the program material edited on such a system is played on one with full range, bumps at the edits at low frequencies may be heard.

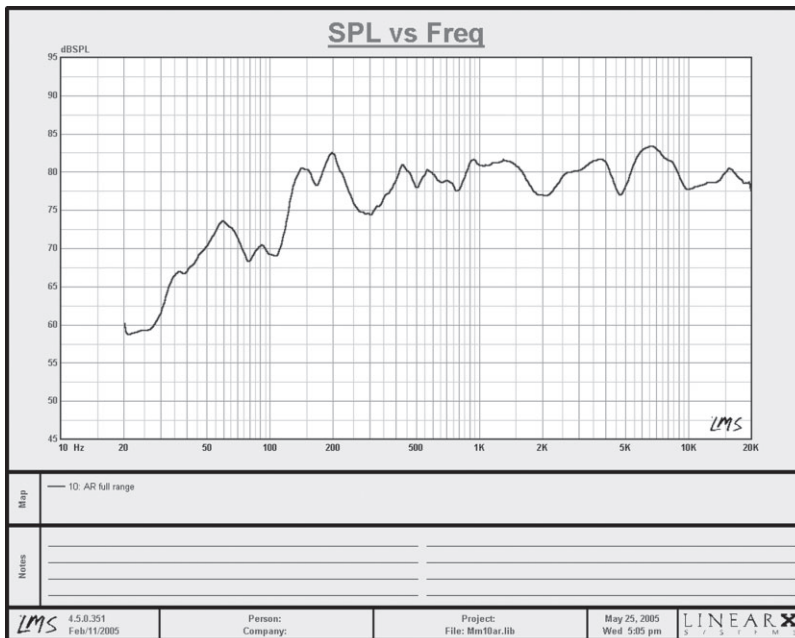
The way around this is to use a common-bass subwoofer for the two channels. This works for several reasons. It allows the smaller monitors to be placed relative to the picture and the subwoofer placed on the floor up against a wall or corner. With a low-enough crossover frequency such as 80 Hz, with the bass below that redirected to the sub, the system works because we generally don't hear the direction of bass as the frequency goes down. Sure we localize the higher harmonics of the bass drum hit, but that's not bass! This is no huge investment today and can save hours of frustration in mixing.

The way to use a common-bass subwoofer requires electronics called *bass management* (Fig. 11-2). This system does the filtering and redirection necessary. It is routine in home receivers, and some professional systems have the capability, either as an outboard unit or built into the subwoofer. A bass manager that works for these purposes is the Martinsound ManagerMax. A bass manager that works for these purposes is the Martinsound ManagerMax.

The next problem is the frequency response of the monitor-room combination. Small rooms are awful in the midbass, because of strong standing waves. Put out some program material with strong bass content in your room. Now move up against the back wall. Almost 99 percent of the time the bass goes up dramatically. Move into a corner: it goes up even more. This is no reflection on the monitor or its manufacturer; it is the interaction between the loudspeaker and the room. For most people, this problem is currently insoluble, and all you can do is buy a decent loudspeaker that will be smooth at the higher frequencies and hope that the lower ones



Figure 11-2 A bass manager. It high-pass filters five main channels, then sums and low-pass filters the five channels, adding the LFE signal to the sum and directing it to one or more subwoofers.



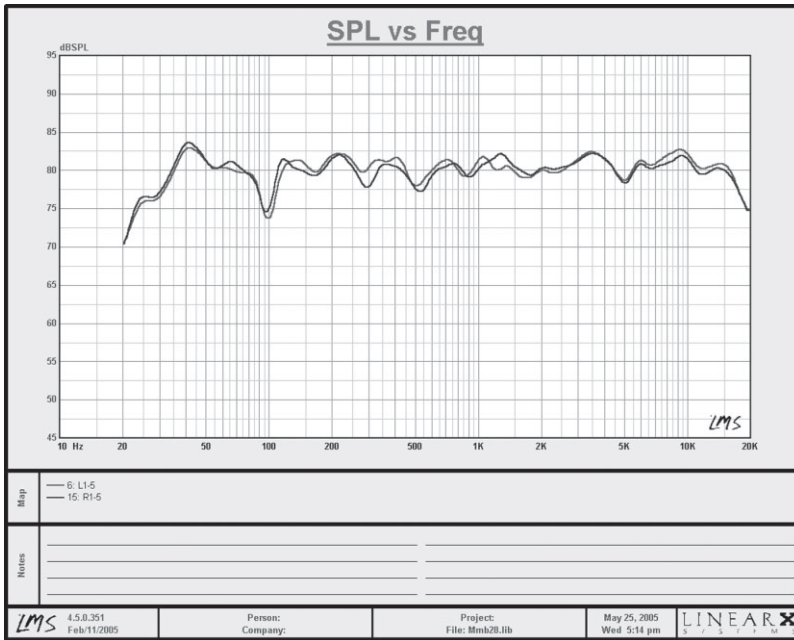
(a)

Figure 11-3a The frequency response of a typical edit room monitor shows that it lacks bass and has uneven frequency response across the band. Measured at a typical listening location in an edit room.

will be all right. Today, home receivers are available that include auto-equalization that goes a very long way toward presenting a frequency response more closely matching what is heard on a professional dub stage (Fig. 11-3). For two-channel work, many models may be used. For 5.1 discrete channel work, most receivers do not have discrete analog inputs for multi-channel audio. One does: the NAD T787. This unit would allow 5.1 inputs and speaker outputs with equalization: either 5.1 discrete speakers, or a mixdown to two.

For mixing 5.1-channel sound on the desktop, not just editing and mixing two-channel sound, a correctly spatialized monitor system is necessary. Necessities are the following:

- A minimum of two front loudspeakers (if center is to be rendered as a phantom) or three including a hard center speaker.



(b)

Figure 11-3b The frequency response of an equalized system, showing a wide frequency range and smooth response over the range.

- A minimum of two surround loudspeakers.
- A minimum of one subwoofer.
- A bass manager circuit.
- Setup such that left, center if present, and right relate correctly to the picture, so that the sound and associated picture appear in the same place. An error in placement in the horizontal plane of ± 4 degrees is noticeable by professionals, and ± 15 degrees is annoying. The speakers are optimally at the same height as the picture, but may be lower or higher because we hear vertical direction only about one-third as well as horizontal direction.
- Setup with left and right surround loudspeakers at ± 110 degrees from front center when viewed from above. Surround speakers are usually more elevated than front speakers, mimicking theater arrangements.

- Subwoofer placed for best response. This is often the corner, which produces the most output. If two subs are available, operate them in mono from the same signal and place one in the corner and one half-way along one wall, both at the floor.
- Apply bass management.
- Apply calibration with an auto-equalized system, and additionally calibrate it for absolute level as described earlier in this chapter.

Film versus Video Mixes

The differing reference levels between cinema and home have been described, as well as the fact that a consumer 5.1 mix needs LFE redistributed to some degree from LFE only into LCR in order that the content will not be lost in mixdown. There are two final points to make: (1) theatrical mixes are done to a different frequency response curve than home mixes; and (2) the sync is different for large rooms, by one frame.

The differences between frequency response environments means theatrical mixes played at home will sound bright, whereas ones mixed on a flat desktop monitor reproduced in a cinema will sound dull. So the final mastering stage of a mix made on flat monitors for distribution into a theatrical environment is to add high-frequency shelving equalizers set to +4 dB at 10 kHz.

The consequences of film mixes sounding bright at home are overemphasis on sibilance in speech, too much apparent Foley, and sizzly sounding strings in music. These can be accommodated on receivers and controllers equipped with Home THX by switching in re-equalization, which fixes this problem. But for the problem of the video mix shown theatrically, no fix is available in the cinema, so you have to build this into the master. While listening on your nice flat near-field sound system, film mixes will sound too bright, but they will become right in the correctly tuned theater environment (SMPTE 202 and ISO 2969 standards).

Additionally, if your master is to be shown in large theaters (as opposed to small screening rooms), pull up the sound earlier relative to picture by

one frame for the theatrical master, which will put it in hard sync at 47 feet from the screen.

So we've come fully to the end: you have a mix that sounds correct on your proper monitor, and now it translates into homes and cinemas. But what if you can't afford that proper monitor? The best way to proceed is to play mixes from good mixers, widely available on DVD, through your system and note what, in particular, voices sound like. The voice has the greatest recognition for most of us, and if that sounds right, a whole host of other problems can be forgiven.

Among the best times of Tom Holman's life are sitting in a movie theater watching the screen title crawl at the end of a wonderful film and finding a former student's name associated with a great sound track. This book attempts to bring the techniques that he teaches to a wider audience and to make sound tracks accompanying moving pictures sound better. We wish that someday that sound track will be yours.

Director's Cut

- Distinguish tapes or files by calling them Camera Original, Edit Master, and Delivery (or Studio) Master. Anything else is an intermediate product. The word *master* standing alone conveys almost no information because everything we use in production or postproduction except for final released materials is a master of one sort or another.
- Mono is used for cable television and low-budget video. While conventional two-channel stereo is occasionally used for legacy materials, surround sound rules today.
- LtRt matrixed surround sound delivered over two channels fits into multiple outlet chains, such as conventional broadcast, satellite and cable, standard and high definition video tape (from VHS up through HDCAM and newer formats), as well as those channels that also have a 5.1 capacity: theatrical, DVD, and some HD satellite and cable services. Thus LtRts are the most universal method of delivering surround sound.
- Audio with 5.1 channels offers greater discreteness or specificity (it can handle more signals simultaneously, presenting their direction correctly) and greater surround envelopment (two channels versus LtRt's single mono surround channel), for reproduction through theatrical, DVD, and some HD satellite and cable services.

Director's Cut *(Continued)*

- Mastering for level: Table 11-3 and the files on the web site can be used to set the level of dialogue appropriately. The SMPTE reference level for sound accompanying pictures is -20 dBFS, and it should be used. Higher reference levels found in music recording lead to less interchangeability of product. You can test your product against others by playing good DVDs or Blu-ray discs and matching them for loudness.
- Many distribution channels have 20 dB of headroom (DVD, theatrical, some HD satellite and cable services). For others, do not use all 20 dB of the headroom, because the downstream equipment and media cannot handle that much. To make a universal mix that fits all media, turn down the mix by some 5 dB and add 5 dB of peak limiting, so that the maximum digital recorded level is -10 dBFS, and the resulting product will fit all other reasonable media.
- Monitoring causes lots of trouble. "It sounded all right on the Avid" is frequently heard as a bromide, when what is meant is "why does it sound so bad on your professional monitoring system, Mr. Sound Mixer, when it sounded all right to me?" (under the lousy conditions of monitoring in the editing room). Calibrated monitoring is a small expense compared to other things in the studio, so it should be given due consideration. It is worthwhile to listen to another product that you know well over the monitor system to hear how it translates.
- Bass management is necessary in virtually all monitoring systems, because end users will listen to the program that way.
- There is a reason to tailor mixes specifically for the theatrical or home environments, and methods of doing so are given in the text.

Index

- 2-channel system *see* two-channel
- audio
 - 2:3 field sequences 29, 56
 - 4/-10 box 82
 - 4-channel system 71
 - 5.1-channel system 261, 296, 298–9, 303–4, 308, 314
 - 12-bit recording 19–21
 - 16-bit
 - audio 19, 21–3, 53, 184, 189, 191, 259
 - recording 19–20, 141
 - word length 20–1, 116
 - 20-bit word length 20
 - 24-bit audio 17, 20–1, 23, 53, 312
 - 44.1-kHz standard 19, 182–3, 185, 188
 - 48-kHz sample rate 21, 116, 226
 - 600-ohm loads 78, 80
 - 1080-line video 38, 40
 - 1280 × 720 video 38
 - AAF format 187, 189, 231
 - absolute level 272, 326
 - abstraction 202, 218–20
 - AC-3 10, 22, 41, 172, 299, 314–15
 - accessories 117, 120, 148
 - microphones 99–105, 122
 - sound kit 117–18
 - acoustical energy 119, 139
 - acoustical levels 73, 147 *see also* sound
 - pressure level (SPL)
 - acoustical noise 21, 145
 - action-adventure movies 206, 215, 272, 310
- Adams, Ansel 142
- ADR (automated dialogue replacement) 112–13, 208–9, 211, 234–6, 242, 248, 265
 - tracks 238, 241, 257, 292
- AES format 83, 181, 185, 187–9
- AES3 174, 176, 178,
 - 180–2, 185
 - interfaces 173, 185
- AES31 format 187–9
- AGCs (automatic gain controls) 6–7, 156, 275
- AIFF (Apple Interchange File Format) 186–8
- AIFF-C (Apple Interchange Film Format for Compression) 186, 189
- All That Jazz* 216
- Almost Famous* 146
- ambience 92–6, 111–13, 203, 209, 211–12, 216–18, 233, 245–6
 - premix 270, 293
 - tracks 203, 208, 243, 245
- ambient sounds 86, 195, 202
- American Graffiti* 220
- American Pie* 2 213
- amplitude
 - of signal 18, 26
 - of sound 11, 15, 229, 317
- amplitude-phase matrix 298–9, 302, 304, 317
- analog 2, 17, 77–8, 173–4, 179–80, 225, 289

- analog audio 73, 83, 141, 174, 179
- analog connections 84, 173
- analog connectors 179
- analog copies 2, 296
- analog film masters 75, 306
- analog inputs 324
- analog line levels 75, 173
- analog media 75
- analog outputs 179, 183
- analog phenomena 25–6
- analog recorders 207
- analog recordings 10
- analog satellite systems 76, 306
- analog signals 4, 78, 176
- analog tape 29
 - machines 256, 261, 273
- analog-to-digital converters 9, 68, 173, 183
- analog video 2
 - cameras 27
 - connectors 174
 - formats 2
- anamorphic lenses 36–8
- anamorphic video modes 56
- anamorphic widescreen 36–7, 39
- angles 60, 97, 119, 126, 128, 132, 243
- animal sounds 223, 243
- antennas 110, 136, 318
- A.Pack 314–16
- Apocalypse 125, 222, 224, 285, 304
- Apocalypse Now* 125, 222, 224, 244, 285, 304
- archival copies 167–8
- archival format 43, 168
- artifacts 10, 29, 34, 42, 54, 136, 256
 - audible 10, 42
- aspect ratio 56–7
 - and resolution 36–40
 - widescreen 36–9, 56, 221
- astronauts 245
- ATSC (Advanced Television Systems Committee) 39, 320
- AU (Audio Units) 260, 286
- audible artifacts 10, 42
- audible distortion 9–10, 75, 139, 306
- audible frequency range 11
- audible problems 109–11
- audiences 2, 62, 202–3, 212–13, 219–20, 290, 305–6
- audio *see* *Introductory Note*
- audio channels 16–17, 19, 21, 65, 70, 176, 190
- audio clips *see* sound clips
- audio coding mode 315
- audio compression 41–2, 275, 318
- audio connectors 83–4
- audio data 9–10, 194
- audio editing 84, 204, 225
 - digital 225, 229
- audio equipment 75, 174
- audio files 169, 189, 191–2, 227, 230–1, 257, 265
 - formats 188–9, 200
- audio guidetrack 193–4, 196
- audio inputs 6–7, 53, 64, 166
- audio interfaces 18, 70, 174, 204
- audio levels 61, 141, 180, 305
- audio meters 77, 84, 140–1
- audio mixers *see* mixers
- audio output *see* outputs
- audio plug-ins *see* plug-ins
- audio processes 7, 267, 270–88
 - related to frequency 278–83
 - related to level 272–8
- audio programs 315
- audio recorders 6, 18, 21–2, 40, 50, 69–71, 193
 - dedicated 5–6, 59, 66–7, 94, 189, 192
- audio recording 19, 42, 68, 193
 - 2-channel stereo master 314
 - double-system 192, 194
 - formats 6, 9–10, 172
 - four-channel 17
- audio recordings
 - lossy 42
 - master production 194
 - single-system 68

- audio sample rate *see* sample rate
- audio samples *see* samples
- audio services 315–16
- audio signals 9, 81, 121–2, 178, 319
- audio syncing *see* syncing
- audio transfers, streaming 174
- Audio Units (AU) 260, 286
- audio-video synchronization 52, 115
- audio waveforms *see* waveforms
- AudioSuite 260, 286, 291–2
- automated dialogue replacement
 see ADR
- automatic gain controls (AGCs) 6–7,
 156, 275
- automation 233, 258, 262, 270, 291
- autosync
 - PluralEyes and DualEyes 194–5
 - timecode 193–4, 200
- auxiliary bus 262, 288–9
- AVC 172
- average levels 142, 144, 152, 273, 293
- Avid 193, 196, 198–9, 297, 300,
 322, 328

- background music 63, 276
- background noise 63, 114, 235, 238,
 259, 278, 284
 - levels 20–1, 63, 111, 144–5, 210, 278
- background sound 86, 92, 156–7,
 203, 242
- background tracks 242, 246, 259
- backgrounds 203, 243, 245–6, 271–2
- backup vocalists 250
- backups 43, 49, 71, 123, 141, 167–70,
 191–4
- backwards compatibility 32, 301, 316
- balance 216, 234, 268, 272, 277,
 286, 294
 - among tracks 268, 272
 - octave-to-octave 267, 295
- balanced connectors 83, 185
- bands 135, 152, 210, 278, 319, 324
- bass 11, 105, 119, 128–30, 278, 280,
 323–4
 - management 300, 323, 326, 328
- batteries 79, 82, 109, 120–2, 149, 157
 - consequences of diminishing
 voltage 121
- Battleship Potemkin* 218–19
- BBG windscreen 162
- beaches 62–3
- beat 208, 219, 221–2
- Bergman, *Cries and Whispers* 141, 153
- BGs *see* backgrounds
- bidirectional polar patterns *see* polar
 patterns
- bit depth 17–19, 84, 186, 200
- bit-for-bit copies 296
- bit rates 5, 9, 18, 41–2, 44, 184, 316
- bitstreams 314, 316
- Black Hawk Down* 245
- block diagrams for mixing
 270–2, 289
- Blu-ray 4, 10, 15, 22, 41, 300,
 313–15
 - Dolby Media Encoder 314
- BNC connectors 54, 83, 178, 185
- bonus features scene titles 111–13
- booms 60–1, 88–91, 96–102, 112–14,
 130–2, 165, 279–80
 - Fisher 99
 - fishpoles 99–101, 113–14, 118
 - operators 60, 65, 89, 99, 113–14, 118,
 130–1
 - poles 60, 148
- Bourne Supremacy, The* 216
- Boyz 'N the Hood* 246
- brightness 26–8
- broadband noise reduction 284,
 291–2
- broadcast channel 277, 319
- broadcast television 46, 306
- broadcast WAV 187, 189
- broadcasts 29, 31, 38–9, 249, 298,
 300, 316
 - conventional 46, 327
 - digital 313–15
 - television 36, 45–6, 281, 297

- bumps 203, 236, 245, 259, 267, 291, 322
 - fixing 239–42
- bungee cord 117
- busses 261–2, 288–90, 293
 - auxiliary 262, 288–9
 - FireWire 171, 173–4, 191
 - output 262, 288, 290
- buttonhole microphones 106, 110
- BWF *see* broadcast WAV
- cable services 327–8
- cables 7–8, 53, 66, 117, 121–2, 178, 180
 - Firewire 53, 174, 176
 - microphone 7, 99, 102, 133, 284
 - power 7
- calibration 160, 316, 326
 - level 291, 307–10
- camcorders 2, 4–6, 34–5, 47, 51, 68, 171–4
 - consumer 9–10, 22, 42, 47, 51, 95
 - digital 4, 27, 35, 45, 68, 174, 184
 - HDV and Mini-DVD 172
 - prosumer 10, 170
- camera bodies 86, 94
- camera capabilities 2, 4–5, 21, 23, 41, 51, 148
- camera cranes 218, 220
- camera dollies 202, 213
- camera electronics 154, 312
- camera formats 43, 94
- camera inputs 66, 74, 77, 147–9, 158–60
- camera manufacturers 17, 32, 65, 68
- camera models 23, 51, 171
- camera modes 18, 57
- camera operation 59–60
- camera operators 9, 60–1, 65, 69, 90, 112, 169–70
- camera original clones 295
- camera originals 2, 34, 44, 49, 52, 167, 295–6
- camera perspectives 89–91, 99, 132
- camera preamplifier noise 153, 163
- camera recording 31, 167, 195
- camera time codes 48, 68
- camera workflow 64
- cameras *see also* camcorders
 - analog video 27
 - Canon XL-series 233
 - digital 5, 9, 53
 - digital SLR 4, 16, 22, 35, 170–1
 - digital video 4–5, 26, 30, 32–3, 35–6, 75, 170
 - DV 47–8, 141
 - DVCAM 43
 - file-based 25
 - film 33, 61, 114, 130
 - high-definition 6, 36, 41
 - high-end 40, 48, 114
 - mic preamplifiers 166
 - microphone input 145, 154
 - Mini DV 115
 - motion-picture 40
 - off-camera 213, 235, 238
 - professional 7–8, 35, 48, 68, 74, 115
 - HD 22
 - prosumer 8, 9, 10, 35, 38, 52, 115, 170–1
 - and sound workflow 64–72
 - tape-based 94, 174
 - under- and over-cranked 40–1
- Canon 4, 35, 72, 171–2
- capacitance 120
- capacitors 109, 120, 138
- carabiners 117
- cardioid polar patterns *see* polar patterns
- cards 43, 49, 50, 80, 115, 167–71, 192, 296
 - live 169–70
 - memory *see* memory cards
 - sound 79, 83
 - transferred 169
- cassette formats 41–3
- cassettes 5, 171, 296–7
- CCDs (charge-coupled devices) 26, 33, 38, 44

- CD players 8, 78–80, 83, 139, 174, 179, 182–3
- CDs 19, 22, 26, 84, 183, 188, 198
- chain 20, 65, 74, 141, 158–62, 166, 274
 - and multiple level controls 158–61, 166, 274
- channels 288–9
 - audio 16–17, 19, 21, 65, 69–71, 176, 190
 - dialogue 278, 281
 - discrete 16, 115, 303
 - distribution 26, 31, 273, 328
 - first 95, 97
 - front 297–8, 303, 317–18
 - full-bandwidth 318
 - input 77, 79, 191, 261, 288–90, 299
 - LCR 304
 - left 80, 141, 179, 302
 - LFE 298, 315–16, 319
 - LS and RS 299
 - main 303, 316, 319, 323
 - multiple 262, 288
 - output 67, 261–2, 287–90, 308
 - recorded 16, 18, 299
 - right 179, 304, 317
 - screen 205, 286
 - second 95–6, 141, 157
 - surround 246, 286, 298, 301, 304, 315, 327
 - and tracks 261
- characters, foreground 238
- charge-coupled devices (CCDs) 26, 33, 38, 44
- chest mounting 13, 90–2, 96, 106–7, 132, 145, 280
- chin, and sound shadowing 107, 110, 280
- chorus programs 288, 304
- cinema 1, 15, 16, 64, 141, 201, 316, 326–7
 - digital 5, 27, 35
 - masters 295, 316
- cinema sound system 13–14
- cinematography 3, 142–8, 201
- Circle Surround 299, 317
- circus analogy 207
- cities 133, 135, 215, 220
- Citizen Kane* 218
- clamps 114, 117
- clapperboard slates 64, 69, 72, 115, 192–5, 200
- classical music 276, 285
- clipping 82, 155, 161, 259, 274
 - digital 74, 311
 - distortion 160, 259, 311
- clip removal 259
- clips 49–50, 193–4, 227–8, 233–4, 237, 239, 258–9
 - sound 193, 196, 198, 228, 233, 240, 257
 - video 185, 193–4
- clock input, word 182, 185
- clock time 46, 48
- clocks 45, 51–2, 153, 181–2, 185, 193
 - video 51–3, 56, 115
 - wall 46, 51
- close-ups 60, 63, 88–9, 91, 93, 113, 217
- clothing 90–1, 106, 110, 132, 280
- CMOS (complementary metal-oxide silicon) sensors 26
- codecs 5, 32, 41, 43, 53, 226
 - lossy 10
 - and quality modes 41–2
- color bars 77, 296
- colors 4, 26, 28, 46, 54, 98, 110
- comb filtering 98, 196, 197, 286, 288
- Compact Flash cards 5, 171–2, 190–1
- compatibility, backwards 32, 301, 316
- complementary metal-oxide silicon (CMOS) 26
- Complete Main 315–16
- composers 223, 247
- composition 61, 189, 201, 313
- compression 7, 9, 186, 274–6, 292, 294, 300
 - amount of 275–6
 - audio 41–2, 275, 318
 - data 7, 9, 41

- compression (*Cont.*)
 digital video 44
 hand 274–5
 preset 318
 ratios 41, 275, 313
 time-compression 256, 268
 video 4, 25, 44
- compressors 7, 156, 274, 276–8, 318
- computer audio interfaces 18, 204
- computer-based editing 225, 227
- computer sound cards *see* sound cards
- computers 19, 32, 173–4, 183–4, 191–2, 225–6, 228–30
- concealment 206
- concert halls 63, 285
- condenser microphones 120
- conductors 7, 54, 80, 81, 120, 176, 180
- connections 8, 54, 66, 71, 81, 83–4, 179–81
 analog 84, 173
 balanced and unbalanced 80–4
 digital 200
 audio 176, 180, 182
- connectors 8, 54–5, 73–4, 81–4, 120–1, 173–4, 178–81
 analog 179
 audio 83–4
 BNC 54, 83, 178, 185
 FireWire 174, 191
 phono 73, 83–4, 185
 RCA 8–9, 79, 83–4, 179–80
 RJ-45 178
 TRS 80, 83
 XLR 7–8, 73–4, 77–8, 79, 174, 176–7, 185
- consumer camcorders 9–10, 22, 42, 47, 50, 95
- consumer equipments 73, 81–3, 178, 180
- continuity 202–3, 217, 224, 233, 235, 247, 250–1
- continuous recording 2, 184
- continuous sound 7, 87, 249
- continuous time 213, 249, 251–3, 286
- control rooms 310, 322
- conventions 28, 197, 213–14
 film sound 202
- conversations 62–3, 213–15, 217, 219, 245, 250
- conversion 17, 26, 29–30, 33, 225, 287, 289
 sample rate 183, 188, 198
- converters 9, 173, 181, 259, 273, 312
 analog-to-digital 9, 68, 173, 183
 digital-to-analog 78–80, 173, 225–6
- copies
 analog 2, 296
 archival 167–8
 backup 168, 191
 bit-for-bit 296
 digital 169, 295–6
 one-to-one 296
- copy machines 2, 26
- copying 2, 26, 168–9, 184–5, 198
- costs 27, 35, 44, 226, 228, 284, 322
- costumes 86, 90, 96, 132, 201
- coughs 258, 268–9, 294
- counter frames 46
- coverage 85–9, 91, 93, 97–8, 105, 117, 131
 scene *see* scene coverages
- CPU, host 260
- crew 1, 59–61, 63, 65, 69, 86–8, 244
 sound 9, 62, 66–7, 98, 113–14
- cries 148–53
 and whispers 141, 153, 155–8
Cries and Whispers, Bergman 141, 153
- cross-fades 89, 220, 229–30, 232, 237, 248, 259
- crowds 113, 204
- cues 52, 196, 203, 212, 250, 252
 music 248, 251, 253, 255, 302
- curves 280, 282, 307
- CUT-1, Schoeps 161–2
- DACs *see* digital-to-analog converters
- damages 13, 122–3, 161, 176, 281, 291

- DAT (digital audio tape) 174, 182, 190, 192, 198
- data compression 7, 9, 41
- DAWs (digital audio workstations) 225, 230–1, 259, 262–3, 267, 269–70, 279
- editors 258
- Days of Heaven* 219
- dB *see* decibels
- DC offset removal 259
- de-essers 277, 294 *see also* limiters
- dead-hung microphones 131
- dead spaces 208, 235, 247
- decay, sound 63, 284
- decibels 19–20, 145–8, 152–4, 162–6, 272–7, 308–13, 317–19
- definition 15
- decoders 16, 175, 246, 288, 299–304, 319
- Dolby Stereo Decoder 301
- Deep End, The* 214
- delay compensation 289
- delays 285–6, 289
- deleted scenes 109, 111
- delivery masters 46, 47, 295–7, 313–14, 316
- and sound format on edit 297–304
- delivery media 299–300
- devices 4–5, 21–2, 32, 75–7, 82–4, 189–91, 283–4
- line-level 139
- portable 169, 204
- reverberation 284–6, 288–9
- screw-in 152, 161
- DF (drop frame) 46–7, 56, 116, 321
- dialogue 91–2, 207–9, 236, 276–8, 293–4, 302–4, 310–11
- channels 278, 281
- editing 239–40
- editors 203, 256
- foreground 209–10, 236
- levels 210, 267, 272, 277, 310–11, 328
- normal 141, 310
- normalization 314, 316–17, 319
- premixes 270, 292–3
- production 202, 220
- recording(s) 98, 140, 258
- scenes 63, 203, 208, 248
- smoothing 231, 233, 241, 265
- sync 62, 196
- tracks 203, 207–9, 242, 259, 276
- diaphragms, microphone 103, 105, 120, 123, 152
- digital, basic 26–7
- digital audio 17, 26–7, 29–30, 83–4, 182–3, 189, 312–13
- bitstreams 314, 316
- connections 176, 180, 182
- conventional 296, 313
- editing 225, 229
- interfaces 174, 178, 183
- problems 198–9
- recorders 18
- settings 23
- signals 178
- tape *see* DAT
- workstation editing 228
- digital audio file transfers, problems
- for sound accompanying picture 197–200
- digital audio workstations *see* DAWs
- digital broadcasts 313–15
- digital camcorders 4, 27, 35, 45, 68, 174, 184
- digital cameras 5, 9, 53
- digital cinema 5, 27, 35
- digital clipping 74, 311
- digital editing 25, 229
- systems 2, 34, 53
- digital formats 22, 26, 43
- digital interfaces 55, 174, 180–1
- digital output 174, 178–9, 182
- digital peak meters 115, 139, 306, 309, 311
- digital pictures 236, 293
- digital processing 26, 29, 38
- digital recording(s) 2, 25, 74, 75, 141, 171, 200
- digital reference levels 75, 78–9

- digital satellite television 313
- digital signals 10, 25, 173–4, 176
- digital SLRs 4, 16, 22, 35, 170–1
- digital sound 17–21, 173
- digital-to-analog converters (DACs) 78–80, 173, 225–6
- digital-to-analog output boards 305
- digital transmissions 136–7
- digital video 3–4, 21, 25–57, 64, 83, 169, 173 *see also* DV
 - camcorders *see* digital camcorders
 - cameras 4–5, 26, 30, 32–3, 35–6, 52, 53, 75
 - professional 10, 50, 171
- capabilities 72
- compression 44
- editing 49, 171
- and film distinguished 32–6
- formats 21–3, 39
- interchangeability 4, 42–4, 312
- landscape 4–5
- off-line and online 44
- and operational matters 41–53
- productions 1, 66, 115, 169, 174, 193–4
- projects 6, 25, 85
- recording 41, 56–7, 226
- technology 1, 25
- transfers 173, 185
- dimensions 10–11, 15, 127, 246, 278
- dinosaurs 33, 223, 243–4
- dips 280–1, 283
- direct sound 87, 97, 107, 125–6, 138, 161, 219
- directional microphones 92, 104–5, 124–6, 128–30, 136, 161, 195
- directional preference 119, 124
- directional properties 119, 128, 138
- directors 23, 56–7, 59–60, 84, 205, 212, 223–4
- DirectX 260, 286
- disc 2, 32, 34, 42, 49, 56–7, 227
 - optical 5, 25, 43, 167–8, 170, 190, 265
- discrete surround 297–8
- disposable microphones 109
- distances 10, 11, 98, 114, 124–5, 211, 218, 243
- distortion 74, 121, 137, 144, 148, 152, 154–7
 - audible 9–10, 76, 139, 306
 - bad 95, 139
 - clipping 160, 259, 311
 - gross 74, 148
 - overload 161–2
 - potential 75, 121, 155
- distribution
 - channels 26, 31, 273, 328
 - formats 26, 36
 - theatrical 36, 295, 310
- dither 19–20, 141, 287
- documentaries 3, 6, 59–60, 62, 86, 91–2, 238–9
- Dolby 296, 300, 314–15, 319
 - surround modes 317
- Dolby Digital 10, 21–3, 41, 172, 299, 313–15
- Dolby Media Encoder, Blu-ray 314
- Dolby Pro Logic II 299, 317
- Dolby Stereo 16, 301
- Dolby TrueHD 41
- dolly shots 128, 131
- double-system recording 5–6, 17, 27, 64, 68–71, 191–2, 195
 - audio 192, 194
 - management 189–92
 - setups 64, 69–71, 72, 174, 189–90
 - workflow 84
- downbeat 219, 222
- downward expanders 278
- Dr Zhivago* 215
- dream states 216–17
- dreams 216–18, 245, 286
- drift 52–3, 72–3, 199
- Driving Miss Daisy* 273
- drop frame *see* DF
- dry sound 285
- DSLRS 4, 27, 35, 45, 68, 72, 172

- DTRS 296, 319
- DTS format 41, 299–300, 314–15
- DTS-HD 41, 313–14
 - Master Audio streams 314
- dual mono 96–8
- DualEyes 194–5
- dub stages 85, 204–5, 307, 318, 322, 324
- duck 89, 258, 268
- DV 5, 19, 22, 26, 37, 171–2, 190 *see also*
 - digital video
 - camcorders *see* digital camcorders
 - cameras 47–8, 141
 - prosumer 38
 - film festivals 305–6
 - formats 53, 56, 185
 - standard-definition 18, 38, 226
 - frame 38
 - Mini DV 19, 22, 39, 43, 51, 168, 171–2
 - standard 171
 - tape 20, 37, 53, 56, 168, 171
 - formats 4, 17, 19, 41, 173
- DVCAM 19, 22, 37, 52, 168, 171–3
 - cameras 43
- DVCPRO 22, 37, 41, 52, 171–3
- DVCPRO-HD 22, 172
- DVD 4, 10, 15, 22, 41, 313–16, 327–8
- DVD Studio Pro 300, 314
- dynamic microphones 123
- dynamic processes 274, 291
- dynamic range 14–15, 19–21, 75–8,
 - 141–2, 146, 153–7, 312–13
 - available 78, 137
 - capabilities 15, 19
 - compression 7
 - control 273, 314
 - wide 141, 144, 155, 157, 273,
 - 287, 318
- DZC 20 152, 162–3
- ear 11, 13–15, 94, 142, 222, 279, 285
- Edit Decision List *see* EDL
- edit masters 44, 270–1, 287, 296, 298,
 - 301, 327
- edit room monitors 322, 324
- edit rooms *see* editing, rooms
- editing 225–65
 - environment 30, 185
 - modes 227
 - non-destructive 227–8
 - non-linear 225–6
 - systems 44, 50, 227
 - process 21, 44, 206, 258–9
 - production sound 234–7
 - random-access 227
 - rooms 42–3, 68, 193, 198, 310,
 - 322, 324
 - software 34, 43, 196, 225, 227,
 - 263, 308
 - sound 87, 187, 196, 228, 231–5,
 - 238–9, 249
 - sound effects 242–5
 - stage 259, 268
 - systems 47–9, 52–3, 167–8, 170–1,
 - 173–4, 181–3, 185–6
 - computer-based *see* editing,
 - software
 - non-linear 44, 49, 227
 - off-line 44, 47
 - picture 197, 230–1, 263–4
 - sound 2, 263, 265
 - visual waveform 228–9
- editorial operations 197–8
- editors 95–6, 196–7, 236–7, 239, 256–8,
 - 267–8, 322
- DAW 258
- dialogue 203, 256
- picture 111, 198, 208, 231–2, 239,
 - 265, 267
- sound 96, 198–9, 205, 208, 231, 238–9,
 - 263–4
- film 263
- supervising 211, 217, 224
- video 97, 248
- edits
 - dialogue 294
 - and fade files 229–30
 - picture 87, 198–9, 233, 247
 - seamless 230, 247

- edits (*Cont.*)
 - sound 49, 98, 203, 210, 231, 248, 265
 - prelap 248
- EDL (Edit Decision List) 44, 49–50, 187, 230
- electrical energy 119, 138–9
- electrical voltage *see* voltage
- electrodynamic microphones 158
- electron guns 31, 34
- electronics 20, 55, 145, 152–3, 155, 161–3, 323
 - camera 154, 312
 - microphone 152, 163, 166
- electrostatic microphones 120–1, 123, 138, 158
 - Logic Audio 260
- encoders 175, 299–302, 316
- energy
 - acoustical 119, 139
 - electrical 119, 138–9
 - sound 120, 126
- envelopment 16, 303, 327
- environment 21, 30, 43, 93, 243–4, 267, 326
 - editing 30, 185
 - monitor 272, 294
 - multi-camera 194
 - studio 1, 70, 301
- equalization 110–13, 233, 239–40, 267, 269, 278–82, 291–5
 - characteristic effects 279
- equalizers 240, 278–9, 281, 283, 326
- errors 46, 175, 302, 325
 - sync 200
- establishing shots 62–3, 92, 131, 249–52
- excessive noise 125–6, 141, 311
- excessive reverberation 62, 64, 86, 98, 125–6
- expanders, downward 278
- export 49, 197, 199–200, 228
 - improper setup 199
- extensions 159, 187–8, 256, 298–9
 - file 186–7, 197
- exterior 13, 111–13, 149, 213, 245–6, 249–52, 254–5
 - perspective 211
- external hard drives 174, 190–1
- external inputs 6, 65
- external microphones 18, 99, 120–1, 158
- external mixers 67, 95, 115, 122, 157–9, 268–9, 308–9
- external power supplies 154, 158, 164
- fade files 188, 229–31, 263, 265
- fade-in 230, 248
- fade-out 228, 230, 248
- faders 15, 79, 234, 258, 268–9, 274, 286
 - channel level 79, 310
- fades 199, 229–30, 236
- Fahrenheit 9/11* 212
- FCC (Federal Communications Commission) 133
- feature films 61, 89, 93, 96, 141, 199, 233
- fiction filmmaking 3, 60, 85–6, 97, 113, 117, 132
- field mixers 67, 153–4
 - portable 122–3
- field recorders 22, 71, 191, 204
 - production 189, 204
- fields 4, 25, 28, 30–5, 38–9, 45, 307
 - magnetic 7, 82, 178
 - second 31, 33
 - sequences 29, 56
 - sound 15, 95, 125, 129, 301–2, 304–5, 317
 - video 28, 30–1
- file-based cameras 25
- file-based recording 5, 17, 168
- file-based systems 4, 25
- files
 - audio 169, 189, 191–2, 227, 230–1, 257, 265
 - formats 188–9, 200
 - digital video 313
 - export *see* export

- extensions 186–7, 197
- fade 188, 229–31, 263, 265
- formats 36, 43, 170, 186, 188–9, 200
- management 167, 230–1
- new 2, 168, 195
- OMF 189
- operations 197
- project 230–1, 265
- transfers 53, 167, 169, 171–88, 191–2, 200
- types 188, 265
- video 167, 188–9, 191–2
- fill see presence*
- film 3–5, 27–36, 50–1, 125–6, 173, 201–2, 227–9
 - cameras 33, 61, 114, 130
 - and digital video distinguished 32–6
 - feature films 61, 89, 93, 96, 141, 199, 233
 - frames 28–9, 36
 - independent films 3, 167
 - look 32–6
 - magnetic 229, 261
 - masters, analog 75, 306
 - mixes 290, 326
 - production 115, 119, 122, 136, 209
 - sound 122, 156, 201–2, 206, 221, 298
 - conventions 202
 - styles 208–24
 - v video mixes 326–8
- Film Light Compression 318
- Film Standard Compression 318
- filmmakers 1, 3, 15, 87, 92, 111, 219–20
- filmmaking 85, 117, 130, 133, 201, 205, 280
- filters 135, 161–2, 206, 213–14, 281–3, 319
 - comb 197, 288
 - high-pass 162, 166, 281, 291, 323
 - low-pass 141, 281–2, 291, 319, 323
 - neutral density 148, 166
 - notch 283, 289, 291
- Final Cut 183, 193, 196
- FireWire 141, 171, 174, 183–4, 192
- 6-pin 184
 - 400 171, 183–4
 - 800 184
- bus 171, 173–4, 191
- cables 53, 174, 176
- connectors 174, 191
- interconnects 183–4
- interface 22, 190–1
- fishpoles 99–101, 113–14, 118
- fixed-frequency system 111, 135
- flash memory cards 5, 70, 167
- flashbacks 217
- flicker 27–8, 31
- floor 99, 113, 125, 143, 323, 326
 - noise 20, 75, 146, 153–4, 163–4, 259, 312
 - weighted 141, 146
- FM (frequency modulation) 135–7, 297
- foam windscreens 104–5
- focal length 4, 60, 114
- focus 33, 61, 270
- Foley 93, 165–6, 204, 207, 216–18, 270–1, 293
 - effects 93, 153, 204–5, 246–7
- footsteps 12, 276–7
 - stages 204, 246
- footsteps 112–13, 204
 - Foley 12, 276–7
- Forbidden Planet* 15
- Foreground
 - action 210, 277
 - characters 238
 - dialogue 209–10, 236
- foreground sound 155–6, 259
- formal structure 85–6
- formats 4–5, 21–3, 25, 39–40, 43, 51–2, 186–9
 - AAF 187, 189, 231
 - AES 83, 181, 185, 187–9
 - AES 31 187–9
 - archival 43, 168
 - audio 10, 41–2, 188, 314
 - audio files 188–9, 200
 - audio recording 6, 9–10, 95, 172

formats (*Cont.*)

- cameras 43, 94
- cassettes 41–3
- common 21, 184
- digital 22, 26, 43
- digital tape 170, 300
- digital video 19, 21–3, 25, 37, 39, 56
- distribution 26, 36
- Dolby 41, 299–300, 314–15
- DTS 41, 299–300, 314–15
- DV 53, 185
 - tape 4, 17, 19, 41, 173
- encoding 10, 22
- files 36, 43, 170, 186, 188–9, 200
- interlaced 30, 32–3
- lossy 6, 10, 21, 23
- new 167, 188
- OMF 187, 189
- PAL 34, 45, 51
- professional 41, 48, 168
- recording 18–19, 25, 32, 189
- sound 25, 297–304
- standard-definition DV 18, 38, 226
- tape 5, 10, 18, 21, 34, 42, 168
- tape-based 25, 51, 53, 189
- video 2, 5, 40, 47, 56

frame boundaries 34, 263

frame lines 61, 91, 114

frame rates 27–30, 38, 40, 46, 53, 56, 65, 142

- higher 28, 33
- integer 40, 72

frames 26–8, 30–6, 45–6, 51–3, 60–1, 91, 263–4

- counter 46
- film 28–9, 36
- full 38, 40
- single 28, 45
- video 28, 30–1, 33, 35, 46, 51, 195

framing 60–1, 86, 93, 95

freeways 62–3

frequency 11–12, 14, 18, 110–11, 135, 278–81, 283–5

- audio processes related to 278–83

- bands *see* bands
- combinations of level and 284
- equalization 278–81
- high frequencies 10, 110, 130, 132, 277, 280–1, 289
- higher frequencies 11, 128, 133, 135, 323
- highest frequencies 18, 124
- low frequencies 144, 222, 224, 281, 303, 322
- modulation (FM) 135–7, 297
- radio frequencies 111, 133, 135, 137
- range 11–12, 17–18, 135–6, 155, 214, 221, 278–9
 - of hearing 18
 - wide 114, 325
- response 81, 97, 142, 146, 154, 280, 322–6
 - flat 144, 277
- spectrum 133, 283, 287
- weighting 14, 142, 144

frequency-agile systems 111, 135

front channels 297–8, 303, 317–18

front speakers 316, 325

full resolution 44, 45, 49–50, 167, 171, 295

futz 221

Gaffer's tape 117

gain 7, 15, 155–7, 257–8, 268, 274–6, 311

- controls 79, 137, 140–1, 153, 274–5, 278, 292
 - automatic 7, 156, 275
- function 258, 268
- plug-ins 239, 268
- reduction 276, 292, 313
- staging 20, 79, 274, 311
- unity 79, 160, 269, 287, 308

generation loss 2, 10, 26

GenLock 48, 57

gross distortion 74, 148

ground 80, 121, 123, 149, 151, 215

ground loops 180

grouping 232, 262–3

- guidetracks 6, 69, 71, 94, 193–6, 234
- hair 90, 132
- hairpiece tape 110, 117
- hammer transients 273
- hand compression 274–5
- hard clipping 74
- hard disks, storage capacity 44
- hard drives 17, 43–4, 55–6, 70, 167–71, 191–2, 225–6
 - external 174, 190–1
 - file systems 43
 - internal 190–1
- hard effects 93, 205–7, 212, 224, 232–4, 242–3, 293
- hard sync 193–4, 196–7, 200, 205, 235, 327
- hardware 29, 51, 182–3, 188, 229, 268–9, 300–1
- harmonics 155, 283
- HD cameras, professional 22
- HD satellite services 327–8
- HDCAM tape 17, 21–2
- HDMI 55
- HDTV (high-definition television) 31, 36, 55, 298
- HDV 10, 22, 171–2
- headphone level controls 67
- headphone monitoring 67, 71
- headphones 9, 60, 66–7, 83, 114, 175, 235–6
 - outputs 66, 71
- headroom 74–8, 141, 152–3, 161–2, 273, 306–7, 312–13
 - and reference levels 74–6
- heads-to-tails 237, 257
- hearing 13–15, 20, 139, 144, 216, 287, 315–16
 - damage 13–14, 152
 - threshold 139, 153
- helicopter blades 243
- Hi-8 tape 296
- high-definition cameras 6, 36, 41
- high-definition standards 32, 36, 38
- high-definition video 1, 5, 38, 41, 44, 56
- high-pass filters 162, 166, 281, 291, 323
- highs 277, 279–80, 282, 291
- hiss 21, 112, 145, 166
- holes in sound 144, 278, 291, 313
- Hollywood 93, 96, 205–6, 215, 267, 274, 317
 - productions 5, 111
 - studios 75, 99, 206
- home masters 295, 316
- home mixes 326
- home playback 273
- home receivers 323–4
- home systems 16, 278, 295, 297, 310, 316, 326–7
- Home THX 295, 326
- host CPU 260
- Host Time Division Multiplex (HTDM) 260, 286, 292
- hours counter 48–9
- HTDM (Host Time Division Multiplex) 260, 286, 292
- hum 7, 81, 180, 283
- hyper-realism 217 *see also* realism
- hyper-reality 214–15, 224, 244
- hypercardioid microphones 126–8, 130, 138, 156
- I, Robot* 223
- identification, verbal 69
- IEEE 1394 175, 184
- IHF Match Box 82
- iLink *see* FireWire, 400
- images 7, 26–8, 31–2, 35–9, 43, 88, 142
 - phantom 263, 297, 301, 304, 322
 - quality 38, 42
 - scanned 31–2, 34
 - sensors 4, 26, 33–6
 - visual 60, 63
 - widescreen 36, 38
- impedances 80–1, 149, 178, 180

- import 49, 173–4, 188, 192, 200, 308
- incoming scenes 247, 249, 251
- infrasonic overload 161
- input capabilities 23, 139, 162, 166
- input channels 77, 79, 191, 261, 288–90, 299
- input level 7, 95, 144, 154, 275, 292, 309
- input overload 146–7, 153, 166
- input trim controls 79, 309
- inputs
 - audio 6–7, 53, 56, 64, 166
 - camera 66, 74, 77, 147–9, 158–60
 - external 6, 66
 - mic-level 73
 - microphone 7–8, 16, 20, 139, 141, 153–5, 159–60
 - word clock 182, 185
- instant playback 53
- integer frame rates 40, 72
- interchange 74–5, 122, 186–7, 199
- interchangeability 4, 42–4, 312
- interconnecting video 53, 56–7
- intercutting 199, 202–3, 205, 208, 217, 224, 291
- interfaces 55, 173–4, 178, 180, 185, 258
 - AES3 173, 185
 - computer audio 18, 204
 - dedicated audio 174
 - digital 55, 174, 180–1
 - FireWire 22, 190–1
 - professional 174, 178
- interference 29, 84, 88, 90, 111, 135, 158
- interference tube *see* shotgun microphones
- interior shots 245
- interiors 112–13, 251–5
- interlaced scanning 30, 33–4, 45
- interlacing 28, 30–4, 45, 56
- intermediate masters 296
- internal hard drives 190–1
- interviews 63, 65, 90, 93, 238–9, 253
- isolation 82, 244–5
- jam sync 48, 56, 68, 193
- jungle scenes 222, 244, 304
- Jurassic Park* 223, 243
- JVC camcorders 172
- large rooms 318, 326
- lavalier microphones 89–91, 96–7, 106, 109, 131–3, 280, 282
- lavalier mounting, tape for 117
- lavalier placement 96, 280
- lavalier tracks 197, 291
- lavs *see* lavalier microphones
- layering 204, 206, 223, 234, 243
- left channel 80, 141, 179, 302
- left signal 80, 121, 301
- left surround 261, 288, 303, 308, 317
- lenses 4, 36, 60, 114, 148, 209
 - anamorphic 36–8
 - cinema 4, 36
 - zoom 4, 60, 94
- L_{Aeq} 14, 143
- level 73–80, 143–5, 155–60, 258–9, 267–78, 292–4, 308–13
 - combinations of frequency and 284
 - line levels 7–8, 73–4, 77, 78–80, 154, 160, 190
 - microphones 73–4, 77–8, 82, 158–9
 - peak 74, 95, 139–44, 152, 258, 273, 306
 - speaker 73–4
- level calibration 291, 307–10
- level controls 65–7, 79, 148, 160–1, 272–3, 305–6, 308–9
 - and chain 158–61, 166, 274
 - headphone 67
 - main recording 145, 155
 - manual 6–7, 27, 65, 67
 - mixer 158–9
 - multiple 158–61
- LFE (Low Frequency Effects) 261, 298, 303, 315–16, 319, 321, 326
- libraries, sound effects 99, 205, 232, 244, 247, 265, 304

- light 25–6, 108, 142, 148, 201, 212–13, 252
 - meters 142
- limitations 2, 5–6, 18–19, 40, 96, 186, 206
 - caused by routing 288–90
 - voice 290
- limiters 136, 138, 156, 274–7, 315
 - see also* de-essers
- limiting 276–7
- line-level inputs 8, 18, 20, 73, 139
- line-level outputs 9, 73
- line-level signals 8, 79, 154
- line levels 7–9, 73–4, 77, 78–80, 154, 160, 190
- linear PCM 10, 22–3, 189, 299, 313–14
- lip sync 52
- lips 91, 132, 196, 235, 280
- listening location 298, 304, 324
- little endian information 188
- live monitoring 53
- localization 15–16, 294, 303
- locations 21, 59–64, 85, 88–9, 98–9, 111–12, 209–10
 - listening 298, 304, 324
 - noisy 62, 112, 238
 - quiet 62, 63, 92, 113, 203
 - scouting 59, 61–4, 84, 112, 144
 - selection 3, 61–2
 - windy 138
- locking of pictures 183, 232–3, 293
- long play (LP) 42, 56, 297
- loops 88, 107, 233, 236–7, 294
 - ground 180
- lossy format 6, 10, 21, 23
- loud passages 155–6
- loudness 73, 136, 258, 272–3, 294, 305–6, 311
- loudspeakers 73, 83, 119, 220–1, 297, 309, 323–4
 - monitor 174, 322
 - surround 317, 325
- Love Actually* 220, 248–55
- low-cut 161, 281
- Low Frequency Effects *see* LFE
- low resolution 44, 171
- LP (long play) 42, 56, 297
- LtRt (left total, right total) 298–9, 302, 314, 317, 327
- Mac programs 260
- magic surrounds 302
- magnetic fields 7, 82, 178
- magnetic film 229, 261
- main channels 303, 316, 319, 323
- male voices 279–81, 291
- manual level controls 6–7, 27, 65, 67
- Marantz 22, 190–1
- marker 69
- market 4–5, 17, 27, 36, 135, 137–8, 296–8
- master audio 71, 194, 296, 314
- master mixes 293
- master production audio 194
- mastering 1, 22, 273, 293–4, 301, 319, 326
 - for DVD, Blu-ray, digital broadcast, and digital satellite television 313–21
 - for level 305–6
 - stages 302, 313, 326
- masters 26, 34, 49, 113, 297, 301, 303
 - delivery 46, 47, 295–7, 313–14, 316
 - digital 306
 - edit 44, 270–1, 287, 296, 298, 301, 327
 - home 295, 316
 - one size fits all 295
 - one-size-fits-all approach 311–13
 - theatrical 295, 327
 - word 327
- materials 34, 40, 43, 45–6, 52, 174, 199–200
- matrices 246, 297–9
 - amplitude-phase 298–9, 302, 304, 317
 - surround 298–300
- maximum recordable level 20, 139

- maximum undistorted level 76, 146, 166
- media 1–3, 15–17, 48–9, 75, 141–2, 155–6, 295–9
 - analog 75
 - audio 189, 192
 - capabilities 15
 - delivery 299–300
 - dynamic range 78, 142, 146, 155, 159
 - improper and no labeling of 199
 - internal 191
 - management 43, 167–200
 - recording 4–5, 20, 40–4, 64, 74, 167, 190–1
 - solid-state 53, 170–1
 - source 50, 319
 - tapeless 25, 65
 - types 43, 170–2
 - capacity 190
 - video 68, 192, 200
- memory 181–2, 222, 228 *see also* RAM (random access memory)
- memory cards 17, 25, 42–4, 52, 167–71, 190–2, 200
 - Compact Flash 5, 171–2, 190–1
 - flash 5, 70, 167
 - Panasonic P2 172
 - SD 171, 191
 - SDHC 5, 172, 190
 - SDXC 172
- metadata 49, 186–7, 189, 313, 319
- meters 15, 74, 142, 272, 287, 309, 311
 - audio 77, 84, 140–1
 - peak *see* peak meters
 - sound level 144–5, 307, 309–10
 - VU 75, 79, 309, 311
- mic preamps *see* microphone preamplifiers
- microphone attenuate (MIC ATT) switch 139, 175
- microphone preamplifiers 65–7, 139, 147, 155, 161, 166, 174–5
- microphones 88–99, 106–11, 119–25, 127–30, 138–9, 144–9, 152–4
 - accessories 99–105, 122
 - body 120, 132, 161
 - boom *see* booms
 - buttonhole 106, 110
 - cables 7, 99, 102, 133, 284
 - channels 69–71, 98
 - radio 110, 137
 - chest mounting 13, 90–2, 96, 106–7, 132, 145, 280
 - condenser 120
 - dead-hung 131
 - diaphragm 103, 105, 120, 123, 152
 - differences depending on type of polar pattern 128–33
 - directional 92, 104–5, 124–6, 128–30, 136, 161, 195
 - disposable 109
 - dynamic 123
 - electrodynamic 158
 - electronics 152, 163, 166
 - electrostatic 120–1, 123, 138, 158
 - external 99, 158
 - hypercardioid *see* polar patterns
 - inputs 7–8, 16, 139, 141, 145–6, 153–5, 159–60
 - lavalier 89–91, 106, 109, 111–12, 131–3, 280, 282
 - mounting 106–9
 - level 73–4, 77–8, 84, 158–9
 - lines 149, 158
 - mixers 73, 154, 159–60
 - models 119–20, 152, 161, 281
 - multiple 8, 204
 - Neumann 154
 - off-mic sound sources 87, 91, 106, 234, 236, 238
 - omnidirectional *see* polar patterns
 - on-board 6, 16, 18, 195
 - on-camera 20, 68, 87–8, 92–5, 97, 131, 194–6
 - on-mic sound sources 91, 106, 124–5, 236, 247

- outputs 74, 139, 145, 166
 - and production sound 139–66
- perspectives 87, 89, 91, 114, 118, 204, 208
- pickup capsule 152, 161–2
- placement 5, 89
- planted 89, 131
- polar patterns *see* polar patterns
- production sound 119–38
- professional 121, 139
- radio 64–5, 90–2, 96–7, 109–10, 112–14, 132–8, 146
 - lavaliere 89, 112
- Schoeps 112, 122, 152–4, 157, 161–4
- selection 61, 119, 131
- Sennheiser 122, 165–6
- single 65, 157
- smallest 124, 126
- stereo 95, 121, 126, 165
- supercardioid *see* polar patterns
- technique 62, 85–6, 131
 - and scene coverage 88–92
- types 103, 122, 282
- video 126, 128
- midbass 278–80, 323
- Midnight in the Garden of Good and Evil* 125
- midrange frequencies 128, 130, 132, 144, 277–8, 303
- Mini-DVD camcorders 172
- MiniDV 22, 43, 51, 168, 171–2
 - camcorders 18
- Minnetonka Audio SurCode 314–15
- mise en scene 201
- mix *see* mixers; mixes; mixing
- mix busses 261
- mix rooms 269, 310
- mixdown 69–71, 316–17, 324, 326
- mixers 67, 69–71, 77–8, 121–2, 154–5, 158–64, 190–1
 - expensive 158
 - external 67, 95, 115, 122, 157–9, 268–9, 308–9
 - field 67, 153–4
 - human 156
 - level controls 158–9
 - microphone 73, 154, 159–60
 - mixer-recorders 69, 191
 - music 313
 - outboard 166, 190
 - outputs 71, 74, 158–60, 191
 - over-the-shoulder 65–6
 - portable field 122–3
 - production sound 99, 111, 118, 146
 - re-recording 146, 205, 220
 - Shure 122, 154–5, 164
 - well-trained 274
- mixes
 - balance 234
 - film 290, 326
 - final 10, 64, 97, 261, 270, 291
 - home 326
 - master 293
 - surround 21–2
 - theatrical 326
 - two-channel 191
 - universal 328
- mixing 30, 66, 267–94
 - complex 261
 - consoles 83, 205
 - control surface 269
 - eight-channel 71
 - film 293
 - hourglass 270–1
 - how to 290–4
 - modes 258, 268
 - multichannel 5
 - music 313
 - on-camera 77
 - postproduction 15, 96–7
 - software 319
 - stem 261, 294
 - virtual 270
- modes 19, 30–3, 35, 38, 40, 42, 47–8
 - anamorphic video 56
 - cameras 18, 57
 - and codecs 41–2
 - editing 227

- modes (*Cont.*)
 - motion picture 171
 - recording 4
 - movie 9, 35
 - progressive scan 32, 38
 - recording 4, 10, 18, 42
 - scan 32, 38
 - standby 47
 - video 4, 42, 56–7
- modulation 188, 228
- monitoring 8–9, 54, 60, 71, 115, 159–60, 295–328
 - headphone 66, 71
 - live 53
 - one-size-fits-all approach masters 311–13
- monitors 52, 114, 159–60, 268–9, 300–1, 309, 322–4
 - edit room 322, 324
 - flat 318, 326
 - level 145, 272, 300
 - path 302
 - picture 204, 322
 - speakers 198, 322
 - systems 272, 290, 305, 307, 311, 328
 - video 4, 61
 - X curve 307, 318
- mono 15, 83, 95, 186, 297–8, 301, 326–7
 - dual 96–7
 - programs 297, 303
 - signals 83, 96
 - surround 315
- montage 202, 208, 216, 218–20, 244, 248, 250
- Moonrise over Hernandez* 142
- motion pictures *see* movies
- Motorola 186
- MOTU Audio Systems (MAS) 260
- movie modes 9, 35
- movies 2, 4, 125, 219–20, 243–4, 246, 306–7
 - action-adventure 206, 215, 272, 310
- MPEG-2 5, 41, 171–2
- MPEG-4 41, 172
- MPEG Layer II audio 21
- multi-camera environment 71, 194
- multi-camera productions 167, 193
- multi-channel recording 15, 94, 246, 286, 297, 302, 320
- multipath 110, 136–7
- multiple level controls 158–61
- Murch, Walter 207, 220–1, 232
- music 209, 216–22, 232–3, 247–52, 254–5, 261, 270–1
 - background 63, 276
 - classical 276, 285
 - cues 248, 251, 253, 255, 302
 - cutting 247
 - editing 229, 247
 - industry 272, 312
 - live 152
 - mixers 313
 - mixing 313
 - mysterioso 254
 - recording 130, 328
 - scored 219, 247
 - source 202, 210, 212, 220, 247, 249–50, 252–3
 - studios 105, 250
 - temp 231–2
 - tracks 207, 294
- music cue score 254–5
- Music Light Compression 318
- Music Standard Compression 318
- mute 259, 262, 309
- narration 125, 285
- narrators 209
- natural roll-off 307
- NDF (non-drop-frame) 40, 46–7, 50–1, 56, 116, 321
- near-sync 196, 205
- Neeson, Liam 249–50, 252, 255
- network 273, 312–13
 - television 297
- Neumann 154–5, 164–6

- neutral density filters 148, 166
 - noise *see also Introductory Note*
 - acoustical 21, 145
 - background 63, 114, 235, 238, 259, 278, 284
 - level 20–1, 63, 111, 144–5, 210, 278
 - boom panning 161
 - camera 94, 291
 - camera preamplifier 153, 163
 - clothing 91, 110
 - computer 322
 - crowd 317
 - excessive 125–6, 141, 311
 - exposure 14
 - floor 20, 75–6, 146, 153–4, 163–4, 259, 312
 - weighted 141, 146
 - gate 258–9, 277–8, 291
 - handling 96, 128
 - high-frequency 141
 - level 21, 144, 154–6, 284
 - loud 23
 - loud crowd 317
 - low-frequency 103, 141, 161
 - low-level 141
 - pink 278–9, 287, 307, 309, 321
 - reduction 136, 259
 - broadband 284, 291–2
 - environmental 124
 - room 90, 96, 112, 285
 - sources 62–3, 94, 126, 130, 235
 - undesired 263, 291
 - wind 105, 112, 128, 161, 163, 166
- noise gate 258
 - noisiness 62, 73, 89, 92, 139, 155–7, 166
 - noisy locations 62, 112, 238
 - nominal pause mode 256
 - nominal reference levels 75, 77
 - non-destructive editing 227–8
 - non-drop-frame *see* NDF
 - non-interlaced scanning 32
 - non-linear editing 225–6
 - systems 44, 50, 227
 - normalization 239, 257–8, 268
 - North America 38–9
 - notch filter 283, 289, 291
 - NTSC 30, 34, 38, 40, 45, 56, 72
 - devices 32
 - television 45, 56, 318
 - video 27, 29–30, 32–5, 56
 - ocean 62
 - octave-to-octave balance 267, 295
 - off-axis sound 124, 126, 128, 130, 154, 156–7, 221
 - off-line systems 44, 46, 47, 50
 - off-mic sound sources 87, 91, 106, 234, 236, 238
 - off-screen sound 69, 201, 213, 222, 224, 294
 - design 214–15
 - off-speed playback 206, 223
 - OMF format 187, 189
 - omnidirectional microphones 104–5, 124, 126, 128–9, 138, 166
 - omnidirectional pickup
 - capsules 162
 - omnidirectional polar patterns 123–4, 126–7, 161
 - on-axis sound 126, 157, 195
 - on-board microphones 6, 16, 18, 195
 - on-camera microphones 20, 68, 87–8, 92–5, 97, 131, 194–6
 - on-camera mixing 77
 - on-mic sound sources 91, 106, 124–5, 236, 247
 - on-screen sound 87, 214–15, 285
 - one size fits all masters 295
 - online sessions 49–50, 167
 - online systems 44, 50
 - operators 7, 20, 40, 86, 90, 112, 146
 - booms 60, 66, 89, 99, 113–14, 118, 130–1
 - optical discs 5, 25, 43, 167–8, 170, 190, 265 *see also* Blu-ray; CDs; DVD
 - orchestral scores 202, 276
 - orchestration 247, 249

- originals
 - camera 2, 34, 44, 49, 52, 53, 167, 295–6
 - film 30
 - tape 43
- Out of Africa* 216
- outboard equipment 270, 284, 286, 289, 291
- outboard mixers 166, 190
- output busses 262, 288, 290
- output channels 67, 261–2, 287–90, 308
- output levels 121, 158, 162, 258, 275, 292
- output voltage 80, 82, 123, 139
- output volume 308
- outputs 67, 78–82, 147–9, 154–5, 178–9, 261–2, 288–9
 - analog 179, 183
 - audio 67, 77, 83
 - digital 174, 178–9, 182
 - final 115, 295
 - headphones 67, 71
 - line-level 8, 9, 73
 - microphones 73, 139, 145, 166
 - mixer 69–71, 73, 158–60, 191
 - signal 148
 - surround 302, 304
- outtakes 208, 235, 237
- over-the-shoulder mixers 65–6
- over-the-shoulder shots 235–6
- overload 74–6, 137, 139, 147, 155, 161–2, 318
 - distortion 161–2
- pads 148–50, 152–5, 157–8, 161–6
 - pre-electronics 152, 158
- pain, threshold of 11, 13, 14, 20, 152
- PAL 30, 50, 51, 56
 - formats 34, 45, 50
 - video 51, 56
- pan pots (panoramic potentiometers) 243, 262
- Panasonic 163–4, 172
 - camcorders 172
 - cameras 145, 148, 153–4, 171
 - HPX170 171–2
 - HDC 172
 - HVX200 40, 171–2
 - P2 cards 172
- panning 33, 161, 262, 286, 288, 293, 303–5
- panoramic potentiometers 243, 262
- pans 33, 61, 84, 202, 289, 317
- pass-bys 211
- PCM 6, 10, 21–3, 188, 299, 314
 - linear 10, 22–3, 189, 299, 313–14
 - uncompressed 9–10, 22–3, 42, 44, 172
- peak levels 74, 95, 139–44, 152, 258, 273, 306
- peak meters 272, 287, 291, 306–7, 309
 - digital 115, 139, 306, 309, 311
- peak-to-average ratio 273
- peaks 69, 143–4, 152–3, 272, 280–1, 291–3, 311–13
 - highest 273, 311–12
 - instantaneous 144, 152, 257
 - short-term 13, 148
- performance 2, 136–7, 156–7, 164, 234–5, 257–8, 275–7
- perspectives 16, 88–90, 96–7, 131, 211, 213, 245–6
 - camera 89–91, 99, 132
 - distant 210
 - microphone 87, 89, 91, 114, 118, 204, 208
- PFX 209, 271, 291, 293
- phantom images 263, 297, 301, 304, 322
- phantom power 121–3, 148–9, 154, 158, 163–4, 263, 324
- phone calls 214, 250
- phone plugs 79
- phones 4, 9, 11, 14, 119, 213, 282
- phono connectors 73, 83–4, 185
- phonograph records 225, 297
- pickup capsule 152, 161–2
- picture change 227, 247
- picture department 198–9

- picture editing systems 197, 230–1, 263–4
- picture editors 111, 198, 208, 231–2, 239, 265, 267
- picture edits 87, 198–9, 233, 247
- picture monitors 204, 322
- picture quality 25, 42, 44, 316
- picture-sound sync 263–4
 - resolution 264–5
- pink noise 278–9, 287, 307, 309, 321
- pitch 11, 29–30, 183, 188, 198, 211, 256–7
 - correct 183, 287
 - correction 29, 287
 - perfect 30
 - shifting 51, 223, 240, 243, 256, 288, 304
- pixels 26, 31, 36–9
 - non-square 37, 39
- playback 8–9, 16, 30–1, 43, 52–3, 55–6, 200
 - home 273
 - instant 53
 - level 21, 258
 - off-speed 206, 223
 - techniques 295
- plug-ins 206, 257, 259–60, 264–5, 267–8, 286–9, 291–2
 - gain 239, 268
 - and processes 249–60
 - reverberation 240, 284, 286
- plugs 79, 81–3, 121, 148, 153, 181–2
 - see also* connectors
- PluralEyes 71, 194–5
- polar patterns 119, 123–33
 - bidirectional 126–7
 - cardioid 105, 126–7, 129–30
 - differences between microphones
 - depending on type of 128–33
 - hypercardioid 126–8, 130, 138, 156
 - omnidirectional 123–4, 126–7, 129, 161
 - supercardioid 126, 128, 130
- polarity 122
- police 112, 211, 215
- pop suppression 105
- popper stoppers 105
- pops 129, 222, 296, 321
- portable devices 169, 204
- portable field mixers 122–3
- positions 16, 27, 86–7, 89, 91, 263, 299–300 *see also* locations
 - crew 59
 - poor 100–1, 110, 132
- postproduction 2–3, 30, 42–4, 85–6, 93–5, 112–14, 210–11
 - high-end 26, 284
 - mixing 15, 96–7
 - process 231, 295
 - sound 109, 111, 157, 205, 264, 297
 - studios 306, 320
 - techniques 111
 - tools 210
- power 3, 7, 15, 120–3, 133, 149–50, 157–8
 - cables 7
 - phantom 121–3, 148–9, 154, 158, 163–4, 263, 324
 - processing 200, 230
 - supplies 120, 138, 163
 - external 154, 158, 164
- powering 121–3
- pre-electronics pads 152, 158
- preamplifiers 20, 147, 153, 161, 163, 166
 - microphone 66–7, 139, 147, 155, 161, 166, 174–5
- Premiere 193, 260, 286
- premixes 270–2, 292–3
- preproduction 52, 53, 63, 68
 - planning 59–84
- presence 239–40 *see also* room tone
- presentation
 - Bonus Feature 111–13
 - linear 239
 - realistic 213, 216
 - scaled 267
- Pro Tools 113, 196, 204, 231, 239, 260, 291–2

- processes 114–15, 142–3, 205, 225–6, 231–2, 238–9, 292–4
 - audio 7, 267, 270–88
 - dynamic 274, 291
 - and plug-ins 249–60
 - sound 294
- producers 2, 33, 47, 62, 65, 297–8, 305–6
 - program 295, 317
- production audio 23, 189, 194
 - masters 194
 - recording 193–4
- production dialogue 202, 220
- production effects 92–3, 118, 209, 234
- production field recorders 189, 204
- production sound 5, 15, 85–166, 204–5, 209, 232–5, 265
 - common problems 114–15
 - example 109–13
 - fine editing 234–7
 - general considerations 85–118
 - logs 115
 - DV 116
 - microphones 119–38
 - dealing with output 139–66
 - mixers 99, 111, 118, 146
 - non-voice items 98–9
 - original 208, 233
 - recordings 2–3, 21, 86, 95–6, 124, 284
 - primary 192, 195
 - tracks 196, 203, 209, 232, 234–5, 238, 291–2
- production tracks *see* production sound, tracks
- professional cameras 7–8, 35, 48, 68, 74, 115
 - digital video 10, 50, 171
 - HD 22
- professional equipment 29, 79, 80
- professional formats 41, 48, 168
- professional interfaces 174, 178
- professional microphones 121, 139
- Professional Sound Corporation (PSC) 48 149, 163–4
- professional systems 43, 81–2, 323
- program materials 74, 77–8, 144, 226, 273, 303, 322–3
- program producers 295, 317
- programs 1, 46, 59, 272–3, 295–7, 305–6, 310–12
 - Mac 260
 - mono 297, 303
 - Windows 260
- progressive scan modes 32, 38
- progressive content 32, 34
- progressive scanning 31–4, 45, 56
- progressively scanned video 31–2
- project files 231, 265
- prosumer equipment 5, 8, 9, 10, 35–6, 38, 51–2, 170–1
- PSC (Professional Sound Corporation) 48 149, 163–4
- psychoacoustics 303
- pulldown 29–30, 72, 199
- pulse code modulation *see* PCM
- punches 206, 232
- quality 1–3, 9–10, 17–18, 26–7, 59, 62, 81
 - modes and codecs 41–2
 - technical 5, 59, 85
 - video 41, 170
- radio frequencies 111, 133, 135, 137
- radio microphones 16, 64–5, 90–2, 96–7, 109–10, 112–14, 146
 - lavaliere 89, 112
 - radio part 133–8
- Raiders of the Lost Ark* 85, 202, 211, 243
- RAM (random access memory) 225–6
- random access editing 227
- random access memory *see* RAM
- range
 - dynamic *see* dynamic range
 - frequency *see* frequency, range
 - full 21, 214, 322
- ratio control 275–6
- RCA connectors 8–9, 79, 83–4, 179–80

- re-recording mixers 146, 205, 220
- real time 173–4, 209, 227–9, 232, 239, 260, 265
- Real Time Audio Suite (RTAS) 260
- real-time native processing 260
- real-time streaming transfer 169, 171
- realism 208–13, 215, 217, 224
- reality
 - shifting levels of 219–21
 - stretched 213–15, 224
- receivers/receiving equipment 110–11, 133–7, 178, 182–3, 190, 246, 318
- Record Run 47, 48, 57
- recorded level 23, 139, 141, 155, 257, 273–4, 292–3
- recorders 6, 17, 20, 23, 69–71, 141, 190–3
 - analog 207
 - audio 6, 18, 21–2, 40, 50, 69–70, 193
 - dedicated 5–6, 59, 66–7, 94, 189, 192
 - double-system 195
 - field 22, 71, 191, 204
 - handheld 190
- recording chain 77–8, 141
- recording formats 18–19, 25, 32, 189
 - audio 6, 9–10, 95, 172
- recording media 4–5, 20, 40–4, 65, 75, 167, 190–1
- recording medium 5, 20, 40, 42–4, 65, 75, 190–1
- recording modes 4, 10, 18, 42
- recording studios 204, 248–9
- recording time 41–2, 46, 47, 49, 191
- recording(s)
 - 12-bit 19–20
 - analog 10
 - backup 141
 - continuous 2, 184
 - dialogue 98, 140, 258
 - digital 2, 25, 74, 75, 141, 171, 200
- double-system 5–6, 17, 27, 64–5, 68–72, 191–2, 195
 - audio 192, 194
 - management 189
 - setups 65, 69–71, 72, 174, 189–90
- file-based 5, 17, 168
- four-channel 18
- long 52
- primary 141, 195–6
- single-system 6, 64–7, 159
- stereo 95, 246
- two-channel 20, 23, 69–71
 - stereo 15, 246
- voice 125, 274
- redundancy 95, 98, 118, 167
- reference levels 74–6, 77, 80, 82, 115, 143, 305–8
 - digital 75, 78–9
 - and headroom 74–6
 - nominal 75, 78
 - SMPTE 75, 76, 273, 291, 313, 328
- reformatting 167, 169, 200
- release times 276, 278, 292
 - short 275–6
- REO Speedwagon 152
- repetitions 29, 63, 237
- resistors 109, 149–51
- resolution 4–5, 17, 19–20, 54, 53, 226
 - and aspect ratio 36–40
 - full 44, 49–50, 167, 171, 295
 - low 44, 171
 - picture-sound sync 264–5
- restaurants 63, 210
- reverb *see* reverberation
- reverberation 63–4, 86, 89–90, 98–9, 124–6, 285–6, 304–5
 - devices 284–6, 288–9
 - excessive 62, 64, 86, 98, 125–6
 - plug-ins 240, 284, 286
 - recorded 87, 256
 - time 63, 284–5
- reverberators 210, 262, 285, 292
- reverse 256
- RF Overmodulation Protection 318
- right surround 261, 288, 303–4, 308
- rights 232, 247
- ring 79–80, 83, 121
- RJ-45 connectors 178

- Road to Perdition* 220
 roll-off, natural 307
 room noise 90, 96, 112, 285
 room tone 88, 117, 202–3, 237, 241, 243, 245
 rooms 63, 144–5, 253–4, 274, 279, 307, 323
 acoustics 131, 279
 control 310, 322
 editing 42–3, 68, 193, 198, 310, 322, 324
 Foley 204
 large 318, 326
 mix 269, 310
 quiet 94, 143–4
 small 112, 279, 284, 318, 322–3, 326
Rope 217 (film by Alfred Hitchcock)
 routing 265, 288, 291–3, 308
 limitations caused by 288–90
 ruggedness 99, 103, 119
 Rydstrom, Gary 209, 215, 243, 256
- S/PDIF 174, 178–82, 185
 sample rate 17–19, 21, 29–30, 51, 52, 182–6, 188–9, 198
 48-kHz 21, 116, 226
 audio 29, 52
 conversion 183, 188, 198
 locked and unlocked 51–3
 standard 19, 199
 samples 29, 51, 52, 188–9, 191, 200, 230, 263
 satellite 300, 313, 316, 319, 327
 analog 75, 306
 HD 327–8
Saving Private Ryan 215
 scanlines 31–2
 scene changes 203, 219, 247–9, 251–2, 254–5
 scene coverage 88, 97, 233
 and microphone technique 88–92
 scene titles, bonus features 111–13
 scenes 59–64, 85–9, 109–13, 201–5, 208–12, 237–8, 244–52
 deleted 109, 111
 dialogue 63, 203, 208, 248
 funeral 249
 incoming 247, 249, 251
 jungle 222, 244, 304
 opening 211, 216
 Schoeps microphones 122, 152–4, 157, 161–4
 scored music 219, 247
 scores 202, 220, 222, 232–3, 249–50, 252–3, 265
 music cue 254–5
 orchestral 202, 276
 scouting of locations 59, 61–4, 84, 112, 144
 screeches 257
 screen channels 205, 286
 screens 28, 31, 40, 52, 53, 243–4, 246, 286
 screw-in devices 152, 161
 scripts 94, 99, 115, 216, 274
 scrubbing 228
 SD (Secure Digital) 55, 171–2, 190–1
 SDDS (Sony Dynamic Digital Sound) 299–300
 SDHC memory cards 5, 172, 190
 SDII (Sound Designer II) 186, 189
 SDXC memory cards 172
 seamless edits 230, 247
 second channel 95–6, 141, 157
 second sticks 69
 Secure Digital *see* SD
 Sennheiser microphones 122, 165–6
 sensations 15–16, 280
 sensors 26, 35
 image 4, 26, 33–6
 sequences 27–8, 45–6, 216–19, 240, 245, 248–9, 256
 field 29, 56
 time code 46, 48
 set-top boxes 318
 setups 6, 40, 65, 67, 77, 85, 325
 double-system 64–5, 72, 174, 189–90
 studio 70

- shock mounts 102–3, 105, 148
- shooting 4, 7–8, 34, 61–4, 91–3, 169–70, 235–7
 - situations 3, 20–1, 60, 94, 143, 239
 - time of 16, 93
- shotgun microphones 128, 138, 157, 165
 - long 128, 130
 - short 128, 130
- shots
 - close-up *see* close-ups
 - dolly 128, 131
 - establishing 62–3, 92, 131, 249–52
 - interior 245
 - over-the-shoulder 235–6
 - wide 64, 85, 88–9, 92, 112–13, 131, 133
- Shure mixers 122, 154–5, 164
- signal flow 159, 262, 272, 291
- signal paths 221, 301, 305–6
- signal processing 239, 249, 288–9
- signal-to-noise ratio 76
- signals
 - AES3 180
 - amplitude 18, 26
 - analog 4, 78, 176
 - audio 9, 81, 121–2, 178, 319
 - color 54, 56
 - digital 10, 25, 173–4, 176
 - interlaced 31, 34
 - left 81, 301
 - LFE 323
 - line-level 8, 79, 154
 - microphones level 73, 78
- mono 83, 96
 - right 81, 121, 301
 - S/PDIF 179, 181
 - test 287, 303, 310–11
- silence 144–5, 153, 203, 229, 321
- sine-wave tones 308–9, 321
- sine waves 287, 307–8
- singers 250, 252
- single-lens reflex (SLR) *see* SLR digital cameras
- single-speaker surround 315
- single-system recording 6, 64–7, 159
- skin 98, 110, 293
- slates 64, 68–9, 72, 115–16, 192–6, 200, 209
- sleeves 80, 83, 121
- SLR digital cameras 4, 16, 22, 35, 170–1
- small productions 67, 68
- small rooms 112, 279, 284, 318, 322–3, 326
- SMPTE reference level 75, 76, 273, 291, 313, 328
- SMPTE time code (*see* time code)
- snats 181–2, 185, 200
- sneezing 258, 268
- softest sounds 14–15, 20
- software 185, 193–4, 197–8, 230, 268–70, 300, 305
 - devices 29, 259
 - editing 34, 43, 196, 225, 227, 263, 308
 - encoding 314
 - switches 182
- Solaris* 217–18
- solid-state media 53, 170–1
- solo 262
- songs 220, 252, 286
- Sony 22, 171–2, 190
- sound *see also* *Introductory Note*
 - and camera workflow 64–73
- sound amplitude *see* amplitude
- sound cards 79–80, 83
- sound clips 193, 196, 198, 228, 233, 240, 257
- sound crew 9, 62, 66–7, 98, 113–14
- sound decay 63–4
- sound department 198
- sound design 1, 3, 85, 153, 201–24, 246
 - as art 222–3
 - on-screen 214–15
 - spotting 223–4
- sound designers 186, 203, 209, 223–4, 257, 264
- Sound Devices 22, 67, 71, 122, 153, 190–1

- sound editing 87, 115, 187, 228, 231–5, 238–9, 249
 - systems 2, 263, 265
 - as compared with picture editing systems 263–4
 - and pictures 263–4
- sound editors 96, 198–9, 205, 208, 231, 238–9, 263–4
 - ability 231
 - supervising 211, 217, 224
- sound edits 49, 98, 203, 210, 231, 248, 265
- sound effects 99, 204–5, 209, 211–12, 215, 242–5, 256–7
 - editing 242–5
 - libraries 99, 205, 232, 244, 247, 265, 304
 - moving objects 224
 - recordists 211
 - simple 231
- sound fields 15, 95, 125, 129, 301–2, 304–5, 317
- sound formats 25, 297–304
- sound insertion 240, 259
- sound kit accessories 117–18
- sound level meters 144–5, 307, 309–10
- sound levels 14, 142–3, 152
- sound localization *see* localization
- sound mixers *see* mixers
- sound persons 1, 8, 47, 59–61, 66, 69–70, 84
- sound perspectives 88, 91, 132, 235–6
- sound postproduction 109, 111, 157, 205, 264, 297
- sound pressure level (SPL) 11, 13–15, 142–3, 145–8, 152–5, 163–5, 309–10
- sound quality *see* quality
- sound recorders 69–70, 72, 94
- sound recording 14, 25, 59, 62–3, 69, 88, 93
 - production *see* production sound
- sound sources *see* sources
- sound sync *see* sync
- sound systems 264, 278, 287, 293, 295, 303, 322
 - cinema 13–14
 - flat near-field 326
- sound tracks 3, 11, 20, 62, 85, 285, 327
 - four dimensions of 10–17
 - multiple 213, 219
- sound transitions *see* transitions
- sound waves 11, 15, 25, 211
- Soundtrack Pro 183, 196, 204, 249
- source machines 181, 183, 185, 293
- source media 49, 319
- source music 202, 210, 212, 220, 247, 249–50, 252–3
- source tapes 44, 48–50, 183
- source tracks 49, 220
- sources 15–16, 79–80, 124–6, 129–30, 197–8, 235–6, 311–13
- SP (standard play) 42
- space 10–11, 15–16, 41–2, 94–5, 246, 284–6, 296
 - definition 15
- spaciousness 16
- speakers
 - front 316, 325
 - level 73–4
 - monitor 198, 322
 - surround 304, 315, 317, 325
- speech 7, 12–14, 87, 129, 143–5, 274–6, 278
- Speech Compression 318
- speed changes 72, 256
- SPL *see* sound pressure level
- split tracks 197, 234, 238, 291
- spotting 223–4
- staggered levels 118, 157
- standard-definition 4–5, 32, 38, 41, 56
 - DV formats 18, 38, 226
 - television 31
 - video 41, 56
- standard play (SP) 42

- standards 5, 30–1, 35, 53, 56, 305, 326
 - for audio 5–10
 - de facto 267
 - high-definition 32, 36, 38
 - minimum 5–6
- standing waves 274, 279, 323
- Star Wars* (223, 225)
- stem mixing 261, 294
- stereo 16, 80, 83, 95, 226, 297–8, 301
 - microphones 95, 121, 126, 165
 - receivers 297
 - recordings 95, 246
 - two-channel 94–8, 288, 297–8, 304, 313, 317, 327
 - recordings 15, 246
 - sources 302
- sticks 69, 193, 195
- storage
 - cost of 44, 226
 - digital audio 189
- story 3, 86, 98, 218, 224, 231–2, 239
- storytelling 1, 3, 224, 242, 286
 - elements 62, 87, 203
- streaming transfers 171–88, 191–2, 200
- streets 124, 130, 143, 221, 235, 254
- stretched reality 213–15, 224
- studio environment 1, 70, 301
- studio equipment 182
- studio master clock 182
- studio masters *see* delivery masters
- studio setups 70
- studios 53, 78, 83, 93, 221, 316, 327–8
 - Hollywood 75, 99, 206
 - postproduction 306, 320
 - recording 204, 248–9
- styles 3, 112, 208, 232, 261, 272, 290
- subharmonic synthesis 257
- subwoofers 316, 319, 323, 325–6
- supercardioid microphones 126, 128, 130
- surrealism 216–18, 224
- surround 16, 91, 96, 204–5, 216, 299, 303–6 *see also* surround sound
 - channels 246, 286, 298, 301, 304, 315, 327
 - decoders 246, 288, 302
 - left 261, 288, 303, 308, 317
 - magic 302
 - matrix 298–300
 - mixes 21–2
 - mono 315
 - right 261, 288, 303–4, 308
 - single-speaker 315
 - sound 15–16, 298–9, 301, 303–5, 313, 327
 - discrete 297–8
 - fields 294, 302, 313
 - speakers 304, 315, 317, 325
 - switches 139, 158–9, 166, 175, 258
 - software 182
 - sync 29–30, 35, 51–3, 68–70, 72–3, 193–200, 262–4
 - audio-video 52
 - dialogue 62, 196
 - drifts 52, 56, 72
 - errors 200
 - hard 193–4, 196–7, 200, 205, 235, 327
 - jam 68, 193
 - near-sync 196, 205
 - picture-sound 263–4
 - problems 35, 91, 115
 - references 64, 68, 193–6, 200
 - sound 93–4, 131, 209, 216, 250–1, 264
 - testing 72–3
 - Synchro Arts 198, 257
 - synchronization *see* sync
 - syncing 6, 113, 193–6, 200
 - jam 68, 193
 - manual 194
 - synthesizers 286–7
- tablets 4, 9
- tails 227–8, 231, 233, 237, 256–7, 263, 268
- tails-to-heads 237, 257
- talkers 11, 86, 98, 101, 105–6, 109, 129–30

- tape 2, 43, 47–9, 52–3, 167–71, 173–4, 295–6
 - analog 29
 - digital 174, 192
 - DV 20, 37, 53, 56, 168, 171
 - Gaffer's 117
 - hairpiece 110, 117
 - HDCAM 17, 21–2
 - new 49, 168
 - source tapes 44, 48–50, 183
 - VHS 45, 297, 306
- tape-based cameras/camcorders 94, 174
- tape-based formats 25, 51, 53, 189
- tape-based workflows 5, 43
- tape for lavalier mounting 117
- tape formats 5, 10, 18, 21, 34, 42, 168
 - digital 170, 300
- tape machines 79, 82, 179, 293, 312
 - analog 256, 261, 273
- tapeless media 25, 65
- tapeless workflows 43, 173
- Tascam 22, 83, 190
- telecine 29, 34, 72, 173, 199
- telephones *see* phones
- television 1, 27–30, 32, 45, 69, 88, 142
 - black-and-white 28–9
 - broadcast 36, 45–6, 281, 297
 - color 28–9, 98
 - digital 135, 319
 - digital satellite 313
 - high-definition 31, 36, 55, 298
 - network 297
 - NTSC 45, 56, 318
- temp music 231–2
- Terminator 2* 15, 273
- test signals 287, 303, 310–11
- theatrical distribution 36, 295, 310
- theatrical masters 295, 327
- theatrical mixes 326
- threat levels 209, 222
- threshold controls 275, 278
- thresholds
 - hearing 139, 153
 - level 276, 284
 - of pain 11, 13, 14, 20, 152
 - of tickle 11, 13, 15
- timbre 90, 96, 98, 245, 275, 278–81, 294
- time
 - clock 46, 48
 - continuous 213, 249, 251–3, 286
 - real 173–4, 209, 227–9, 232, 239, 260, 265
 - recording 41–2, 46, 49, 191
- time-based devices 284–8
 - reverberation 284–6, 288
- time-based effects 286
- time code 40, 44–51, 65, 68, 115–16, 193–6, 198
 - autosync 193–4, 200
 - camera 47, 48, 69
 - common 69
 - DF and NDF 47
 - generators 48–9
 - recording 45, 48, 50
 - sequences 46, 47, 48
 - sync 68, 194–5
- time compression-expansion 256
- Time Division Multiplex (TDM) 258, 284, 290
- timelines 196, 293
- titles 29, 202, 219–21, 250, 286
- To Wong Foo, Thanks for Everything, Julie Newmar* 111–13
- tonal sounds 216, 237, 283
- tone arm 225
- tones 26, 77, 111, 115–16, 154, 296, 308–9
- Top Gun* 244
- Touch of Evil* 221
- track layout 197, 209, 233, 321
- track splitting 234, 265
- tracks
 - ambience 203, 208, 243, 245
 - audio 196, 248, 315
 - background 242, 246, 259

- and channels 261
- character 238
- dialogue 203, 207–9, 242, 259, 276
- grouping 262, 265
- lavaliere 197, 291
- multiple 16, 219, 241
- music 207, 294
- optical 156, 281
- separate 204, 242, 262
- source 49, 220
- split 197, 234, 238, 291
- wild 93
- transducers 119, 139
- transduction 120, 123
- transfers 2, 29, 35, 50, 51, 169–70, 191–2, 200
 - data 55, 191
 - digital audio 180
 - file 53, 167, 169, 171–88, 191–2, 200
 - film 34, 249
 - streaming 171–88, 191–2, 200
 - video 29, 169, 173–4, 176, 191
- transformers 80–2, 180–1, 185
- transitions 216, 220, 231, 248, 250, 252, 265
- transmissions 277, 319
 - digital 136–7
- transmitters 110–11, 135–8, 178, 277
- treble 11–12, 119, 123, 132, 142, 278–9, 285
- trims 227–8, 310
- TRS connectors 80, 83
- tunes 210, 221, 287, 292
- TV *see* television
- two-channel audio 10, 22–3, 69–71, 296–8, 300, 314–15, 322
 - delivery 298–9
 - library sources 304
 - stereo 94–8, 297–8, 304, 313, 317, 327
 - recordings 15, 246
 - sources 302
- ultrasonic levels 176
- unbalanced audio 81, 83
- United States 34, 36, 46, 74, 205, 283
- unity 258, 308
 - gain 79, 160, 269, 287, 308
- unlocked audio 52–3, 56, 185
 - sample rate 51–2
- user bits 50–1
- VCRs (videocassette recorders) 318
- verbal identification 69
- Vertical Limit* 214
- VHS tape 45, 297, 306
- Video
 - 1080-line 38, 40
 - 1280 × 720 38
 - analog *see also* analog video
 - basic introduction 27–41
 - high-definition 1, 5, 38, 41, 44, 56
 - interchangeability 4, 42–4, 312
 - interconnecting 53–7
 - interlaced 28, 30–4, 45, 56
 - look as compared with film 32–6
 - NTSC 27, 29–30, 32–5, 56
 - operational matters 41–53
 - PAL 51, 56
 - scanned 31–2
 - standard-definition 41, 56
 - widescreen 36–9, 56, 221
- video cameras 4, 30, 33, 61, 64, 68, 77
 - analog 27
 - digital 4, 32, 36, 52, 53, 140
 - professional 68
 - digital 10, 50, 171
 - single digital 194
- video clips 185, 193–4
- video clocks 51, 52, 56, 115
- video compression 4, 25, 44
- video editors 97, 248
- video fields 28, 30–1
- video files 167, 188–9, 191–2
 - digital 313
- video formats 2, 5, 40, 47, 56
 - digital 25, 37
- video media 69, 192, 200
 - management 191

- video microphones 126, 128
- video mixes, v film 326–8
- video modes 4, 42, 56–7
 - anamorphic 56
- video monitors 4, 61
- video production 3, 16, 42, 69, 73, 115, 119
- video tape 64, 173, 327
- video technology, digital 25
- video transfers 29, 169, 173–4, 176, 191
 - streaming 174, 176, 191
- viewfinders 53, 61, 114
- virtual mixing 270
- Virtual Studio Technology (VST) 260
- visual waveform editing 228–9
- VocALign 257
- vocals 105, 223, 249–50
- voice coil 123
- voice limitation 290
- voice recordings 125, 276 *see also*
 - dialogue
- voices 2, 98, 125–6, 214, 276–9, 286, 290–1
 - actor 29, 51, 87, 137
 - male 279–81, 291
 - raised 277
- voltage 26, 73, 78–9, 82, 121–3, 139
 - dividers 149–50
 - levels 79, 82, 139, 308
 - output 80, 82, 123
- voltmeters 82, 308
- volume 104, 269, 277, 305–6
 - graphing 234, 239, 258, 268, 270, 308
 - output 308
- vowels 155, 196, 276
- VU meters 75, 79, 309, 311

- walkie-talkies 214
- walls 15, 63, 125, 322–3, 326
- WAV 186–9, 318
 - broadcast 187, 189
- waveforms 9–10, 19, 69, 71, 186, 194–6, 228–9
- waves, standing 274, 279, 323
- wedding receptions 252
- weddings 64, 97, 249
- weighting 14, 91, 141–2, 146, 307, 309
- Where Eagles Dare* 215
- whispers 13, 141, 153–5, 275, 285
 - and cries 141, 153, 155–8
- wide shots 64, 85, 88–9, 92, 112–13, 131, 133
- widescreen video 36–9, 56, 221
- wild lines *see* wild track
- wild track 93, 208, 235, 291
- wind 103–5, 112, 119, 128–9, 138, 161–3, 166
 - features for best performance 162–6
 - machines 208, 235
- Windjammer 104, 162, 164, 279
- Windows programs 262 *see also*
 - software
- windscreens 102–5, 118, 148, 163–4, 279
 - foam 104–5
 - silk windscreens 104, 279
- wireless receivers 77
- Wonderful Horrible Life of Leni Riefenstahl, The* 219
- word clock input 182, 185
- word length 17–22, 184, 189, 226
- workstations 184, 187, 206, 225, 231, 268, 293
- worldizing 220–1, 240, 249

- X curve monitors 307, 318
- XLR
 - cables 148, 163–4
 - connectors 7–8, 73–4, 83–4, 149–51, 174, 176, 185
 - extensions 160
 - inputs 18, 77, 190

- zoom lenses 4, 60, 94